

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/168966>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Utilising Ecology and Molecular Techniques to Identify
Alternative Hosts, Green Bridges, and Population Structure
of a Serious Aphid Pest

Dion Garrett

January 2022

A thesis submitted to the University of Warwick in fulfilment of the
requirements for the degree

of Doctor of Philosophy of Life Sciences

Project supervisors:

Dr. James Bell, Rothamsted Research

Dr. Ramiro Morales-Hojas, Rothamsted Research

Dr. Graham Teakle, University of Warwick

Dr. Rosemary Collier, University of Warwick

Emma Garfield, Gs Fresh



Declaration

This thesis contains my own work which has not been submitted for a degree at another University. The illustrations at the beginning of each chapter and throughout the thesis are also my own work.

Acknowledgements

Firstly, I would like to thank all of my supervisors, James Bell, Ramiro Morales-Hojas, Graham Teakle, Claire Donovan and Emma Garfield. James has been fantastic throughout the PhD. From words of encouragement, modelling assistance, general discussions, and the occasion reality check. Anyone who manages to get James as a supervisor would be very lucky indeed. Ramiro for his continued support on the molecular aspects of the project, technical assistance, helpful tips, and advice – plus having to endure the monthly minute reports. Graham for his excellent knowledge of molecular theory and support throughout the entire project. Rosemary for her practical knowledge and seemingly endless knowledge with anything field related. Claire for helping me create strong links to my industry partner, Gs Fresh. Emma for her support throughout the project, creating and reinforcing strong ties to many Gs Fresh colleagues and her invaluable assistance with the project and working alongside Gs Fresh.

The Insect Survey team have been amazing since day one. Thanks go to Alex Greenslade's plethora of Hemipteran knowledge and assistance with any aphid question I had. I would also like to thank Paul Verrier for help with the Insect Survey database, Tracey Kruger and Lynda Alderson for their support throughout the project, Gemma Ford for technical assistance and just being Gemma. Chris Shortall for support, both technical and moral. Mike Hall for being a great bloke and providing me with the pop-up suction traps used in the project (including setting them up on site) – and lending me random tools at a moment's notice!

Both the University of Warwick and Rothamsted Research have been great throughout and made the PhD experience enjoyable. I would like to thank Marion Elliot for providing me with bountiful supply of *N. ribisnigri* cultures from the University of Warwick. Donna Fellowes for her PhD support throughout (Rothamsted). All of the Gs Fresh staff that I have interacted with during the 4-year project, you made my project much more enjoyable. The Gs Fresh agronomist, David Norman, who provided invaluable insight into *N. ribisnigri* behaviour. I would also like to thank the bioinformatics team (Rothamsted), especially Rob King, David Hughes and Emma Bailey. Rob provided me with the knowledge and experience to confidently *de-novo* assemble the *Nasonovia ribisnigri* genome. I would also like to mention and thank the bioinformatics sessions held throughout the PhD at Rothamsted Research – without these, my project would have been so much harder.

I would like to thank my family (Mum, Dad, Al, Nan and Grandad) for their support throughout the entire project. Particularly my Mum who has been very supportive once I made the decision to pursue an academic career – I know you had your reservations but look where I am now :). I would like to thank Hannah for making me apply for an undergraduate degree and supporting me throughout it. I most likely would have ended up making books in Bungay for the rest of my life and never embarked on an academic journey to begin with. Finally, Kelly for all her help in numerous components of the PhD. From very long days of fieldwork, counting plants, collecting samples to listening to endless musings about the project and being very understanding when many a weekend was written off due to laboratory work or other PhD-related activity. This thesis would have been nowhere as complete or as good without her help and I doubt I would have made it if it wasn't for her.

Abstract

Nasonovia ribisnigri (Mosley, 1841) is the most damaging pest on outdoor lettuce crops during the summer months. Control measures to ameliorate yield loss are hampered by the cryptic ecology of this pest aphid, with low numbers trapped, and an autecology of feeding on concealed young leaves. A lettuce cultivar introduced in the 1980's provided single gene resistance (Nr-gene) against *N. ribisnigri*. However, over reliance of this cultivar led to resistance-breaking biotypes arising during 2007.

This multifaceted thesis aimed to address unknown areas of *N. ribisnigri* ecology. Over 50 years of UK suction trap data were studied to understand how the changes in climate and agricultural practises have influenced population trends. Different trapping methodology was incorporated to improve successful capture of this cryptic aphid to help improve current forecasting models. Various semi-field and field experiments were established to identify alternative summer host plants that could be used by *N. ribisnigri* as a green bridge between secondary hosts and lettuce. Additionally, the project aimed to assemble a draft genome of *N. ribisnigri* to provide insight into the level of gene flow between populations in England, and to study the potential resistance-breaking mechanism in lettuce cultivars containing the Nr-gene.

Firstly, population trend analysis revealed significant declines in the abundance of *N. ribisnigri* in the UK since 1965. Secondly, multiple trapping methods elucidated the low dispersive behaviour of *N. ribisnigri*. Thirdly, the importance of green bridge vegetation in connecting occurrence of this host-alternating species from its primary host to the lettuce crop was expounded with field surveys and gut content analysis. This work suggests *N. ribisnigri* is using an intermediate secondary host plant prior migration onto outdoor lettuce later in the year. Next, the first draft reference genome and transcriptome for *N. ribisnigri* were *de novo* assembled. Subsequently, population genetic analysis revealed a low level of gene flow between populations in England and highlighted a divide between the East and West populations of *N. ribisnigri*. Finally, RNA-seq analysis of susceptible and resistant *N. ribisnigri* biotypes revealed a single candidate resistant gene likely responsible for the resistance-breaking mechanism in lettuce cultivars containing the Nr-gene.

The work in this thesis has already been applied by growers, this may be extended to improve the monitoring and field control of *N. ribisnigri*. Promising avenues have been opened for further research and development which may be extended to other invertebrate pests, including dietary analysis to identify previous host plants in aphids, techniques in population genetics, and analysis of resistance mechanisms.

Contents

1.....	2
2. CHAPTER 1: Introduction.....	2
1.1 Aphids as crop pests	2
1.1.1 Aphid lettuce pests	4
1.2 Taxonomy of the currant-lettuce aphid, <i>Nasonovia ribisnigri</i>	5
1.3 Distribution.....	9
1.4 Virus transmission in lettuce and associated hosts.....	9
1.5 <i>Nasonovia ribisnigri</i> Biology and lifecycle.....	10
1.5.1 Winter (primary) hosts.....	11
1.5.2 Summer (secondary) hosts	11
1.5.3 Lettuce host biology.....	14
1.6 Organic and conventional practices	15
1.7 Monitoring and control of <i>Nasonovia ribisnigri</i> in lettuce	16
1.7.1 Insect traps.....	16
1.7.2 Biological control and IPM	18
1.8 Resistance in <i>N. ribisnigri</i>	21
1.8.1 Insecticide resistance in <i>Nasonovia ribisnigri</i>	21
1.8.2 Resistant-breaking <i>Nasonovia ribisnigri</i> biotypes (Nr-gene in lettuce)	22
1.9 Aphid genetics	24
1.10 Thesis aims and objectives	26
1.11 References.....	29
3. CHAPTER 2: Long-term trends in the abundance of <i>Nasonovia ribisnigri</i> in the UK; forecasting and improving trapping methodology	44
2.1 Introduction.....	44
2.1.1 Migration biology.....	44
2.1.2 Sampling aphids during flight.....	45

2.1.3	Forecasting migration	45
2.2	Materials and methods.....	47
2.2.1	Experiment 1: <i>Nasonovia ribisnigri</i> – developing a model to describe annual abundance trends	47
2.2.2	Experiment 2: First flight model for <i>Nasonovia ribisnigri</i>	50
2.2.3	Experiment 3: Investigating the height at which <i>Nasonovia ribisnigri</i> alates fly during appetitive flight	51
2.2.4	Experiment 4: Citizen science and social media	57
2.3	Results	58
2.3.1	Experiment 1: <i>Nasonovia ribisnigri</i> – developing a model to describe annual abundance	58
2.3.2	Experiment 2: First flight model for <i>Nasonovia ribisnigri</i>	62
2.3.3	Experiment 3: Investigating the height at which <i>Nasonovia ribisnigri</i> alates fly during appetitive flight	69
2.3.4	Experiment 4: Citizen science	71
2.4	Discussion	74
2.5	References	84
4.	CHAPTER 3: Alternative summer hosts, green bridges, overwintering biology and using a novel molecular method to detect previous host plants of <i>N. ribisnigri</i>	92
3.1	Introduction	92
3.2	Experiment 1 & 2: Field margins, alternative hosts, and green bridges	93
3.3	Experiment 3: Identifying host plants in <i>Nasonovia ribisnigri</i>	94
3.4	Experiment 4: Overwintering biology	96
3.5	Materials and methods.....	99
3.5.1	Experiment 1: Field margins surveys to identify potential <i>N. ribisnigri</i> host plants	99
3.6	Experiment 2: Alternative host plant plots to encourage <i>N. ribisnigri</i> colonisation.....	105
3.6.1	Study site.....	105
3.6.2	Alternative plant host plot experiment design	108

3.6.3	Experiment 3: Using molecular techniques to identify the host plant barcode (ITS2) of <i>Nasonovia ribisnigri</i> to identify potential green bridge reservoirs.....	111
3.6.4	Experiment 4: Overwintering biology	115
3.7	Results	118
3.7.1	Experiment 1: Field margins surveys to identify potential <i>N. ribisnigri</i> host plants	118
3.7.2	Experiment 2: Alternative host plant plots to encourage <i>N. ribisnigri</i> colonisation	122
3.7.3	Experiment 3: Using molecular techniques to identify the host plant barcode (ITS2) of <i>N. ribisnigri</i> to identify potential green bridge reservoirs	125
3.7.4	Experiment 4: Overwintering biology	127
3.8	Discussion	131
3.9	Conclusion	136
3.10	References.....	138
5.	CHAPTER 4: Genome assembly and transcriptomic analyses to study the resistance-breaking mechanisms in host plant resistant biotypes to <i>Nasonovia ribisnigri</i>	143
4.1	Introduction.....	143
4.2	Materials and methods.....	147
4.2.1	Genome assembly and annotation	147
4.2.2	RNA-seq methods and differential gene expression analysis	152
4.2.3	Validation of RNA-seq data by qRT-PCR.....	155
4.3	Results	158
4.3.1	Genome <i>de novo</i> assembly and annotation	158
4.4	Discussion	175
4.5	References	180
6.	CHAPTER 5: Population genetic analysis of <i>Nasonovia ribisnigri</i> reveals a temporal divide between the East and West of England	190
5.1	Introduction.....	190
5.2	Materials and methods.....	193
5.2.1	Study organism and site locations	193

5.2.2	DNA extraction.....	195
5.2.3	Microsatellite discovery	195
5.2.4	Validation, PCR amplification and sequencing of microsatellites	196
5.2.5	Population analysis	200
5.3	Results	202
5.3.1	The population structure of <i>Nasonovia ribisnigri</i>	203
5.3.2	Temporal analysis of <i>Nasonovia ribisnigri</i>	208
5.4	Discussion	215
5.5	Conclusion	219
5.6	References	221
7.	Chapter 6: General Discussion.....	228
6.1	New insights into the field behaviour of <i>Nasonovia ribisnigri</i>	228
6.1.1	Declines in <i>Nasonovia ribisnigri</i> populations in the UK	228
6.1.2	Migration and Flight Behaviour	230
6.1.3	Alternative host plants.....	232
6.2	Population processes and resistance mechanisms in <i>Nasonovia ribisnigri</i>	233
6.2.1	Population genetics.....	233
6.2.2	Genome and insights into resistance-breaking biotypes.....	234
6.3	Future trends and management of <i>Nasonovia ribisnigri</i>	235
6.3.1	Concluding remarks	236
6.4	Key findings.....	236
6.4.1	Recommendations for growers	237
6.4.2	Recommendations for research.....	237
6.5	References	238
8.	Appendix 1: Chapter 2 Supplementary material	241
9.	Appendix 2: Chapter 3 supplementary material	275
10.	Appendix 3: Chapter 4 supplementary material	283

9.1	Supplementary 1: A step-by-step guide to hybrid de novo genome assembly in the command line (Bash script)	283
9.1.1	Annotation of the <i>N. ribisnigri</i> genomes.....	291
9.1.2	Constructing the transcriptome using RNA-Seq data	295
9.2	Supplementary 2: k-mer graphs for the <i>N. ribisnigri</i> genomes	296
9.3	Supplementary 3: RNA-seq analysis using DeSeq2 in RStudio	299
9.4	Supplementary 4: PCR and qRT-PCR primers designed from RNA-seq validation of gene 2698 and 3 house-keeping genes (HKGs).....	331
11.	Appendix 4: Chapter 5 supplementary material	337
9.5	Supplementary 1: Chapter 5 – Population genetics DNA yield analysis between sites and years	337

List of figures

Figure 1.1: Morphology of <i>Nasonovia ribisnigri</i> highlighting the main characteristics of a winged adult (alate), non-winged adult (apterous) and a nymph. Figure produced using Adobe Illustrator.	6
Figure 1.2: The currant-lettuce aphid, <i>Nasonovia ribisnigri</i> , highlighting some of the characteristic features used for taxonomic identification. A) Alate <i>N. ribisnigri</i> feeding on lettuce, B) Alate <i>N. ribisnigri</i> under x 20 magnification using a stereomicroscope, with finger-like cauda, siphunculi and distinct abdominal markings and, C) Apterous <i>N. ribisnigri</i> both adults and nymphs with mixture of yellowish green and red morphs.	8
Figure 1.3: <i>Nasonovia ribisnigri</i> feeding on a host plant (primary or secondary). Unlike chewing insects (e.g. caterpillars, grasshoppers, beetles) which cause large tissue damage, aphids inflict only slight physical damage. Using their long flexible stylet, they can navigate through the host plant cuticle, epidermis, and mesophyll to reach the phloem is which they feed. Figure produced using Adobe Illustrator.	10
Figure 1.4: The holocyclic (sexual) lifecycle of the currant-lettuce aphid, <i>Nasonovia ribisnigri</i> . Primary host is various members of the <i>Ribes</i> spp. Genus (e.g. Blackcurrant (<i>R. nigrum</i>) and gooseberry (<i>R. uva-crispa</i>)) and secondary host are members of the Asteracea Family (e.g. Chicory (<i>Cichorium intybus</i>) and lettuce (<i>Lactuca sativa</i>)). Figure produced using Adobe Illustrator.	13
Figure 2.1: The distribution of the RIS 12.2m suction traps used in the analysis of long-term trends. For regional analysis, the sites were split into four regions as follows: North (7-sites), South (12-sites), East (6-sites) and West (6-sites).	48
Figure 2.2: Study site map in Ely, Cambridgeshire. Fields are displayed in red with field names. Locations of pop-up suction traps, water traps, yellow sticky traps and Malaise traps are circled in black. Additional water traps, yellow sticky traps and Malaise traps were situated in an outdoor lettuce field (Hobbs field – circled in red) nearby to determine whether traps would have better success at capturing <i>N. ribisnigri</i> . ..	52
Figure 2.3: A) Pop-up suction trap schematic render showing the complete set up with handles and wheels for mobility. B) Pop-up suction trap schematic breakdown highlighting each constituent part, including the stainless-steel mesh netting insert. Both diagrams were created using Fusion 360 software, courtesy of Gemma Ford (RIS).	54

Figure 2.4: The 2m pop-up suction trap (left) and the 6m pop-up suction trap *in situ* at the Cambridgeshire farm. Pop-up traps were 5m apart from one another, with no buildings or trees to obscure insect sampling or affect the boundary layer. 55

Figure 2.5: A yellow sticky trap after 3-days of sample collection in an outdoor lettuce field. B) A malaise trap set up *in situ*. Flying insects are intercepted by the fine mesh in the centre of the malaise trap and instinctively move up into the collection bottles located at the top of the trap. It is possible to collect bidirectional samples, to test for differences, as visualised in the image. However, only 1 bottle was used for both directions in the study..... 56

Figure 2.6: Trends of total *N. ribisnigri* abundance in the UK between 1965-2020. The black line represents the trend (i.e. spline), with significant periods of decline indicated in orange. Points and whiskers represent year random effects for each year. A) log-linear year effect of *N. ribisnigri* abundance between 1965-2020. B) non-linear year effect of *N. ribisnigri* abundance between 1965-2020. A 95% confidence interval around the spline for each model is highlighted in blue. All trends are scaled and relative to the starting year of 1965. Red arrow indicates when lettuce cultivars resistant to *N. ribisnigri* were introduced. 59

Figure 2.7: Trends in total abundance of *N. ribisnigri* in the UK between 1965-2020 for each region. The black line represents the trend, with significant periods of decline indicated in orange. Points and whiskers represent year random effects for each year. A) **north** log-linear year effect between 1965-2020. B) **north** non-linear year effect between 1965-2020. C) **south** log-linear year effect between 1965-2020. D) **south** non-linear year effect between 1965-2020. E) **east** log-linear year effect between 1965-2020. F) **east** non-linear year effect between 1965-2020. G) **west** log-linear year effect between 1965-2020. H) **west** non-linear year effect between 1965-2020. A 95% confidence interval for each model are highlighted in blue. All trends are scaled and relative to the starting year (1965) at 1. Red arrow indicates when lettuce cultivars with resistance to *N. ribisnigri* were introduced. All x-axis start at the same position and y-axis for the West is slightly larger due to the size of the error bars between 1970 – 1975. 61

Figure 2.8: Phenology of the currant-lettuce aphid (*N. ribisnigri*) captured and recorded in all active 12.2m RIS suction traps between 1966-2016. Blue arrows indicate time spent on the winter (primary) host and

red arrows indicate when *N. ribisnigri* will be on the summer (alternative) host, including lettuce. Two peaks are in abundance are apparent and a mid-summer crash is evident between them, which is typical for many heterecious aphids..... 63

Figure 2.9: The correlation matrix plot of log daily counts of aphid species captured in the RIS suction traps showing a low correlation with *N. ribisnigri* between 1965 and 2019. 64

Figure 2.10: The correlation matrix plot of log weekly counts of aphid species captured in the RIS suction traps showing a low correlation with *N. ribisnigri* between 1965 and 2019. 65

Figure 2.11: Records of the first flight of the currant-lettuce aphids captured between 1965-2019. The x-axis shows the number of days until the first *N. ribisnigri* was captured in a given year, with the y-axis denoting the year of capture. Only the data from Broom’s Barn, Kirton, Preston, Hereford, Rothamsted Tower, Starcross, Writtle and Wye were used for correlating first flight with degree days..... 67

Figure 2.12: Geographical distribution of 12.2m RIS suction traps that caught *N. ribisnigri* in a given year (day of first flight) between 1965-2019 (54 years). Coloured symbols represent the day when *N. ribisnigri* was first captured at that given location.. 70

Figure 1.13: Total numbers of aphids caught in the different types of trap during 2018 and 2019 at the Cambridgeshire study site. A) Total numbers of aphids caught during 2018 in the 2 m pop-up suction trap, 6 m pop-up suction trap, yellow water trap, white water trap, green water trap and Malaise trap, B) Total numbers of aphids caught during 2019 in the 2 m pop-up suction trap, 6 m pop-up suction trap, yellow water trap, white water trap, green water trap and Malaise trap, C) Total numbers of *N. ribisnigri* captured during 2018 in all types of trap, D) Total numbers of *N. ribisnigri* caught during 2019 in all types of trap.....70

Figure 2.14: Distribution of volunteers throughout the UK. Map created using batch geocoding using doogal (Bell, 2021).....72

Figure 1.15: Number of responses from volunteers and samples obtained. Aphids found included: the peach-potato aphid *Myzus persicae*, potato aphid *Macrosiphum euphorbiae* and the glasshouse-potato aphid

<i>Aulacorthum solani</i> . Other insects found included: planthoppers and leafhoppers (Auchenorrhyncha), adult and larvae of lacewings (Neuroptera) and ladybird larvae (Coleoptera).....	73
Figure 2.16: Number of responses from volunteers and samples obtained. Aphids found included: the peach-potato aphid <i>Myzus persicae</i> , potato aphid <i>Macrosiphum euphorbiae</i> and the glasshouse-potato aphid <i>Aulacorthum solani</i> . Other insects found included: planthoppers and leafhoppers (Auchenorrhyncha), adult and larvae of lacewings (Neuroptera) and ladybird larvae (Coleoptera).....	82
Figure 3.1: The lifecycle of <i>N. ribisnigri</i> with each experiment in this chapter superimposed around it. This demonstrates how each experimental component addresses the current unknown scientific knowledge gaps throughout the <i>N. ribisnigri</i> holocyclic lifecycle. Red arrows indicate the duration and time frame of each experiment.	98
Figure 3.2 A map of the organically managed farm near Ely (Hainey), Cambridgeshire, UK (52°21'N, 0°16'E). Grid overlay highlighting how sampling sites (circled) were identified and chosen. Red circles – AES planted; yellow circles – Semi-natural; green circles – grass. Different margin types on the site are shown in different colours (Green – grass, Red – Flower rich margins (AES), Purple – Organic bird seed mix, Yellow – Bird mix). Only flower-rich margins (Red) were used for the AES margin type. Source: Gs Fresh.....	101
Figure 3.3 A map of the conventionally managed farm near Wisington, Norfolk, UK (52°32'N, 0°27'E). Grid overlay highlighting how sampling sites (circled) were identified and chosen. Red circles – seeded mixes; yellow circles – Semi- natural; green circles – grass. Source – Gs Fresh.....	101
Figure 3.4 Examples of quadrats in A) a flower-rich margin (AES); B) a grass margin; C) Semi-natural habitat.....	101
Figure 3.5 Using a sweep net to sample for invertebrates along one of the 20m transects at the conventional farm in Norfolk (Wisington).....	101
Figure 3.6: The organically managed farm in Ely, Cambridgeshire, UK (52°21'N, 0°16'E). Red area – AES planted; yellow areas – Semi-natural; green areas – grass – expand these descriptions. Black arrows show where the alternative host plant plots were located throughout Experiment 2 in 2020. Field locations: Mill Ground, Sale Ground, Bridge Ground and 21/8 Acre.....	106

Figure 3.7: The conventionally managed farm near Wisington, Norfolk, UK (52°32'N, 0°27'E). Black arrows show where the alternative host plant plots were located throughout Experiment 2 in 2020. Field locations: G3, R54, R20 and B(S). 107

Figure 3.8: Plot layout for the alternative plant host plot experiment. Fifteen host plant plugs were directly planted into the prepared soil as 3 rows of five plants. The alternative host plant plots were 24m long and 4m wide and divided into 5 sub-plots (4m each) to accommodate each host plant (four blocks for both site 1 and site 2, eight in total). The plots in which each host plant species was planted was randomised (Random, 2020). 109

Figure 3.9: A) newly planted plot on site 1 near Hainey in Cambridgeshire (organic). B) the plants were watered for the first 4-weeks to allow them to establish (site 2, near Wisington site in Norfolk (conventional)). 110

Figure 3.10: A schematic diagram of the priming sites within the second internal transcribed spacer 2 (ITS2), and respective flanking regions (5.8S and 26S). UniPlantF primer is within 5.8S and UniPlantR begins on the last bp of ITS2 and continues into 26S. Schematic adapted from Moorhouse-Gann *et al.*, 2018).. 116

Figure 3.11: Lettuce plants infested with high numbers of *N. ribisnigri* nymphs (A). and alates (B) found in a lettuce field in August 2018, during an outbreak at Hainey farm (Organic) (Site 1)..... 117

Figure 3.12: Field cage enclosing primary hosts (*Ribes* spp.) and secondary host (*Lactuca sativa*)..... 120

Figure 1.1: A) number of grass (light green) and broadleaf (dark green) species identified in each type of margin (grass, semi-natural, seeded) during the 3-month study at the Hainey (organic) site. Data shown for each margin type are the combined total from two quadrats for each of the three survey locations (six quadrats in total). B) Percentage cover of both grass (light green) and broad leaf (dark green) species in each margin type (grass, semi-natural, seeded) at the Hainey (organic) site.....120

Figure 3.14: A) number of grasses (light green) and broadleaf (dark green) species identified in each margin type (grass, semi-natural, seeded) during the 3-month study at the Wisington (conventional) site. Data shown for each margin type is the combined total of two quadrats for each of the three survey locations (six quadrats in total). B) Percentage cover of both grass (light green) and broad leaf (dark green)

species identified in each margin type (grass, semi-natural, seeded) at the Wissington (conventional) site.
..... 121

Figure 3.15: A) Transect R54 at the site near Wissington (conventional) with chicory (*C. intybus*) and B) close-up view of an apterous *N. ribisnigri* on *C. intybus* in transect R54 at the same location. Table 3.2: Total number of *N. ribisnigri* identified on chicory (*C. intybus*) alternative host plant plots..... 123

Figure 3.16: Sequence percent identity of the ITS2 lettuce plant barcode barcodes in *N. ribisnigri* after periods of starvation. In this experiment, a lettuce positive control was used throughout DNA extract, PCR and sequencing to verify results. Number of replicates in indicated on the z axis (e.g. rep #1).. 124

Figure 3.17: Overwintering phenology and colonisation of *N. ribisnigri* on winter host plants (currant – *Ribes* spp.) from summer host (lettuce – *Lactuca sativa* cv. Maureen) between 02OCT18-12DEC19. BC - Black currant (*Ribes nigrum*); GB – Gooseberry (*Ribes uva-crispa*); RC – Red currant (*Ribes rubrum*); WC – White currant (*Ribes rubrum*).....124

Figure 3.18: Overwintering phenology and colonisation of *N. ribisnigri* on winter host plants (currant – *Ribes* spp.) from summer host (lettuce – *Lactuca sativa* cv. Maureen) between 06MAR19-30APR19. Data from 01JAN19-05MAR19 not shown due to no activity from *N. ribisnigri*. BC - Black currant (*Ribes nigrum*); GB – Gooseberry (*Ribes uva-crispa*); RC – Red currant (*Ribes rubrum*); WC – White currant (*Ribes rubrum*). Alates first observed on 25APR19 on Black currant..... 125

Figure 4.1: A) BUSCO analysis score of both *N. ribisnigri* genomes (WT = Nr:0, susceptible; R = Nr:1, Resistant) and previously published aphid genomes (obtained from BIPAA genomic resources) using an Arthropod gene set (Insecta_odb9) (n= 1562). C: Complete, S: Single-copy, D: Duplicate, F: Fragmented, M: Missing. 162

Figure 4.2: A) *Nasonovia ribisnigri* transcript length distribution of 76,782 contigs of de novo assembled transcriptome. Individual contigs are ordered on X-axis based on increasing size. B) BUSCO analysis score of the *N. ribisnigri* transcriptome and the pea aphid (*Acyrtosiphon pisum*) transcriptome (obtained from BIPAA genomic resources) using an Arthropod gene set (n= 1562). C: Complete, S: Single-copy, D: Duplicate, F: Fragmented, M: Missing. 162

Figure 4.3: Gene expression change in *N. ribisnigri* biotypes feeding on resistant (Nr-gene) and susceptible (Nr-gene absent) lettuce host plants. The shrunken log2 fold change (removing the noise associated with

log2 fold changes from low count genes) of each gene between *N. ribisnigri* able to feed on resistant lettuce (Nr:1) and *N. ribisnigri* unable to feed on resistant lettuce (Nr:0). Fold changes were considered significant if P values were <0.05. Differentially expressed (DE) genes are coloured as blue circles and DE genes which fall out of the window are indicated by open triangles. The remaining grey circles are genes which are not significantly expressed between the resistant and susceptible biotype..... 164

Figure 4.4: Principal Component Analysis of the RNA-read count data (*featurecounts*) for each *N. ribisnigri* biotype grouped into three groups (Susceptible, Resistant_Kt (Kent_CL) and Resistant_UK (UK631)). Susceptible denotes all five *N. ribisnigri* cultures that are unable to feed on lettuce containing the *Nr-gene* (Nr:0)..... 166

Figure 4.5: Volcano plot of the 18,872 genes in both resistance-breaking (Nr:1) and susceptible (Nr:0) *N. ribisnigri* biotypes. 689 differentially expressed (DE) genes were identified with a P value of <0.05, highlighted in either blue or red. The volcano plot used the susceptible (Nr:0) *N. ribisnigri* biotype as a foundation for the plot and therefore DE genes are shown to be down regulated in the Nr:0 biotype and up regulated in the resistance-breaking (Nr:1) biotype. All up-regulated DE genes are highlighted in red and all down-regulated DE genes are highlighted in blue.. 166

Figure 4.6: Heatmap of the top 32 DE genes (P <0.00001) between *N. ribisnigri* resistant-breaking (Nr:1) and susceptible (Nr:0) biotypes, showing clear division. RES = resistant-breaking (Nr:1); SUS = susceptible (Nr:0). SUS1 = Nr4, SUS2 = Nr8, SUS3=Nr29, SUS4=4850a, SUS5=WT Kent, RES1=Kent_CL (Eluarde), RES2=Kent_CL (Pinokkio), RES3=UK631 (Eluarde), RES4=UK631 (Pinokkio). Gene identification (ID) is indicated on the right-hand side, along with functional information. 167

Figure 4.7: (A) 3-way Venn diagram of the differentially expressed genes in three contrasts (Kent_CL/susceptible; UK621/susceptible; Kent_CL/UK631) identified by DeSeq2. Differentially expressed genes were considered significant and used in the diagram if P <0.05. (B) RNA-seq counts for gene 2698 between resistant-breaking (RES) and susceptible (SUS) biotypes..... 167

Figure 4.8: Ct values (±SE) obtained for the three different HKGs under two experimental conditions. Green circles represent susceptible *N. ribisnigri* feeding on susceptible lettuce, orange circles represent

resistant <i>N. ribisnigri</i> feeding on resistant lettuce. Three biological and three technical replicates were used to validate the HKGs.....	171
Figure 4.9: Comparison of gene-2698 expression levels using qRT-PCR between four conditions of <i>N. ribisnigri</i> cultures feeding on resistant (cv. Eluarde) and susceptible (cv. Pinokkio) lettuce cultivars. Both Kent_CL and UK631 are resistant-breaking biotypes able to feed on resistant lettuce cultivars. WT Kent is a susceptible biotype unable to feed on the resistant lettuce cultivars containing the Nr-gene. RR – resistant <i>N. ribisnigri</i> fed on resistant lettuce cultivar (cv. Eluarde), RS – resistant <i>N. ribisnigri</i> fed on susceptible lettuce cultivar (cv. Pinokkio), SS – susceptible <i>N. ribisnigri</i> fed on susceptible lettuce cultivar (cv. Pinokkio). In total, three biological and three technical replicates were used to compare gene-2698 expression level. *** - $P < 0.001$. Both <i>RPS9</i> and <i>RPS18</i> were used for the qPCR normalisation for the gene expression analysis of gene-2698..	171
Figure 5.1: The Rothamsted Insect Survey (RIS) suction trap sites throughout the UK that were used in the study.....	194
Figure 5.2: Jitter box plot of all DNA extractions using Qiagen Qlamp DNA Micro Kit on individual <i>N. ribisnigri</i> . Samples below 4ng/ul (indicated by red line on graph) were unable to be used in the study as quantity was below minimum threshold. DNA yields were varied between sites and by year and was particularly low in Starcross, with DNA quantity dropping after the first five years of being stored in the RIS archive (Appendix 5).....	203
Figure 5.3: Principal coordinate analysis of the microsatellite variation in the currant-lettuce aphid (<i>Nasonovia ribisnigri</i>). Loose groupings of the West population (Hereford, Preston, Starcross and Wellesbourne) and the East (Brooms Barn, Kirton, Rothamsted, Writtle, Wye and York) can be seen.	204
Figure 5.4: Genetic structure bar plot inferred using Bayesian analysis in Structure software with two clusters ($K = 2$). Each bar represents an individual, with the colour of the bar representing the likelihood of membership to either population cluster.	205
Figure 5.5: Population pairwise F_{ST} of <i>Nasonovia ribisnigri</i> between population sites between 2003-2020 showing the level of genetic differentiation. The scale bar on the x axis highlights the pairwise F_{ST} value,	

the darker the square, the higher the F_{ST} value. X indicates no significant F_{ST} value between two populations (significance set to $F_{ST} > 0.20$, $P = < 0.05$).. 211

Figure 5.6: Population pairwise F_{ST} of *Nasonovia ribisnigri* between population sites and separated between year groups showing the level of genetic differentiation; (A) 2003-2007; (B) 2008-2014; (C) 2015-2020. The scale bar on the x axis highlights the pairwise F_{ST} value and X indicates no significant F_{ST} value. The UK map to the right of each pairwise F_{ST} analysis indicates population site locations with the thicker dashed line highlighting the East/West population divide (genetic barrier) and the thinner dashed line highlighting the main disparities between East year groupings and the general divides in the East population.. 213

Figure 1.7: Population pairwise F_{ST} of *Nasonovia ribisnigri* between population sites between 2003-2020 showing the level of genetic differentiation between the East (A) and the West (B). 03_07 = 2003-2007; 08_14 = 2008-2014; 15_20 = 2015-2020. The scale bar on the x axis highlights the pairwise F_{ST} value, the darker the square, the higher the F_{ST} value. X indicates no significant F_{ST} value between two populations (significance set to $F_{ST} > 0.20$, $P = < 0.05$).....213

List of tables

Table 2.1: Forecasting model summary statistics: No. of sites = number of sites included in the model; Day cut-off = number of Julian days of *N. ribisnigri* first flight in a given year (model discards any 'first flight' data after this date; T value = T value of the intercept between Julian day and degree day. *Silwood park was removed from model 6-8 and replaced with Kirton. 68

Table 3.1: Total number of *N. ribisnigri* identified on chicory (*C. intybus*) alternative host plant plots. 123

Table 3.2: Blastn results from NCBI nucleotide collection database. Sequence length is unclipped length from Sanger sequencing results. No samples out of the 12 sequenced contained any unsure nucleotides (N). ID: ITS1 and 2 *L. sativa* has been abbreviated from Figure 2. ALB – alate body only; AB – apterous body only; ALH – alate head/stylet only; AH – apterous head/stylet only. 126

Table 3.3: Blastn results from NCBI nucleotide collection database. Sequence length is the unclipped length from the Sanger sequencing results. None of 8 sequenced samples contained any unsure nucleotides (N).....	127
Table 4.1: Near isogenic culture lines of <i>Nasonovia ribisnigri</i> which were used for DNA extraction and <i>de novo</i> genome assembly. Susceptible = not resistant to the Nr-gene in lettuce varieties (Nr:0); resistant = resistant to the Nr-gene in lettuce varieties (Nr:1). Isogenic lines are indicated numerically i.e. (1).....	148
Table 4.2: <i>Nasonovia ribisnigri</i> cultures that were used for the RNA-seq data extraction and sequencing. These data were used for the transcriptome assembly and RNA-seq analysis for differentially expressed (DE) gene analysis. Nr: 0 = <i>Nr-gene</i> susceptible; Nr: 1 = <i>Nr-gene</i> susceptible; Pyrethroid R = pyrethroid resistant; Insecticide R = insecticide resistant; (S) = susceptible (does not contain <i>Nr-gene</i>); (R) = resistant (contains <i>Nr-gene</i>). For the differentially expressed (DE) gene analysis using DeSeq2, RNA-seq data was grouped into three groups: blue (susceptible), orange (resistant_Kent) and green (resistant_UK).....	153
Table 4.3: Conditions used for qRT-PCR experiment for <i>N. ribisnigri</i> . Nr:0 = susceptible biotype (unable to feed on resistant lettuce containing the Nr-gene. Nr:1 = resistant biotype (able to feed and reproduce on resistant lettuce containing the Nr-gene). UK631 were cultured on both susceptible (Pinokkio) and resistant (Eluarde) lettuce lines. Susceptible <i>N. ribisnigri</i> cultures were only cultured on susceptible (Pinokkio) lettuce lines.....	155
Table 4.4: Primer sequences and amplicon characteristics of resistance genes and house-keeping genes for qRT-PCR validation experiment for <i>N. ribisnigri</i>	158
Table 4.5: Summary statistics of the currant-lettuce aphid, <i>N. ribisnigri</i> genomes. Nr8 and WT_Kent are susceptible (Nr:0) biotypes and UK631, Kent CL and Ely are resistant (Nr:1) biotypes. Only the best genome assemblies for the susceptible (Nr:0) and resistant (Nr:1) were selected and used for downstream annotation and analysis, highlighted in red.	159
Table 4.6: HISAT2 results of number of reads which aligned to the Nr_8 WT genome annotation using feature counts and mapping statistics. Nr:0 = <i>N. ribisnigri</i> unable to feed to lettuce containing the Nr-gene (susceptible); Nr:1 <i>N. ribisnigri</i> able to break the host plant resistance and feed on lettuce containing the Nr-gene (resistant).	163

Table 4.7: Differentially expressed (DE) genes in resistance-breaking biotypes of <i>N. ribisnigri</i> (Nr:1) feeding on both resistant (nr-gene present) and susceptible (nr-gene absent) lettuce host plants. Both up-regulated and down-regulated DE genes are shown for the resistance-breaking (Nr:1) biotype. Gene names are derived from contig number (e.g. 2698) and transcript ID (0.8). Both up-regulated and down-regulated genes are ordered by <i>P</i> -value significance identified from DeSeq2 using LFC estimates. Fold change values were considered significant if $P < 0.0001$	169
Table 5.1: Msatcommander search parameters for microsatellite discovery and the number of microsatellites discovered and passing initial Primer3 parameters.	195
Table 5.2: Experimental design of the <i>N. ribisnigri</i> cultures that were used for microsatellite validation. Nr: 0 = <i>Nr-gene</i> susceptible; Nr: 1 = <i>Nr-gene</i> susceptible; Pyrethroid R = pyrethroid resistant; Insecticide R = insecticide resistant; (S) = susceptible (does not contain <i>Nr-gene</i>); (R) = resistant (contains <i>Nr-gene</i>).	196
Table 5.3: The fluorescent dye parameters used for microsatellite analysis.....	197
Table 5.4: The 15 identified and developed microsatellites for <i>Nasonovia ribisnigri</i> and their F/R primers sets. Highlighted in bold are the eight microsatellites used for this study. Ta(°C): Annealing temperature; N _A : Number of alleles.	199
Table 5.5: Number of <i>Nasonovia ribisnigri</i> individuals collected from the Rothamsted Insect Survey (RIS) archive and used in the current population genetics study. York suction trap site was not established pre 2008 and therefore no samples were collected.	200
Table 5.6: Genetic diversity estimates for the 10 <i>N. ribisnigri</i> sites; combined sites; and East and West populations between 2003-2020. <i>N</i> – number of gene copies (number of individuals x2), <i>H</i> – number of haplotypes, <i>H_O</i> – mean observed heterozygosity for all eight microsatellite loci, <i>H_E</i> – mean expected heterozygosity for all eight microsatellite loci, <i>H_O/H_E</i> – percent of loci which had significantly different <i>H_O</i> and <i>H_E</i> over all eight microsatellite loci, <i>F_{IS}</i> – population specific inbreeding coefficient, and <i>pHWE</i> – number of loci for each population which departed from Hardy-Weinberg equilibrium (HWE).	208
Table 5.7: Hierarchical analysis of molecular variance (AMOVA) for <i>Nasonovia ribisnigri</i> . (A) Genetic variation of all individuals between 9 sites among year groupings (2003-2007; 2008-2014; 2015-2020); (B) East geographical cluster determined by Structure comprising of Brooms Barn, Kirton, Rothamsted, Writtle	

and Wye among year groups ; (C) West geographical cluster determined by Structure comprising of Hereford, Preston, Starcross and Wellesbourne among year groups. York was excluded from temporal analysis as the suction trap was installed in 2008 and therefore no data was collected between 2003-2007.

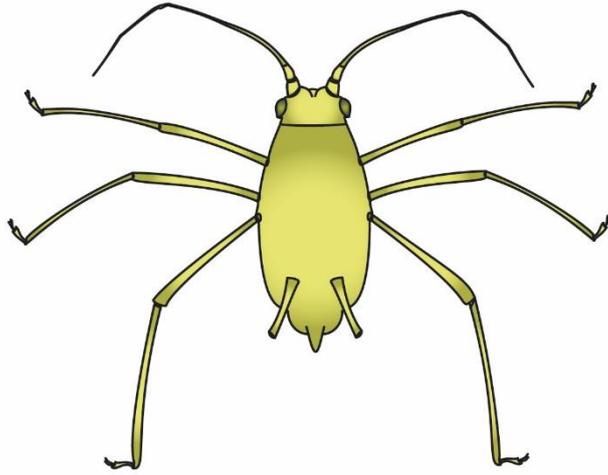
..... 209

Table 5.8: Pairwise F_{ST} values of all individuals collected in different year groups (A), between individuals in the East (B) and individuals in the West (C). N shows the number of individuals in each group. 210

Table 5.9: Hierarchical analysis of molecular variance (AMOVA) for *Nasonovia ribisnigri*. Genetic variation of all individuals between the East and West grouping, as identified in Structure. East geographical cluster comprises of Brooms Barn, Kirton, Rothamsted, Writtle and Wye among all year groups. West geographical cluster comprises of Hereford, Preston, Starcross and Wellesbourne among all year groups.

d.f = degrees of freedom..... 214

Chapter 1



CHAPTER 1: Introduction

Aphids (Hemiptera: Sternorrhyncha) are small phytophagous insects which reside within three extant families within the Aphidoidea. The group comprises over 5000 species throughout the world, although 80% of species live in temperate regions (Blackman & Eastop, 2000; Favret, 2014). Aphids feed on sap from phloem or xylem vessels using their stylet and, have either a single (monophagous) or multiple (polyphagous) host plant range. Their size can range from ~0.75 mm to up to 7.8 mm (Bell *et al.*, 2012), with the largest known aphid to be the giant bark aphid (*Longistigma caryae*) (Harris, 1841) (Denmark, 2003), along with the giant willow aphid (*Tuberolachnus salignus*) (Mordvilko, 1909) (Leather, 2014). Within the UK, there are currently an estimated 670 species/subspecies of aphids in the Aphididae, Adelgidae and Phylloxeridae (Baker *et al.*, 2015), although aphid taxonomy is yet to be resolved which may see this estimate rise.

1.1 Aphids as crop pests

The horticulturally important species of aphids primarily reside in the Aphidinae subfamily (Blackman & Eastop, 1984). Despite a relatively small global species diversity compared to many other insect groups, a small number of aphid species occupying both tropical and temperate regions severely compromise horticulture (feeding on over 100 horticulturally and economically important crops (Dixon *et al.*, 1987; Martin, 1983)). Aphids affect their host plants in a variety of ways. Firstly, they are phloem feeders and penetrate the host plant with their stylet to extract nutrients, diverting resources away from the plant. This in turn, can stunt plant growth and make the plant more susceptible to abiotic stresses (e.g. drought). Secondly, infected aphids can transmit a plethora of plant viruses, with nearly half of all plant viruses (275 out of 600) transmitted by aphids (Katis *et al.*, 2007). Thirdly, because of aphids excreting honeydew on the host plant epidermis, sooty moulds grow on the surface impeding photosynthetic activities (Annan *et al.*, 1994; Hughes, 1976).

It is difficult to accurately estimate the level of damage and economic losses caused by aphids in any one year because of the large interannual variation in both aphid numbers and phenology as well as variation in the type and diversity of crops and horticultural conditions (Dedryver *et al.*, 2010). For example, milder winters are known to promote aphid development and facilitate earlier migration from their primary host to the potential crop (secondary host) (Bejer-Peterson, 1962; Bell *et al.*, 2015; Saldaña *et al.*, 2007) whereas; cold winters significantly affect aphid mortality (Carter, 1972), especially when feeding aphids are exposed to sub-zero temperatures (Powell, 1974). The

introduction of the Russian wheat aphid, *Diuraphis noxia*, in the USA during 1986 and subsequent rapid spread throughout wheat and barley production areas resulted in direct annual yield losses of \$274 million in 1988, highlighting the potential economic impact of horticulturally important aphids (Michaud and Sloderbeck, 2005). Substantial economic costs are associated with aphids, with estimated economical costs of \$241 and \$482 million per year in Australia alone as a result of direct feeding and virus damage, respectively (Valenzuela and Hoffmann, 2015).

The success of aphids in both horticulture and the environment are a result of their ability to exploit and quickly adapt to a wide range of conditions (Williams and Dixon, 2007). This high level of plasticity enables aphids to continuously adapt to the changing environment faster than many other studied insect groups and predicting how they respond to environmental change has proved challenging (Kindlmann *et al.*, 2007). These phytophagous insects can have a large diet breadth, host range and are renowned for their rapid parthenogenic reproduction (Blackman and Eastop, 1984). With unlimited resources, it has been speculated that the world would be 149 km deep in aphids (Harrington *et al.*, 1994). Fortunately, resources are limited, mainly by host plant abundance but also predators, parasitoids and entomopathogenic fungi (Goggin, 2007; Hagen and Van Den Bosch, 1968). The annual aphid life cycle can be holocyclic with two reproductive phases, parthenogenetic in spring-summer and sexual in autumn (Dixon and Kindlmann, 1998). Some populations and species can be anoholocyclic, with aphids reproducing parthenogenetically all year round. The latter produces live young (nymphs) which are already bearing the next generation of clones, in turn enabling very rapid increase in populations and subsequent colonisation of the host plant (Figueroa *et al.*, 2018). Many aphid pests are host alternating (heteroecious), in which they have a winter (primary) host and in the spring migrate onto a summer (secondary) host, the latter of which includes many crops. However, a minority of species do not host alternate and remain on the secondary host as a facultative anholocyclic clone (asexual reproduction) or, never maintain an alternative host phenotype in the UK (Bell *et al.*, 2012). In overcrowded conditions, decrease in host plant quality or a response to a predator/parasitoid attack, asexual aphids can adopt epigenetic switching (Srinivasan and Brisson, 2012) which enables the production of winged (alate) forms that can migrate to a new host plant (Loxdale and Balog, 2018; Sentis *et al.*, 2018). This enables aphids to rapidly colonise multiple hosts plants and further increases their dispersal to uncolonised crop (Simon and Peccoud, 2018).

1.1.1 Aphid lettuce pests

Aphids are major pests of outdoor lettuce (*Lactuca sativa* L). Four species are economically relevant, and these include: currant-lettuce aphid (*Nasonovia ribisnigri*) (Mosley, 1841), lettuce-root aphid (*Pemphigus bursarius*) (Linnaeus, 1758), potato aphid (*Macrosiphum euphorbiae*) (Thomas, 1878) and the peach-potato aphid (*Myzus persicae*) (Sulzer, 1776). Of these, the currant-lettuce aphid, *N. ribisnigri* (Hemiptera: Aphididae), is the most damaging species of aphid on outdoor lettuce and has been reported to occur in very high numbers during mid-summer and early autumn (Liu, 2004; Diaz *et al.*, 2012). A sporadic outbreak of *N. ribisnigri* between July-August 2016 on 60 ha outdoor organic lettuce on a farm in Cambridgeshire resulted in unharvestable crop for 6 weeks, amounting to a loss of ~£250,000 (Garfield, Gs Fresh, unpublished). With ~6,000 ha of land used to cultivate lettuce in the UK; outbreaks of this magnitude would have dramatic effects on the leafy salad industry, damaging not only the crop but more widely affecting the supply chain, impact supplier's reputation and customer satisfaction.

In Europe, *N. ribisnigri* contributes to significant losses to outdoor lettuce in Spain, France, Germany, Switzerland and The Netherlands) (Reinink and Dieleman, 1993). In Switzerland, *N. ribisnigri* has caused 70% yield loss (Sauer-Kesper *et al.*, 2011). During the winter months, lettuce cannot be cultivated outdoors and is limited to indoor growing facilities. However, the demand for fresh salads during these months far surpasses the UK's capability to produce it and therefore most of the lettuce is sourced overseas, particularly from Spanish growers. In Spain, *N. ribisnigri* is regarded as the main aphid species to colonise outdoor lettuce crop (Nebreda *et al.*, 2004). This problem is not just only within Europe and the UK. *Nasonovia ribisnigri* has been recorded as the most damaging species on outdoor lettuce in North America (Palumbo and Hannan, 2002) and in New Zealand (Stufkens and Teulon, 2003). The UK imported 215 thousand tonnes of lettuce in 2019, with a net value of £156Mil (DEFRA, Horticulture Statistics, 2020).

1.2 Taxonomy of the currant-lettuce aphid, *Nasonovia ribisnigri*

Characteristic morphological characteristics of most, but not all aphids, are the cauda, siphunculi, abdominal markings, terminal process, leg colouration, head, and number of rhinaria (sensory pits) found on the antennal segments (Fig. 1.1). Aphids are hemimetabolous in which they undergo incomplete metamorphosis and have no pupal stage (Blackman and Eastop, 1984). There are typically three forms of aphids; nymphs, apterous (wingless adults) and alate (winged adults) however, some aphids do not produce alates. Taxonomically, nymphs, apterous and alates vary dramatically (Fig. 1.1). These diagnostic features vary dramatically between species and in some cases (such as siphunculi) can be greatly reduced or completely absent. First described in 1841 (Mosley), *Nasonovia ribisnigri*, is a small soft-bodied insect in the Order Hemiptera and is classified within the Family Aphididae.

Scientific classification of *Nasonovia ribisnigri*

- **Kingdom:** Animalia
- **Phylum:** Arthropoda
- **Class:** Insecta
- **Order:** Hemiptera
- **Suborder:** Sternorrhyncha
- **Infraorder:** Aphidomorpha
- **Superfamily:** Aphidoidea
- **Family:** Aphididae
- **Genus:** *Nasonovia*
- **Species:** *ribisnigri*

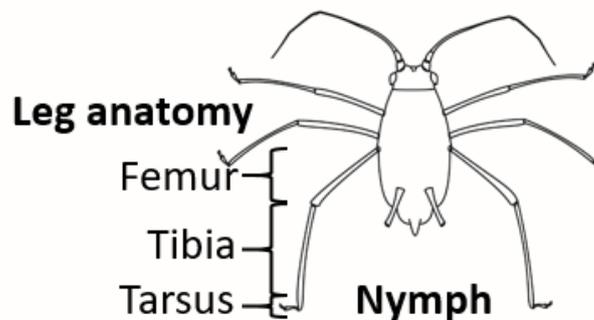
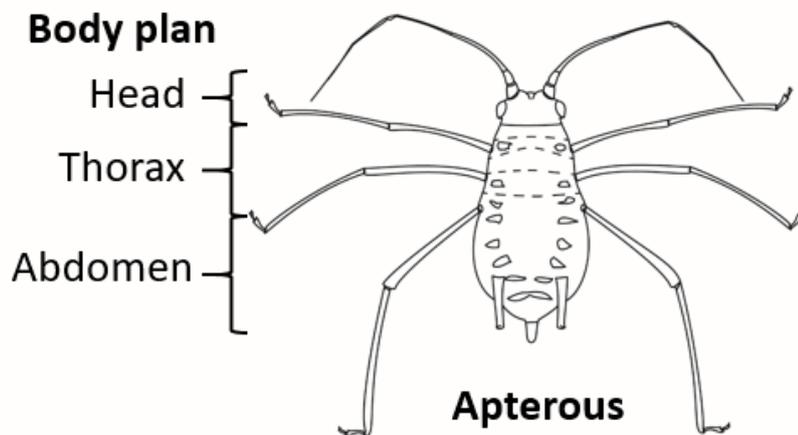
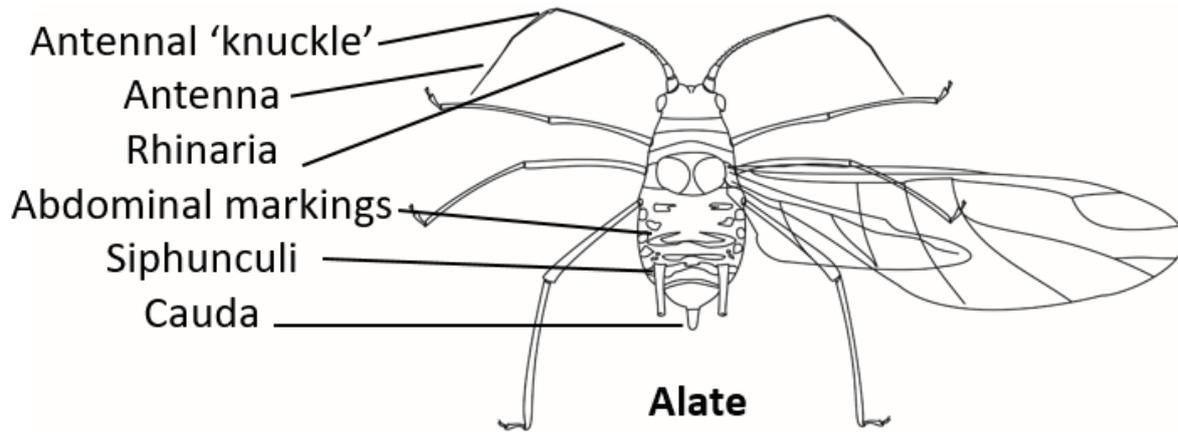


Figure 1.1: Morphology of *Nasonovia ribisnigri* highlighting the main characteristics of a winged adult (alate), non-winged adult (apterous) and a nymph. Figure produced using Adobe Illustrator.

The adults of *N. ribisnigri* exist in two forms, winged (alate) or wingless (apterous). Depending on the host plant and time of year, the colours vary between shades of yellow and green to red. Adult alate and apterous size can range between 1.5-2.5 mm to 1.3-2.7 mm, respectively (Blackman and Eastop, 1984). Figure 1.2 illustrates the main taxonomic characteristics of *N. ribisnigri* for identifying both the alate and apterous forms. Both apterous and alate forms possess a finger-like cauda which is completely dark in alates. The siphunculi of both forms are straight and have dark tips in the apterous

but are uniformly dark in the alates. The abdominal markings are particularly conspicuous in the alate, possessing two characteristic chevrons in the middle of the abdomen (Fig. 1.2B) whereas; the apterous markings consist more of faint flecks of dark which can be difficult to see (Fig. 1.2C). In alates, they possess between 23-66 secondary rhinaria on the third antennal segment, 2-14 on the fourth segment and none on the fifth segment. The antenna of the alates is long and have a large 'knuckle' on the fourth antennal segment prior to the fifth antennal segment. They also possess darkened femurs around the femur/tibia joint.

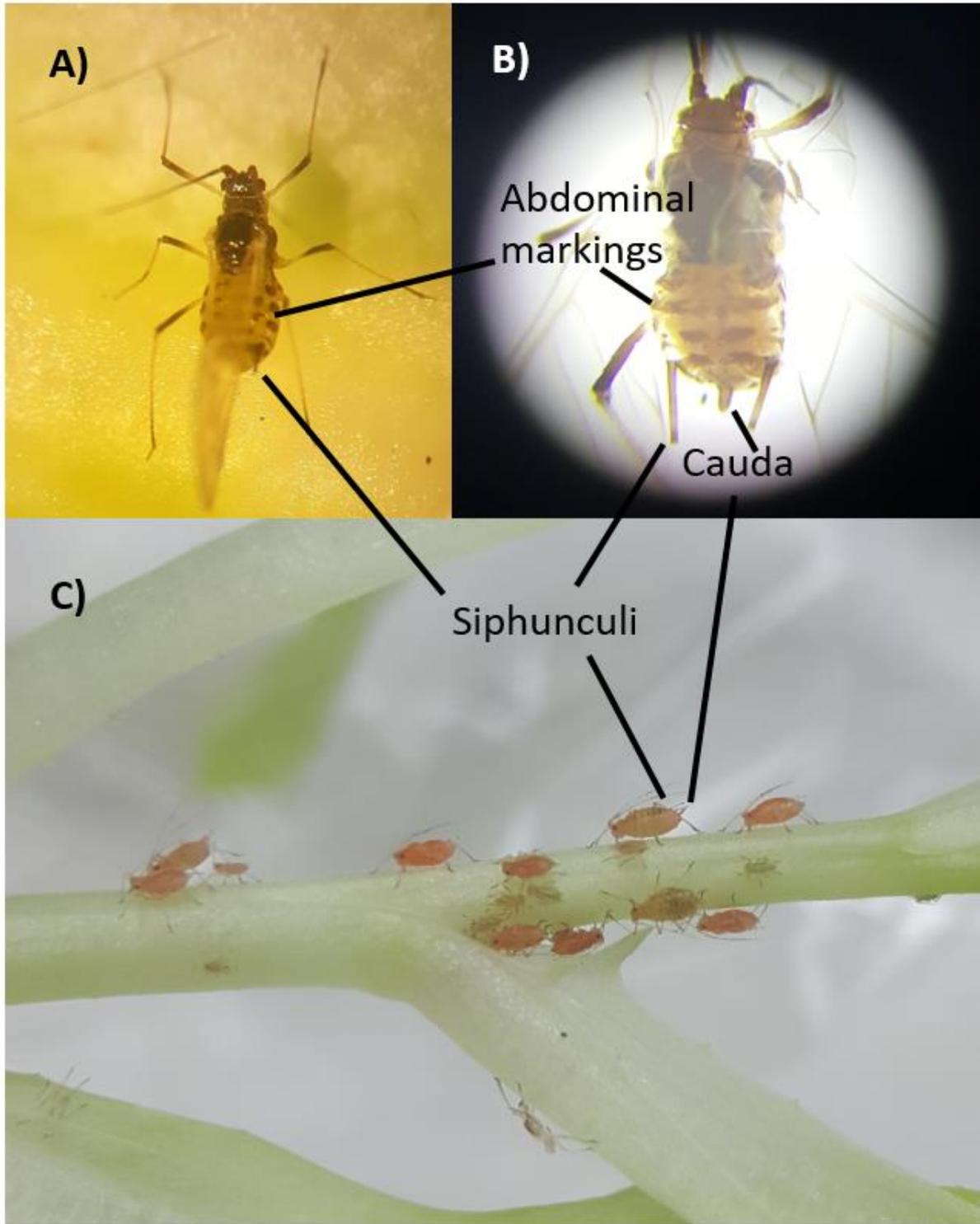


Figure 1.2: The current-lettuce aphid, *Nasonovia ribisnigri*, highlighting some of the characteristic features used for taxonomic identification. A) Alate *N. ribisnigri* feeding on lettuce, B) Alate *N. ribisnigri* under x 20 magnification using a stereomicroscope, with finger-like cauda, siphunculi and distinct abdominal markings and, C) Apterous *N. ribisnigri* both adults and nymphs with mixture of yellowish green and red morphs.

1.3 Distribution

Though it is most common in Europe, *N. ribisnigri* has been identified in North America, South America, and Oceania (CABI/EPPO, 2003). Since 2003, *N. ribisnigri* has been reported in Central Asia and the Middle East (Blackman and Eastop, 1984). Additionally, *N. ribisnigri* have further been reported in New Zealand and Tasmania (Dominiak *et al.*, 2009; Stufkens and Teulon, 2001).

1.4 Virus transmission in lettuce and associated hosts

Aphids are well known for their ability to transmit a wide variety of non- semi- and persistent plant virus types (De Bokx and Piron, 1990; Webster *et al.*, 2018; Fereres and Moreno, 2009). Transmission is effective when the stylet has penetrated beyond the epidermis, through the intercellular spaces and into the phloem, which may only take a matter of minutes (Fig. 1.3). Virus transmission on lettuce by *N. ribisnigri* is generally regarded as insignificant due to their inability to transmit many pathogens, including the economically damaging lettuce mosaic virus (LMV) (Nebreda *et al.* 2004; Moreno *et al.* 2007). However, it has been demonstrated that *N. ribisnigri* can transmit lettuce necrotic leaf curl virus (LNLVC), for which it is thought that the stylets are the potential retention sites (Verbeek *et al.* 2017). The other two main aphids that frequently feed on outdoor lettuce are the peach-potato aphid (*M. persicae*) and the pea aphid (*Acyrtosiphon pisum*) (Harris, 1776) (Reinink *et al.*, 1989). These aphids have a much larger host range than *N. ribisnigri* and, as a result, transmit significantly more viruses (Srinivasan and Alvarez, 2007). The peach-potato aphid (*M. persicae*) is known to be a highly polyphagous cosmopolitan species, with over 50 different known families that can use as a secondary host (summer host) and is a known vector of over 78 different plant viruses (van den Heuvel *et al.*, 1994). In contrast, *N. ribisnigri* has been shown to only accept secondary hosts in the families Asteraceae (daisy family, which include lettuce), Brassicaceae, Scrophulariaceae (figwort) and Solanaceae (nightshade/potato family), and to be associated with only a few plant viruses (Blackman and Eastop, 1984). On their winter host, which predominantly consists of currant species (*Ribes* spp.), *N. ribisnigri* can transmit gooseberry vein-banding virus (Adams and Thresh, 1987; Blackman and Eastop, 1984). Therefore, the low economic significance of virus transmission by *N. ribisnigri* is likely a result of their alternative host plants not acting as a reservoir for viruses that can be vectored to lettuce.

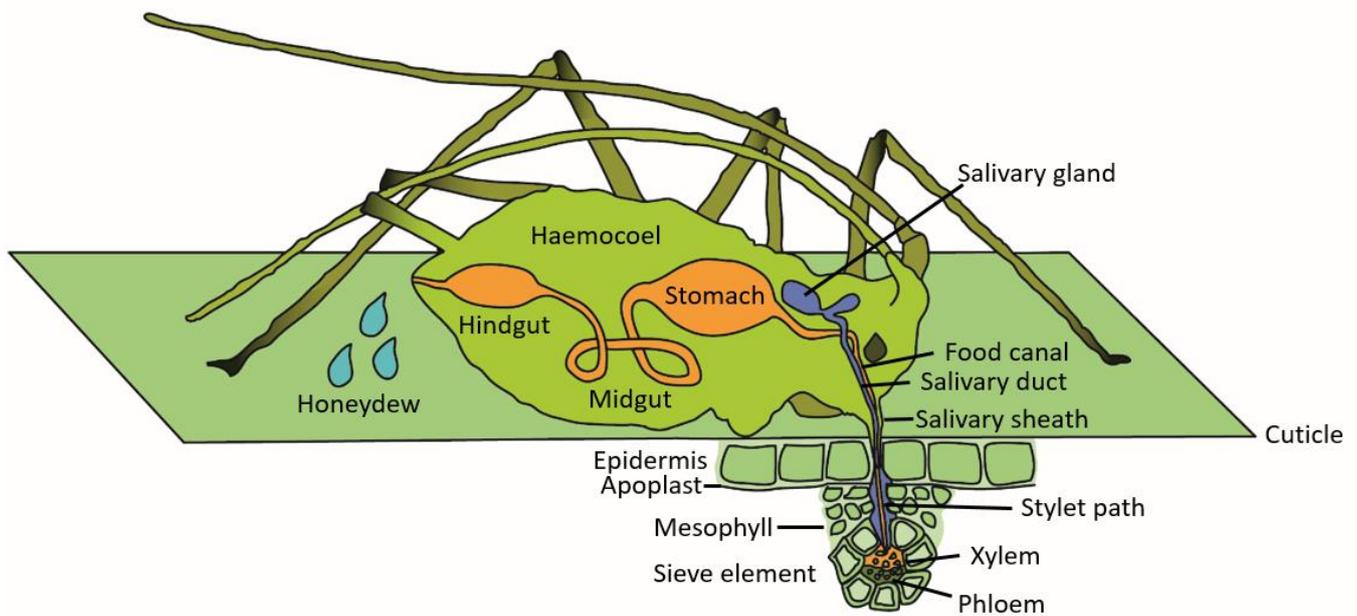


Figure 1.3: *Nasonovia ribisnigri* feeding on a host plant (primary or secondary). Unlike chewing insects (e.g. caterpillars, grasshoppers, beetles) which cause large tissue damage, aphids inflict only slight physical damage. Using their long flexible stylet, they can navigate through the host plant cuticle, epidermis, and mesophyll to reach the phloem is which they feed. Figure produced using Adobe Illustrator.

1.5 *Nasonovia ribisnigri* Biology and lifecycle

The currant-lettuce aphid (*N. ribisnigri*) is a heteroecious (host-alternating) aphid, which is typical for an aphid species that is of horticultural importance (Blackman and Eastop, 1984). *Nasonovia ribisnigri* can either undergo holocyclic (sexual) or anholocyclic (asexual) reproduction (Nebreda *et al.*, 2005). A holocyclic lifecycle involves males and females being produced in autumn, with eggs being laid winter host plant (Blackman and Eastop, 1984). With an anholocyclic (asexual) lifecycle no males are produced, and aphids remain parthenogenic all year and over winter as either nymphs or as adults. There is a final reproduction mode, known as androcycl (Blackman, 1971). In this scenario, only sexual males are produced in the autumn with no mating females. Both anholocyclic and androcyclic aphids have adaption to warmer climates due to the parthenogenic trait that maintains overwintering in an active phase rather than an egg. This trait can be passed through the adrocyclic lifecycle via males. Androcycl has been documented in other horticulturally important aphids (Delmotte *et al.*, 2001; Halkett *et al.*, 2008) but there currently is no information on androcycl in *N. ribisnigri*. However, it is possible that androcycl could occur in areas where both holocyclic and anholocyclic populations exist.

1.5.1 Winter (primary) hosts

The main primary winter host for *N. ribisnigri* are members of the *Ribes* genus which consist of gooseberry and currants (van der Arend, 2003). In the UK midlands, it has been demonstrated that *N. ribisnigri* are able to survive the winter as nymphs and adults on corn speedwell (*Veronica arvensis*) (Hough, 2013). Additionally, other host plants were shown to be suitable alternative winter hosts which include chicory (*Chichorium intybus*), smooth hawks-beard (*Crepis capillaris*), nipplewort (*Lapsana communis*), orange hawkweed (*Hieracium aurantiacum*), mouse-ear hawkweed (*Hieracium pilosella*), spiked speedwell (*Veronica spicata*) and heath speedwell (*Veronica officinalis*) (Hough, 2013). However, these finds were only demonstrated under laboratory conditions and no field observations or experiments have confirmed this to date. Individuals predominantly overwinter as eggs, yet there is anecdotal evidence to suggest they can survive and overwinter as apterae and nymphs in areas with milder winter climates, such as Northern France and Spain. Fundatrix aphids will hatch from eggs in early spring and produce multiple generations parthenogenetically on the primary host (Fig. 1.4). These aphids preferentially feed on the young shoots of the primary host, which can cause leaves to curl and stunt growth and development. *Nasonovia ribisnigri* persists on the primary host until late spring - early summer, depending on local climatic conditions, particularly temperature (cumulative day degrees). Once this threshold or certain cues (e.g., changes in primary host plant quality) are reached, winged alates are produced which migrate from the primary host to a secondary summer host.

1.5.2 Summer (secondary) hosts

For the secondary host, lettuce is the main host plant *N. ribisnigri* associated with this aphid, due to lettuce being a horticulturally important crop. Like the primary host, *N. ribisnigri* preferentially feeds on the young developing shoots and leaves of the secondary host. Successive parthenogenic apterous (wingless) generations of *N. ribisnigri* are produced on the secondary host, including alates which can migrate and colonise other suitable secondary hosts.

Typically, lettuce is not available for the entirety of *N. ribisnigri* secondary host life cycle (April-November in the UK) and only grown in large quantities in certain areas in the UK during this time (predominantly in the east). The four families that *N. ribisnigri* is known to use as a secondary host is Asteraceae (daisy family, which include lettuce), Brassicaceae, Scrophulariaceae (figwort) and Solanaceae (nightshade/potato family) (Diaz and Fereres, 2005). In laboratory experiments, it was

reported that *N. ribisnigri* accepts and can reproduce on mouse-ear hawkweed (*Pilosella officinarum*), Orange hawkweed (*Pilosella aurantiaca*), nipplewort (*Lapsana communis*), smooth hawkbeard (*Crepis capillaris*), speedwell (*Veronica spp.*) and chicory (*Cichorium intybus*) (Hough, 2013). Although it was shown that chicory (*C. intybus*), smooth hawkbeard (*C. capillaris*) and orange hawkweed (*P. aurantiaca*) were more viable regarding the reproduction of both alate and apterous *N. ribisnigri* (Hough, 2013).

In the UK, the typical migration of *N. ribisnigri* to a secondary host (including lettuce) is between April and October (Fig. 1.4). Commercial outdoor lettuce is predominantly grown during a similar time frame and therefore is available for the majority of *N. ribisnigri* secondary host lifecycle. Interestingly, *N. ribisnigri* usually do not become a problem on outdoor lettuce until late summer (August onwards) and is regarded as the main risk period. The earlier months from the initial migration from their winter host to the secondary host is most likely to occur to a non-lettuce alternative host. This is despite areas containing 'suitable' outdoor lettuce hosts, particularly in the east of the UK, where most of the outdoor lettuce is cultivated. There is an evident gap in the literature in understanding why *N. ribisnigri* preferentially feed on non-lettuce host alternatives at the start of lettuce season and subsequently move onto a lettuce host later in the year. These non-lettuce secondary hosts could be acting as a green bridge between the non-lettuce host and lettuce, but further research is required.

Nasonovia ribisnigri have an unusual feeding niche on a host plant whereby they preferentially feed on the young developing shoots and leaves of plants which, in lettuce, is on the heart leaves towards to centre of the plant. The other two main pests of lettuce, peach-potato aphid (*M. persicae*) and the potato aphid (*M. euphorbiae*) preferentially feed on the older outer leaves and are seldom seen feeding in the centre of lettuce (Alvarez *et al.*, 2012; Reuter *et al.*, 1993; Smith and Chuang, 2014). As a result of this feeding behavioural niche, *N. ribisnigri* does not directly compete for space on a lettuce plant with other lettuce-feeding aphids. Since *N. ribisnigri* is a foliage-feeding aphid, it can cause stunted growth in lettuce (Gratwick, 1992). This affects current harvesting practices, due to the need for uniformity amongst lettuce crop when harvesting. The main reason that *N. ribisnigri* is such a huge problem on outdoor lettuce crop is due to their preferential feeding behaviour in the centre of the lettuce (Parker *et al.*, 2002). This makes in-house quality control detection harder and the presence of low numbers within the lettuce head can render the lettuce unmarketable.

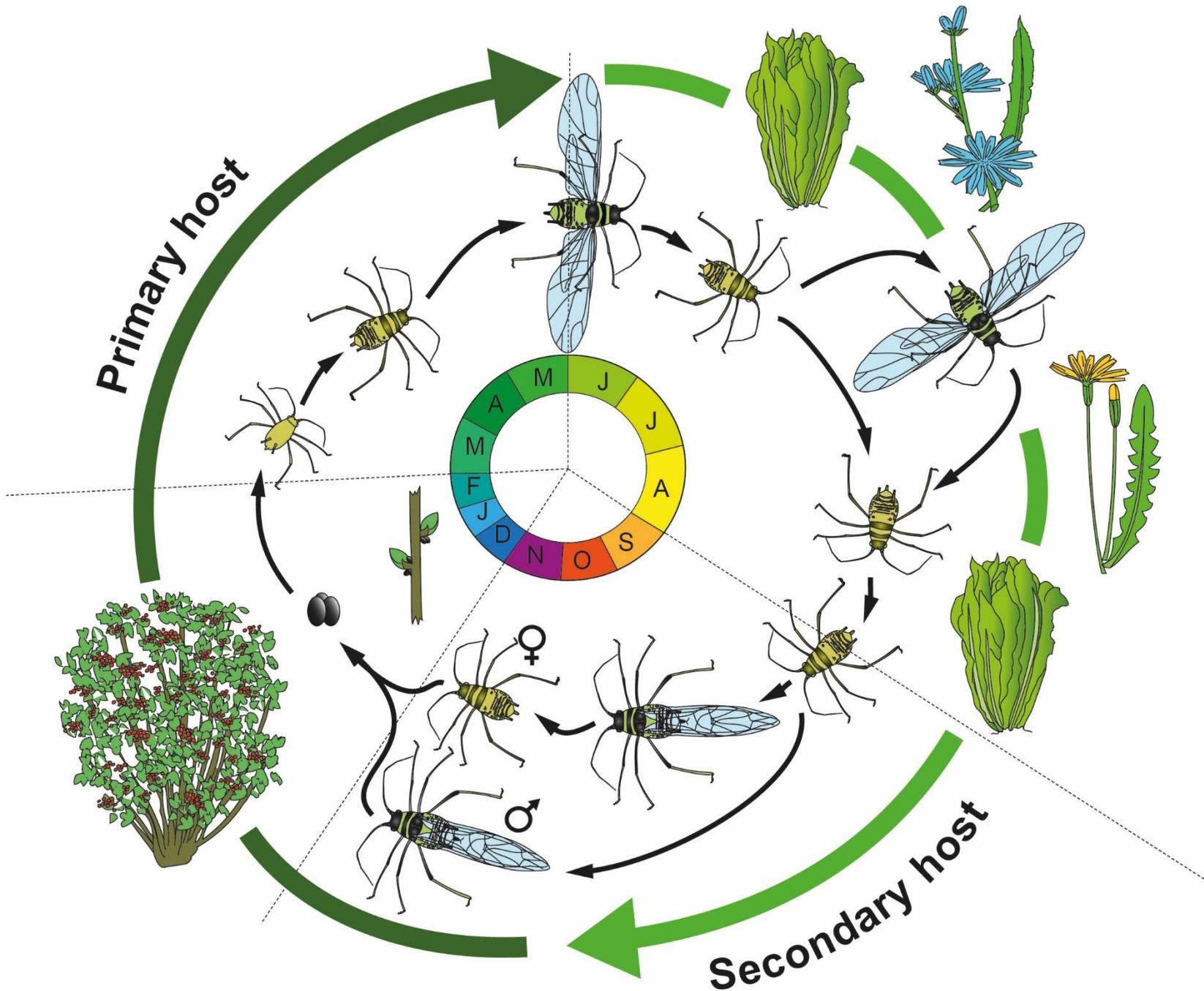


Figure 1.4: The holocyclic (sexual) lifecycle of the currant-lettuce aphid, *Nasonovia ribisnigri*. Primary host is various members of the *Ribes* spp. Genus (e.g., Blackcurrant (*R. nigrum*) and gooseberry (*R. uva-crispa*)) and secondary host are members of the Asteraceae Family (e.g. Chicory (*Cichorium intybus*) and lettuce (*Lactuca sativa*)). Figure produced using Adobe Illustrator.

1.5.3 Lettuce host biology

First described in 1753 by Carl Linnaeus, lettuce, *Lactuca sativa*, is a member of the Asteraceae (daisy) family (Jarvis, 1992). The first recorded case of lettuce cultivation was in Ancient Egypt over 6,000 years ago, in which a tomb painting depicts a long-leafed form of lettuce being cultivated (Lindqvist, 1960). The origin of lettuce is based on three possible theories that it; 1. originated from wild forms of *L. sativa*, 2. originated through direct descent from *L. serriola* (wild lettuce relative) and, 3. originated through hybridisation between different species (Thompson *et al.*, 1941; Lindqvist, 1960). To date, it is still unclear which species were directly involved with the domestication of lettuce but it is clear that *L. serriola* is one or the only direct ancestor (De Vries, 1997; Kesseli *et al.*, 1991).

Lettuce has a wide range of cultivars of different sizes, shapes, colours, and textures. As a result, it is a very versatile crop and used in a variety of dishes worldwide. There are seven lettuce cultivar groups which include Butterhead, Cos, Latin, Crisphead, Cutting, Stalk lettuce and Oilseed lettuce (Rodenburg, 1960; Rulkens, 1987). Typically, lettuce has a spread and height of between 15-30 cm and can have open heads with frilly or ruffled leaves (e.g. Romaine, Cos) and closed heads (e.g. Crisphead, Iceberg), with a large range in between. The lettuce plant root system consists of a main taproot and smaller secondary roots. Varieties can have different climatic tolerances, depending on the location they are cultivated, and enables growers to select plants that are most suited to the environment they will be grown (Michelon *et al.*, 2020; Koudela and Petříková, 2008; Wheeler *et al.*, 1993).

Lettuce is produced in 106 countries worldwide, with China, India and USA sharing the majority of the gross yield (75%), with more than half of the global average residing with China (51.6%) (Shatilov *et al.*, 2019). In the UK, outdoor lettuce crop marketed each year amounts to an average of £148M and is much higher than most other field vegetables grown in the UK (DEFRA, Horticulture Statistics, 2020). The UK is the fourth largest importer of lettuce, with Canada, Germany and USA taking the top three spots with the UK averaging 207M tonnes a year and valued at ~£164M (FAO, 2017). In the UK, most of the lettuce is produced in the eastern region, in which the majority of UK horticulture is located (DEFRA, 2021a).

Lettuce in the UK is only grown outdoors during the spring and summer months, when the weather is favourable for lettuce cultivation. Lettuce has a short growing period in a field, usually between 4-6 weeks during a UK summer. During the winter months, lettuce cannot be cultivated outdoors and is limited to indoor growing facilities. However, the demand for fresh salads during these

months far surpasses the UK's capability to produce it and therefore the majority of lettuce is sourced overseas, particularly from Spanish growers. The UK imported 215 thousand tonnes of lettuce in 2019, with a net value of £156M (DEFRA, Horticulture Statistics, 2020).

1.6 Organic and conventional practices

Lettuce requires a large area for production and, with the additional inherent running costs for growing indoors, makes this method of cultivation economically unviable. Lettuce can be grown under cover, such as using insect-proof meshing and polytunnels, but this again increases the associated growing costs. Indeed, the use of insect mesh and polytunnels enables lettuce crop to be cultivated successfully outdoors with reduced or zero (if organic) chemical input but growing times can be hindered as a result. In addition, insect meshing can create an unfavourable or suboptimal growing condition, particularly in high temperatures, creating a microclimate that can stunt growth and potentially encourage fungal diseases, such as downy mildew, Septoria leaf spot and sclerotinia (Purdy, 1979; Raid, 2004; Young, 2004). As a result, these methods of cultivations are usually restricted to an organic method of farming and therefore cost per lettuce head usually comes at a premium to offset the higher farming input required.

Most of the lettuce grown both in the UK and Europe is grown outdoors, with the majority being cultivated through conventional farming methods (DEFRA, 2021a). Typically, this involves synthetic inputs such as fertilisers and pesticides, to facilitate healthy uniform growth and to control pests such as insects and pathogens on crop, respectively. A two-year study highlighted that a conventional growing method produced an average 20% yield increase but there was no difference in overall food nutritional value (Polat *et al.*, 2010). These additional synthetic inputs can have higher detrimental environmental effects than an organic approach. Considering eight environmental impacts which included human toxicity, terrestrial ecotoxicology and freshwater ecotoxicology, for example, Wongwai *et al.*, (2014) demonstrated that conventional farming of lettuce in Thailand had significantly greater environmental impacts than an organic approach.

It is evident that both conventional and organic lettuce farming practises have associated benefits and disadvantages. In relation to *N. ribisnigri*, organically grown lettuces in the UK generally require an insect mesh which covers the entire crop which acts as a physical barrier to prevent colonisation, especially during late summer when *N. ribisnigri* is more prevalent. For conventional farming, insect meshing is not financially viable.

1.7 Monitoring and control of *Nasonovia ribisnigri* in lettuce

There is a suite of monitoring, control and mitigative strategies to aid at reducing the impact that insects cause to crops. This can range from simple but effective mesh covers to small remote drones delivering targeted chemical or biological control. With pressures from the public, NGOs government and the EU, a grower's repertoire of chemical control is constantly dwindling, largely motivated by a move to 'green horticulture' and a change in regulation following Brexit (DEFRA, 2021b). While the removal of proven environmentally detrimental chemicals is positive, as shown for the neonicotinoids class of agrochemicals (Sánchez-Bayo *et al.*, 2014; Woodcock *et al.*, 2016), there are few suitable replacements. Since 1st January 2021, an independent pesticides regulatory regime is now in operation in Great Britain (GB) following Brexit (HSE, 2021). Treated seed with a product authorised in an EU member state can continued to be used and traded in Great Britain until 31 December 2023. After this point, it is unclear what limitations or changes that will take effect after this transitional period. Consequently, growers may have further restrictions and limitations in what can be applied to crop. Moreover, the use of the same or similar compounds that utilises the same mode of action (MoA) or pathway is likely to facilitate rapid resistance development in target pest insects and render current controls ineffective (Bass *et al.*, 2014; Bass *et al.*, 2015; Sparks and Nauen, 2015; Kandel *et al.*, 2019).

1.7.1 Insect traps

The Rothamsted Research Insect survey (RIS) based in Harpenden, Hertfordshire (UK), have been operating two trap networks throughout each year since 1964 (RIS, 2021). Firstly, a light trap network, currently consisting of around 80 traps across the UK and Ireland providing valuable information on moths. Secondly, a suction trap network, currently consisting of 16 12.2m high suction traps throughout the UK which provides information on aphids and many other migrating insects to researchers, growers, conservation organisations and policy makers. The phenology and cumulative numbers throughout the season aphids important to horticulture is reported in weekly bulletins that provide an effective decision support tool for horticulture, Of the 670 aphid species in the UK, the 12.2m suction traps capture ~400 species, with an average of between 130-140 species captured each year, including ~40 that are known pests of various crops and ornamentals (Bell *et al.*, 2015; Bell *et al.*, 2012; RIS, unpublished). The RIS bulletin particularly reports on the top 21 aphid species which are especially damaging to horticulture, of which *N. ribisnigri* is

included. Other suction trap monitoring networks exist in Europe and other parts of the world such as China (Jiang *et al.*, 2012; Yu-Qing *et al.*, 2011).

A caveat to the utility of data derived from *N. ribisnigri* caught in the RIS suction traps is that it is one of the aphid species that appears in very low numbers, particularly in more recent years. As a result, it is very difficult to determine when its migratory flights from the primary (winter) host to its secondary hosts (which includes lettuce) in the spring and then back from the secondary host to the primary host in autumn occur.

Aphids migrate using two different modes; appetitive flight in which local movements of no more than 200 m are undertaken in still to light winds and within the flight boundary layer below 10 m; long distance migration in which aphids enter the convective boundary layer above the flight boundary layer and may travel several hundreds of kilometres (Parry *et al.*, 2013). Aphids usually use a combination of visual and olfactory cues in host plant locate to 'decide' when to descend from the boundary or jet stream layer (Berry and Taylor, 1968). Despite their size, aphids are not completely at the mercy of the prevailing winds and are more than 'aerial plankton' (Reynolds and Reynolds, 2009). Although they are not strong fliers, with directed flight below wind speeds of 3m/sec (Hassall *et al.*, 2021), aphids have been found to fly considerable distances, with active flights from 7 to 12 hours (Cockbain, 1961). The general consensus for aphid flight is that they control flight direction in the boundary layer, but it is determined by the wind at higher altitudes (Loxdale *et al.*, 1993).

With the passive sampling nature of the suction traps, attractive traps are also used to catch and monitor a suite of insects. Other methods of monitoring aphids exist such as the UK aphid monitoring network, operated by Fera, which use a network of approximately 100 yellow water traps situated in and around seed potato crops through Britain. Typically, aphids respond positively to green and yellow and negatively to red and blue light (Döring and Chittka, 2007) but is not a complete rule. Yellow has been shown to be the most effective colour at attracting aphids and subsequently has been incorporated in a variety of experiments (De Barro, 1991; Kennedy *et al.*, 1961; Roach and Agee, 1972). However, it has been shown that not all Hemiptera are attracted by the colour yellow or green, with some psyllids preferring colours more in the red spectrum of visible light (Farnier *et al.*, 2014).

These current methods of trapping and detecting *N. ribisnigri*, with low numbers caught in Rothamsted Insect Survey (RIS) suction traps and AHDB water traps, makes this aphid difficult to survey and create accurate forecasting models. Without this information, the main method of

detecting *N. ribisnigri* within the crop is to conduct regular monitoring of outdoor lettuce crop. Regular crop monitoring is not an effective or efficient method, as it is very time consuming to survey large areas and only provides information once *N. ribisnigri* is in the crop. With the lack of an early detection system, it is usually too late to implement mitigation strategies, due to well-established populations of *N. ribisnigri* within the crop. There is an evident gap in the knowledge in understanding the migration of *N. ribisnigri* into outdoor lettuce and forecasting the likelihood of an infestation.

1.7.2 Biological control and IPM

Biological control approaches over the past decades are being adopted more frequently and employed very effectively into Integrated Pest Management (IPM) strategies. The use of IPM strategies, focusing on biological control, are proving to be an effective method at minimising and mitigating crop loss in a variety of horticultural systems (Cowan and Gunby, 1996; Hassanai *et al.*, 2008). Since *N. ribisnigri* is regarded as the most damaging species of aphids on outdoor lettuce, there has been considerable research into potential biological control methods, both in laboratory, glasshouse, and semi-field environments.

The introduction of alternative plant resources in a field experiment in Barcelona, Spain, was shown to aid in reducing the number of both *N. ribisnigri* and thrips (*Frankliniella occidentalis*) (Pergande, 1895) by encouraging predators such as hoverflies (*Syrphidae*) and predatory bugs (*Hemiptera*) into the area (Alomar and Gabarra, 2008). Furthermore, with the establishment of predatory insects in the flower patches, *N. ribisnigri* numbers were shown to be reduced below 'economic threshold', unless pesticide treatments were applied. With the accidental introduction of *N. ribisnigri* to California, USA, from Europe during the late 1990's, they soon became an economic pest to outdoor lettuce, especially in Romaine lettuce (Hopper *et al.*, 2010). Intercropping outdoor lettuce with alyssum (*Lobularia maritima*) in California, USA, was shown to encourage hoverflies and provide biological control of lettuce aphids (Brennan, 2013). It was shown however, that lettuce heads were slightly smaller in intercropped lettuce compared to a monoculture of lettuce due to nitrogen competition by alyssum. Intercropping coriander (*Coriandrum sativum*) and chrysanthemum (*Chrysanthemum coronarium*) failed to reduce the level of infestations of *N. ribisnigri* in outdoor lettuce crop, with the severity of infestation being significantly greater for intercropped compared to lettuce monoculture (Pascual-Villalobos, 2006). Similarly, the four most common hoverflies in California's central coast have been shown to be

effective at consuming *N. ribisnigri*. The most voracious species (*Eupeodes fumipennis*) (Thomas, 1869) consumed 507 aphids during larval development in laboratory studies and highlights their potential as good biological control agents (Hopper *et al.*, 2010). Another study, in the central coast of California in organic lettuce, highlighted that it is the combination of hoverfly species creating a 'syrphid complex' which exploit different niches that reduce the number of aphids in lettuce (Smith and Chaney, 2007). It was suggested that in some locations where adequately control *N. ribisnigri* outbreaks failed, this could be due to lack of floral resources to support the necessary level for female hoverflies.

Lacewings (Neuroptera) have also proven to be effective predators at reducing the numbers of *N. ribisnigri* (Quentin *et al.*, 1995; Shrestha and Enkegaard, 2013). Both adults and larvae of lacewings can be used to control aphids, amongst other insect crop pests (Chang, 1996; Hagen *et al.*, 1976; Rocca and Messelink, 2017). However, it is the larvae which are highly voracious and, as a result, are also available for both commercial and domestic use to control insect pests (Isomailov *et al.*, 2020; Tauber *et al.*, 2000).

With the evident beneficial effect of predators reducing *N. ribisnigri* numbers, parasitoid wasps have also been shown to affect *N. ribisnigri* populations. Three commercially available parasitoid wasps, *Aphidius colemani* (Viereck, 1912), *Lysiphlebus testaceipes* (Cresson, 1880) and *Aphelinus abdominalis* (Dalman, 1820) were demonstrated to parasitise *N. ribisnigri* in laboratory conditions (Shrestha *et al.*, 2014). Due to lower levels of parasitisation and mortality observed in *A. colemani* and *L. testaceipes*; only *A. abdominalis* was suggested to have the potential to be used as an effective control against *N. ribisnigri* in the field. Further research highlights the ability of *A. abdominalis* to parasitise *N. ribisnigri* (Shrestha *et al.*, 2017). However, compared to another horticulturally important aphid pest of lettuce, *M. euphorbiae*, *A. abdominalis* preferentially parasitised *M. euphorbiae* over *N. ribisnigri* most likely due to feeding behaviour. Since *N. ribisnigri* feeds predominantly on fresh new growth in tight colonies and *M. euphorbiae* feeds on older leaves in looser colonies on the more outer parts of the lettuce, *M. euphorbiae* are generally more accessible. This is particularly apparent on heart leaf lettuce varieties in which the later growth stage of the lettuce can encapsulate *N. ribisnigri* and thus further hinder parasitisation detection and entry to the inner part of the lettuce. Farsi *et al.*, 2020 highlighted that *Aphidius matricariae* (Haliday, 1834) could also be a suitable biological control agent against *N. ribisnigri* as laboratory experiments demonstrated that the higher density of *N. ribisnigri* colonies increased the amount of parasitisation in *A. matricariae* during their life span. Despite the proven ability of certain parasitoid wasps to reduce the numbers of *N. ribisnigri* and control outbreaks under laboratory

conditions, the effectiveness of control in a field setting is yet to be demonstrated. Furthermore, level of control by proven parasitoid wasps on outdoor lettuce are not likely to be effective on their own. This is especially apparent in whole head lettuce whereby; the leaves form tight layers around one another, preventing easy access for a parasitoid and if more than one aphid species is present, these will likely be preferred over *N. ribisnigri*.

Entomopathogenic fungi are widely used in IPM management approaches to reduce pests below economic threshold (Castrillo *et al.*, 2010; Hesketh *et al.*, 2008). It has been demonstrated that an entomopathogenic fungus, *Verticillium lecanii*, is effective at reducing the numbers of *N. ribisnigri* in laboratory conditions (Fournier and Brodeur, 2000). Out of the three aphids used in experiments (*M. persicae*, *M. euphorbiae* and *N. ribisnigri*), *N. ribisnigri* was significantly more susceptible to *V. lecanii*, with 100% mortality recorded after 2 days at the highest dose. Unfortunately, both greenhouse studies and semi-field experiments failed to provide adequate control (Åsman, 2007; Fournier and Brodeur, 2000). The entomopathogenic fungus *Beauveria bassiana* strain GHA has been shown to be a potential candidate for the control of *N. ribisnigri* (Shrestha *et al.*, 2015). In semi-field conditions, *B. bassiana* was able to reduce the population of *N. ribisnigri* on lettuce by 38% in 5 days, with further declines of 92% and 99% after 11 and 20 days, respectively. Nymph production was significantly affected using the highest dose of *B. bassiana*, with the highest effect observed on fourth instar alate *N. ribisnigri*. In a field study in Argentina, it was shown that the entomopathogenic fungus, *Pandora neophidis*, had successfully infected *N. ribisnigri* in both the edge and centre of lettuce crop, with host density playing an important factor in determining infection levels (Scorsetti *et al.*, 2010). The use of another two biopesticides have shown promise in reducing the populations of *N. ribisnigri*. A biopesticide using *B. bassiana* (Botanigard ES and Naturalis L) further demonstrated the ability of *B. bassiana* to reduce *N. ribisnigri* numbers on lettuce (Prince and Chandler, 2020). However, two others tested biopesticides which incorporated *Cordyceps fumosorosea* (Preferal WG) and *Akanthomyces dipterigenus* (Vertalec) showed a reduction in the progeny in *M. persicae* but had no effect on *N. ribisnigri*.

There are many biological control agents which have proven efficacy at controlling and reducing the number of *N. ribisnigri* on lettuce, some of which are commercially available for the use of controlling other unrelated insect pests. Most of the research conducted was under laboratory or semi-field experimental conditions, with few applying this to a horticultural setting. Additionally, with the few experiments conducted in a semi-field experiment, the results were not as prominent as found under purely a laboratory setting. This could be a result of either experimental design or

external variables (meteorological, abiotic, biotic) or a combination. Some insect pests are currently unable to be effectively controlled by certain IPM methods (Abudulai *et al.*, 2006). Overall, this is true for *N. ribisnigri*, not because of a lack of biological control but because of the inherent growing practises and consumer standards that make the utilisation of IPM difficult to overcome. With the characteristically short growing period of lettuce in a field (between 4-6 weeks during a UK summer) combined with the lag-time for biological control agents to take effect, are often not sufficient to reduce *N. ribisnigri* below threshold numbers prior to harvesting on outdoor lettuce (Smith and Chaney, 2007). In the instance of whole head lettuce and other varieties that grow in this manner, whether it is an aphid or other 'beneficial' insect, it is regarded as a contaminant and can result in consumer rejections.

1.8 Resistance in *N. ribisnigri*

1.8.1 Insecticide resistance in *Nasonovia ribisnigri*

Out of the four main aphid pests of lettuce, *Pemphigus bursarius*, *M. persicae*, *M. euphorbiae* and *N. ribisnigri*; *M. persicae* and *M. euphorbiae* have a diverse number of other crop host plants (Ellis *et al.*, 1995). *Nasonovia ribisnigri* is often the more severe pest of lettuce and much more specific to lettuce than the other two aphid species. Conventionally, *N. ribisnigri* have been controlled using insecticides, however; resistance to a variety of chemicals has been reported throughout Europe and other parts of the world (Rufingier *et al.* 1997; Kift *et al.*, 2004). In five tested insecticides (endosulfan, deltamethrin, carbamate (pirimicarb), methomyl and acephate), two tested *N. ribisnigri* strains from France and Spain had resistance to all (with a fold increase of 680, 26, 62, 46 and 11, respectively) (Rufingier *et al.*, 1997). Field populations collected in UK of *N. ribisnigri* on lettuce demonstrated a varied level of resistance to pirimicarb, with lower but still varied resistance to organophosphates insecticides (Barber *et al.* 2002). When a population of *N. ribisnigri* establish on outdoor lettuce crops, chemical spraying is rendered ineffective due to their preference to infest the lettuce head, where they are inaccessible to non-systemic insecticides (Natwick and Lopez, 2016). This is especially true when *N. ribisnigri* infestations are on developed heart leaves of lettuce and prevent successful penetration of insecticides, which rely on direct contact with the organism to have an effect. The systemic insecticide (spirotetramat), but applied as a foliar application, provides a new mode of action against 'sucking insects' on plants by inhibiting insect growth and development and inhibits the biosynthesis of lipids (Salazar-López, 2015). This has provided conventional growers a new mode of action with no cross-resistance to other insecticides (Brück *et al.*, 2009). Ongoing resistance screening has shown no evidence of

resistance to spirotetramat in the peach-potato aphid (*M. persicae*) (IRAG, 2019). However, in spring 2021, a *M. persicae* clone in Queensland (Australia) were shown to have developed resistance to spirotetramat (Singh *et al.*, 2021).

Other systemic insecticides such as neonicotinoid seed treatments have been used as an alternative method of controlling herbivorous pest insects since the 1990's (Palumbo *et al.*, 2001). Imidacloprid, a systemic seed treatment was effective at controlling aphids for the first few weeks of a crop's life (including lettuce) (PSD, 2003). In recent years, these types of neuro-active insecticides have been under increased scrutiny by both the public and scientific community, which has resulted in partial bans and restrictions throughout the EU. With increasing governmental restrictions and pressures on herbicide and insecticide use, future methods and applications for aphid control are uncertain.

1.8.2 Resistant-breaking *Nasonovia ribisnigri* biotypes (Nr-gene in lettuce)

Plant resistant cultivars to insect pests remain an effective method to control infestations and reduce yield (Akköprü *et al.*, 2015, Asante *et al.*, 2001; de Morais and Pinheiro, 2012). These resistant cultivars play an integral role in IPM approaches and provide another tool in a grower's repertoire. Additionally, the use of resistant cultivars is biologically, economically, and ecologically practicable (Wiseman, 1994). They reduce the overuse and reliance of insecticides, have different target sites and mode of actions to other conventional methods and ultimately increase the longevity of both resistant cultivars and insecticides (Cheng *et al.*, 2010; Feyereisan, 2015).

Plants can range from being completely resistant to susceptible to the damage caused, due to variations in their genetics (Kogan, 1994). In response to aphids feeding, plants exhibit a type of resistance however, the more 'resistant' the plant, the more effective antixenotic or antibiotic effects (or a combination of the two) they have on insects (Kogan, 1994). These two effects either deter an insect from colonising a plant altogether or produces an adverse or antibiotic effect on the insect, respectively (Stout, 2013). These effects can be result of a change in chemical composition and/or plant structure. A third understudied option exists – plants can be 'tolerant' to insect herbivory (Peterson *et al.*, 2017). Tolerance is suggested to be a more sustainable pest management strategy as it only involves a plant response and therefore does not cause insect resistance in target pest insects (Peterson *et al.*, 2017). Unfortunately, lettuce is required to be contaminant free and, as a result, only antibiotic and antixenosis can be used.

For sap-feeding insects such as aphids, two mechanisms of plant response are initiated upon recognition of feeding. A change in host plant chemistry causes a non-specific stress response and the second process involves specific responses by plant resistance genes in response to a feeding insect, known as gene-for-gene interaction (Kaloshian, 2004; Smith and Boyko, 2007). A gene-for-gene response mechanism is only present in plants that possess these *R* genes and has resulted in the development of novel insect pest-specific control methods (Rawat *et al.*, 2012; Zhao *et al.*, 2015). It has been shown that several plant resistance genes exist in several insects such as the the Hessian fly (*Matetiola destructor* (Say, 1817)) (Hatchett and Gallun, 1970 and the brown plant hopper (*Nilaparvata lugens* (Stal, 1854)) (Lv *et al.*, 2014). In cultivated vegetables, several plant resistance genes have been identified and include a gene (*vat*) which confers resistance to the melon aphid (*Aphis gossypii*) (Glover, 1877) (Lombaert *et al.*, 2009) and another gene (*Mi-1*) conferring resistance to *M. persicae* (Bhattarai, 2009).

To tackle the complex issues concerning insecticides and the changing legislation pressures, a *N. ribisnigri* resistant (Nr) cultivar was introduced into the grower's repertoire during the early 1980's which originated from a wild species of lettuce (*Lactuca serriola*) containing natural resistance to *N. ribisnigri* that was subsequently introgressed into lettuce (Eenink *et al.*, 1982). This resistance conferred near to complete resistance to *N. ribisnigri* and was incorporated by many growers to reduce the loss of lettuce to this aphid species. The resistance in plants is usually quantitative trait, which is a result of two or more genes provide partial and varying levels of resistance (polygenic) (Dogimont *et al.*, 2010). The other resistance in plants is monogenic in which resistance is controlled by a single gene. The Nr-gene in resistant lettuce cultivars is an example of a monogenic gene in which introgression into lettuce is inherited dominantly. This resistance associated with the Nr-gene is unlikely to be mechanical nor a change in host plant quality, as other aphid species would likely be affected (both *M. persicae* and *M. euphorbiae* are unaffected), so the resistance mechanisms involved are most likely species-specific (van Helden and Tjallingii, 1993). It has been demonstrated via electrical penetration graph (EPG) that the susceptible (Nr:0) aphids are able to penetrate the lettuce tissue but are unable to feed and therefore move off the unsuitable plant (van Helden, 1995). Additionally, EPG can detect the different activities of an aphid stylet, including, probing, passive uptake of phloem sap, xylem absorption and salivation into the sieve element (Figure. 1.2). As a result, the location of resistance is thought to reside in the phloem of *L. sativa* during sap ingestion from the sieve element, but the exact resistance mechanism or pathway involved is still unknown (ten Broeke *et al.*, 2013). The *N. ribisnigri* biotype which is susceptible to the Nr-gene in lettuce cultivars and the resistance-breaking (*Rb*) biotypes were first

reported in 2007 in Europe. These two biotypes were assigned as Nr:0 and Nr:1 whereby; Nr:0 is susceptible and Nr:1 is resistance-breaking (Thabuis *et al.*, 2011). The *Nr* lettuce cultivars were grown widely (McCreight and Liu, 2012). van der Arend (2003) predicted that due to the potential overuse, high selection pressure, a high level of host specificity and single gene resistance, resistant breaking biotypes of *N. ribisnigri* had the potential to evolve rapidly. To date, there are no other suitable resistant lettuce cultivars against *N. ribisnigri* that have been developed, despite many efforts to change this (Walley *et al.*, 2017; McCreight and Liu, 2012; McCreight, 2008).

1.9 Aphid genetics

The initial meeting in Paris at the International Aphid Genomics Consortium (IAGC) in 2003, highlighted the importance of developing and maintaining genomic resources for studying aphids, particularly those which are horticulturally and economically important. In 2010, the only aphid genomic model available to study aphid genetics and aphid-plant interactions was the pea aphid (*A. pisum*) (International Aphid Genomics Consortium, 2010). This was a milestone for aphid genomics, and the genome is still used 11 years later, in a variety of genomic comparative studies, amongst other genetic analyses (Mathers, 2020; Mathers *et al.*, 2021; Shahid *et al.*, 2021). There are currently 28 aphid genomes, from 12 different species available for genomic analyses: ranging from draft assemblies to chromosome-level assembly and annotation (BIPAA, 2021).

The number of genomes and transcriptomes being assembled and annotated provides researchers with excellent resources to further the knowledge and address a multitude of fundamental questions, which would have been unattainable without these assemblies (Jayasinghe *et al.*, 2021). Assemblies are continuously being improved, with additional sequencing data, new technologies, and manual annotation improvement. A recent paper highlighted the importance of revisiting draft assemblies and incorporated a k-mer based duplication pipeline to improve the accuracy of a previous assembled and annotated genome of the black cherry aphid, *Myzus cerasi* (Walker, 1849) (Mathers *et al.*, 2020). Reusing existing data sets with improved or new bioinformatic approaches improves the quality of genomic resources available for aphids and other taxonomic groups. By improving existing genomes to a higher quality enables researchers to explore and understand fundamental questions such as aphid evolution and adaptation.

The incorporation of population genetics in aphids have elucidated some fundamental questions about aphid ecology and evolution. A New Zealand study on the green spruce aphid, *Elatobium abietinum* (Walker, 1849), demonstrated that this pest aphid of spruce trees had no genetic

variation, despite *E. abietinum* samples being collected across six sites and up to 1200km apart (Nicol *et al.*, 1998). In contrast, the genetic variation between UK populations of *E. abietinum* were high between and within sites. This suggests that the populations of *E. abietinum* in New Zealand which could be result of limited founder population and the absence of holocyclic reproduction. Low level of genetic diversity can seriously reduce the ability of a population to adapt to control strategies and help identify a strategy to mitigate future problems.

The knowledge of aphid pest ecology and life history traits such as sexual and asexual reproduction (holocycly) enables smarter control programmes to be created. Since genetic variation determines the rate of adaption and evolution in a population it is important to understand the level of gene flow and genetic diversity between horticulturally important insects (Barrett and Schluter, 2008). Low genetic diversity creates an evolutionary bottleneck and limits the ability of populations to overcome high selection pressures (such as insecticide resistance) (Leberg, 1992; Van Leeuwen *et al.*, 2008). The development of microsatellite markers for the cereal aphid, *Sitobion avenae* (Fabricius, 1775), has shown that a strong geographical divide between populations in the north and south of France exists (Simon *et al.*, 1999).

Despite superficial similarities, pest aphids can exhibit huge diversity, both inter- and intra-specific, with complex life history traits. From a pest management perspective, it is imperative to understand these nuances between aphid species of interest to implement appropriate mitigative strategies. This is especially important in polyphagous aphid species such as *M. persicae* and *S. avenae* in which it has been shown to have a high level of genetic diversity between populations (Guillemaud *et al.*, 2003; Vialatte *et al.*, 2005). It was shown that populations of *S. avenae* had a high level of genetic diversity between cultivated and uncultivated host plants (Vialatte *et al.*, 2005). This limited gene flow between *S. avenae* in uncultivated plants can generally be disregarded from an IPM approach as these clones do not move onto cultivated cereal crops and the uncultivated host plants are not acting as a reservoir or green bridge. Moreover, the *S. avenae* on uncultivated host plants would likely aid in attracting beneficial insects to help control outbreaks of aphids on crop. However, since the exchange of genes in the populations of *S. avenae* is not entirely separate, the populations on uncultivated host plants could provide a reservoir of genetic diversity and therefore exacerbate the adaptive potential of *S. avenae* on cultivated crop.

In the UK, population genetic analyses have been incorporated to provide insight into aphid pest ecology and evolution. In Scotland, it was found that the most common source of the peach-potato aphid, *M. persicae*, populations were from local overwintering asexual populations (Fenton

et al., 2003). No difference was found in the distribution of genotypes between *M. persicae* collected from potato or brassica crop, suggesting that host-adapted genotypes have little to no advantage in field (Fenton *et al.*, 2003). However, it has recently been shown that global *M. persicae* populations exhibit genetic differentiation based on host-plant association and landscape (Singh *et al.*, 2021). The bird cherry-oat aphid, *Rhopalosiphum padi* (Linnaeus, 1758), was shown to form two genetic clusters which corresponded to a north-south divide in the UK (Morales-Hojas *et al.*, 2019). These two genetic clusters do not correspond to different reproductive forms and suggests that holocyclic reproduction is the dominant form of reproduction in *R. padi* in England. However, it was shown that the dominant reproduction method of *S. avenae* in the UK was anholocyclic, with a spatial expansion of a single genetic cluster (Morales-Hojas *et al.*, 2020). This is potentially a recent response changing horticultural practises and adaption to insecticide resistance.

Therefore, genetic analysis can play an important role in understanding pest aphid ecology and aid in future control and mitigation strategies. No reference genome for *N. ribisnigri* currently exists. To date, only the mitochondrial cytochrome c oxidase subunit I (COI) barcode and a selection of odorant binding proteins (OBP) have been sequenced (Northey *et al.*, 2016). It would be desirable to assemble a reference genome of *N. ribisnigri* since it is the most damaging pest on outdoor lettuce and would provide valuable information in future control strategies. Additionally, the use of population genetics would help elucidate the movements, life history traits and ecology of this cryptic aphid.

1.10 Thesis aims, objectives and research questions

The overall aim of this thesis was to improve the detection and control of the currant-lettuce aphid (*N. ribisnigri*). To achieve this, the project investigated the ecological processes governing the presence of *N. ribisnigri* in lettuce crops, and the autecology and population processes surrounding resistance-breaking biotypes.

Firstly, working alongside an industry partner, this project aimed to improve existing monitoring and forecasting strategies and study alternative summer hosts and overwintering biology *in situ*. Secondly, this project aimed to *de novo* assemble a reference genome of the currant-lettuce aphid, *Nasonovia ribisnigri*, from a susceptible biotype (Nr:0) and resistant biotype (Nr:1) to the Nr-gene found in the lettuce host plant (*Lactuca sativa*). Finally, to complement the newly assembled genome, a *N. ribisnigri* transcriptome was to be assembled using short-read RNA data. Subsequently, with the addition of RNA-seq sequence data, genome-wide changes in gene

expression of two *N. ribisnigri* biotypes (Nr:0 and Nr:1) were to be investigated in order to elucidate potential mechanisms of resistance to the Nr-gene in lettuce.

Research questions

Chapter 2

Key questions: Has the population of *N. ribisnigri* changed over the past 50 years and what could have affected it? With the data available, is it possible to model first flight of *N. ribisnigri* in order to improve existing forecasting models? Can different trapping methodologies improve capture of *N. ribisnigri*?

Aim 1: Assess whether the overall population trend of *N. ribisnigri* in the UK has declined.

Aim 2: Test whether a first flight model can be used to understand migration into crop.

Aim 3: Determine whether *N. ribisnigri* fly at higher densities at heights lower than the 12.2m Rothamsted Research suction traps.

Chapter 3

Key questions: Is *N. ribisnigri* preferentially using an alternative secondary summer host prior to migration onto lettuce later in the summer? Is it possible to detect trace host plant DNA in the midgut of *N. ribisnigri* after a feeding event? Can *N. ribisnigri* move from a known winter host straight onto lettuce (summer host) in the spring?

Aim 1: Identify potential alternative summer hosts of *N. ribisnigri* in different field margin types in both organic and conventional farms in Cambridgeshire and Norfolk.

Aim 2: Incorporate known alternative summer host plant plots to encourage *N. ribisnigri* colonisation and study feeding behaviour in both organic and conventional farms in Cambridgeshire and Norfolk.

Aim 3: Develop a molecular method to identify previous host plants of *N. ribisnigri* using the plant barcode (ITS2) to identify host plants that could be acting as a green bridge onto lettuce.

Aim 4: Semi-field overwintering experiment to observe colonisation of known winter hosts (*Ribes* spp.) and understand whether a lettuce host plant will be accepted the following spring during alate migration (dispersal).

Chapter 4

Key questions: What differentially expressed genes exist between wild type and resistance-breaking biotypes and could any be responsible for the resistance-breaking mechanism?

Aim 1: *De novo* assemble a draft *N. ribisnigri* genome of both the wild type and resistant biotype to the *Nr* resistance gene in lettuce.

Aim 2: Use RNA-seq analysis to study differentially expressed genes between wild type and resistant *N. ribisnigri* biotypes and identify potential candidate genes responsible.

Chapter 5

Key questions: Is the level of gene flow between populations of *N. ribisnigri* in England high or low and has the evolution of resistant-breaking biotypes influenced population structure in regions of high lettuce production (i.e., East of England)?

Aim 1: To understand the population structure of *N. ribisnigri* in England using microsatellite markers, both spatially and temporally.

Aim 2: Determine whether the introduction of resistant lettuce cultivars influenced the population structure of *N. ribisnigri* in areas of high lettuce production (East of England).

Key findings

- Significant declines since 1965 in the abundance of *N. ribisnigri* in the UK have been identified, but there is an East/West divide in trends between populations.
- Population density is low in the field and densities of migrating aphids are comparatively small, which is likely to be due to a weak propensity for flight.
- An intermediate, unidentified alternative secondary host plant is being used as a green bridge prior to *N. ribisnigri* colonising outdoor lettuce.
- A molecular method to identify use of a particular host plant by individual aphids has been developed and two new alternative secondary hosts have been identified.
- A draft genome has been assembled and annotated and the potential gene responsible for the mechanism for overcoming host plant resistance has been identified.

1.11 References

- Abudulai, M., Abatania, L. and Salifu, A.B., 2006. Farmers' knowledge and perceptions of cotton insect pests and their control practices in Ghana. *Journal of Science and Technology (Ghana)*, pp.39-46.
- Adams, A.N. and Thresh, J.M., 1987. Reversion of black currant. *Virus Diseases of Small Fruits.*, pp.133-136.
- Akköprü, E.P., Atlıhan, R., Okut, H. and Chi, H., 2015. Demographic assessment of plant cultivar resistance to insect pests: a case study of the dusky-veined walnut aphid (Hemiptera: Callaphididae) on five walnut cultivars. *Journal of Economic Entomology*, pp.378-387.
- Alomar, O., Arnó, J. and Gabarra, R., 2008. Insectary plants to enhance the biological control of *Nasonovia ribisnigri* and *Frankliniella occidentalis* in lettuce. *IOBC/wprs Bull*, pp.9-12.
- Alvarez, A.E., Broglia, V.G., Alberti D'Amato, A.M., Wouters, D., van der Vossen, E., Garzo, E., Tjallingii, W.F., Dicke, M. and Vosman, B., 2013. Comparative analysis of *Solanum stoloniferum* responses to probing by the green peach aphid *Myzus persicae* and the potato aphid *Macrosiphum euphorbiae*. *Insect science*, pp.207-227.
- Annan, I.B., Saxena, K.N., Schaefer, G.A. and Tingey, W.M., 1994. Effects of infestation by cowpea aphid (Homoptera: Aphididae) on different growth stages of resistant and susceptible cowpea cultivars. *International Journal of Tropical Insect Science*, pp.401-410.
- Asante, S.K., Tamo, M. and Jackai, L.E.N., 2001. Integrated management of cowpea insect pests using elite cultivars, date of planting and minimum insecticide application. *African crop science journal*, pp.655-665.
- Åsman, K., 2007. Aphid infestation in field grown lettuce and biological control with entomopathogenic fungi (Deuteromycotina: Hyphomycetes). *Biological Agriculture & Horticulture*, pp.153-173.
- Baker, E., Dransfield, R. D., and Brightwell, R., 2015. Current British Aphid Checklist. Influential Points. Available at: https://influentialpoints.com/aphid/Checklist_of_aphids_in_Britain.htm (accessed 12 October 2020).
- Barber, M., 2002. Insecticide resistance in the currant-lettuce aphid, *Nasonovia ribisnigri*, University of Bath.
- Barrett, R.D. and Schluter, D., 2008. Adaptation from standing genetic variation. *Trends in ecology & evolution*, pp.38-44.
- Bass, C., Denholm, I., Williamson, M.S. and Nauen, R., 2015. The global status of insect resistance to neonicotinoid insecticides. *Pesticide Biochemistry and Physiology*, pp.78-87.
- Bass, C., Puinean, A.M., Zimmer, C.T., Denholm, I., Field, L.M., Foster, S.P., Gutbrod, O., Nauen, R., Slater, R. and Williamson, M.S., 2014. The evolution of insecticide resistance in the peach potato aphid, *Myzus persicae*. *Insect biochemistry and molecular biology*, pp.41-51.
- Bejer-Petersen, B., 1962. Peak years and regulation of numbers in the aphid *Neomyzaphis abietina* Walker. *Oikos*, pp.155-168.

Bell, J.R., Alderson, L., Iзера, D., Kruger, T., Parker, S., Pickup, J., Shortall, C.R., Taylor, M.S., Verrier, P. and Harrington, R., 2015. Long-term phenological trends, species accumulation rates, aphid traits and climate: Five decades of change in migrating aphids. *Journal of Animal Ecology*, pp.21-34.

Bell, J.R., Taylor, M.S., Shortall, C.R., Welham, S.J. and Harrington, R., 2012. The trait and host plant ecology of aphids and their distribution and abundance in the United Kingdom. *Global Ecology and Biogeography*, pp.405-415.

Berry, R.E. and Taylor, L.R., 1968. High-altitude migration of aphids in maritime and continental climates. *The Journal of Animal Ecology*, pp.713-722.

Bhattacharai, K.K., 2009. *Characterization of Mi-1-mediated defense responses in tomato to root-knot nematodes and potato aphids*. University of California, Riverside.

Blackman, R.L. and Eastop, V.F., 1984. *Aphids on the World's Crops, An Identification Guide*. First edn. Chichester: John Wiley & Sons.

Blackman, R.L., 1971. Variation in the photoperiodic response within natural populations of *Myzus persicae* (Sulz.). *Bulletin of Entomological Research*, pp.533-546.

Brennan, E.B., 2013. Agronomic aspects of strip intercropping lettuce with alyssum for biological control of aphids. *Biological Control*, pp.302-311.

Brück, E., Elbert, A., Fischer, R., Krueger, S., Kühnhold, J., Klueken, A.M., Nauen, R., Niebes, J., Reckmann, U., Schnorbach, H., Steffens, R. and van Waetermeulen, X., 2009. Movento®, an innovative ambimobile insecticide for sucking insect pest control in agriculture: Biological profile and field performance. *Crop Protection*, 838-844.

CABI/EPPO. (2003). *Nasonovia ribisnigri*. [Distribution map]. CABI Distribution Maps of Plant Pests, 1st Edition, Map 644.

Carter, C.I., 1972. Winter temperatures and survival of the green spruce aphid *Elatobium abietinum* (Walker). *Winter temperatures and survival of the green spruce aphid *Elatobium abietinum* (Walker)*.

Castrillo, L.A., Bauer, L.S., Liu, H., Griggs, M.H. and Vandenberg, J.D., 2010. Characterization of *Beauveria bassiana* (Ascomycota: Hypocreales) isolates associated with *Agrilus planipennis* (Coleoptera: Buprestidae) populations in Michigan. *Biological Control*, pp.135-140.

Chang, G.C., 1996. Comparison of single versus multiple species of generalist predators for biological control. *Environmental Entomology*, pp.207-212.

Cheng, X., Chang, C. and Dai, S.M., 2010. Responses of striped stem borer, *Chilo suppressalis* (Lepidoptera: Pyralidae), from Taiwan to a range of insecticides. *Pest management science*, pp.762-766.

Cockbain, A.J., 1961. Fuel utilization and duration of tethered flight in *Aphis fabae* Scop. *Journal of Experimental Biology*, pp.163-174.

Cowan, R. and Gunby, P., 1996. Sprayed to death: path dependence, lock-in and pest control strategies. *The economic journal*, pp.521-542.

- De Barro, P.J., 1991. Attractiveness of four colours of traps to cereal aphids (Hemiptera: Aphididae) in South Australia. *Australian Journal of Entomology*, pp.263-264.
- De Bokx, J.A. and Piron, P.G.M., 1990. Relative efficiency of a number of aphid species in the transmission of potato virus YN in the Netherlands. *Netherlands Journal of Plant Pathology*, pp.237-246.
- de Morais, A.A. and Pinheiro, J.B., 2012. Breeding for resistance to insect pests. In *Plant Breeding for Biotic Stress Resistance*. Springer, Berlin, Heidelberg. pp.103-125.
- De Vries, I.M., 1997. Origin and domestication of *Lactuca sativa* L. *Genetic Resources and Crop Evolution*, pp.165-174.
- Dedryver, C.A., Le Ralec, A. and Fabre, F., 2010. The conflicting relationships between aphids and men: a review of aphid damage and control strategies. *Comptes rendus biologiques*, pp.539-553.
- DEFRA (a) (Department for Environment, Food and Rural Affairs)., 2021. DEFRA Statistics: Agriculture Facts England Regional Profiles March 2021.
- DEFRA (b) (Department for Environment, Food and Rural Affairs)., 2021. The Path to Sustainable Farming: An Agricultural Transition Plan 2021 to 2024.
- DEFRA (Department for Environment, Food and Rural Affairs)., 2020. Horticulture Statistics 2020.
- Delmotte, F., Leterme, N., Bonhomme, J., Rispe, C. and Simon, J.C., 2001. Multiple routes to asexuality in an aphid species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, pp.2291-2299.
- Denmark, H.A., 2005. Giant Bark Aphid, *Longistigma caryae* (Harris) (Insecta: Hemiptera: Aphididae). *EDIS*, 2005(1).
- Diaz, B.M. and Fereres, A., 2005. Life table and population parameters of *Nasonovia ribisnigri* (Homoptera: Aphididae) at different constant temperatures. *Environmental entomology*, pp.527-534.
- Diaz, B.M., Barrios, L. and Fereres, A., 2012. Interplant movement and spatial distribution of alate and apterous morphs of *Nasonovia ribisnigri* (Homoptera: Aphididae) on lettuce. *Bulletin of entomological research*, pp.406-414.
- Dixon, A.F.G. and Kindlmann, P., 1998. Population dynamics of aphids. In *Insect Populations In theory and in practice*. Springer, Dordrecht. pp. 207-230.
- Dixon, A.F.G., Kindlmann, P., Leps, J. and Holman, J., 1987. Why there are so few species of aphids, especially in the tropics. *The American Naturalist*, pp.580-592.
- Dogimont, C., Bendahmane, A., Chovelon, V. and Boissot, N., 2010. Host plant resistance to aphids in cultivated crops: genetic and molecular bases, and interactions with aphid populations. *Comptes rendus biologiques*, pp.566-573.
- Dominiak, B.C., Links, I.J., Fletcher, M.J., Worsley, P. and McDougall, S., 2009. Detection and spread of currant-lettuce aphid '*Nasonovia ribisnigri*' (Mosley) (Hemiptera: Aphididae) in New South Wales. *General and Applied Entomology: The Journal of the Entomological Society of New South Wales*, pp.27-30.

- Döring, T.F. and Chittka, L., 2007. Visual ecology of aphids—a critical review on the role of colours in host finding. *Arthropod-plant interactions*, pp.3-16.
- Döring, T.F., 2014. How aphids find their host plants, and how they don't. *Annals of Applied Biology*, pp.3-26.
- Eenink, A.H., Groenwold, R. and Dieleman, F.L., 1982. Resistance of lettuce (*Lactuca*) to the leaf aphid *Nasonovia ribis nigri*. 1. Transfer of resistance from *L. virosa* to *L. sativa* by interspecific crosses and selection of resistant breeding lines. *Euphytica*, pp.291-299.
- Ellis, P.R., Tatchell, G.M., Collier, R.H., Chandler, D., Mead, A., Jukes, P.L., Vice, W.E., Parker, W.E. and Wadhams, L.J., 1995. Integrated pest management of aphids on outdoor lettuce crops. In *BCPC SYMP. PROC. 1995*.
- FAO., 2017. World Programme for the Census of Agriculture 2020. FAO Statistical Development Series No. 1. Rome.
- Farnier, K., Dyer, A.G. and Steinbauer, M.J., 2014. Related but not alike: not all Hemiptera are attracted to yellow. *Frontiers in Ecology and Evolution*, p.67.
- Farsi, A., Kocheili, F., Mossadegh, M.S. and Rasekh, A., 2020. Functional and numerical responses of the parasitoid wasp, *Aphidius matricariae* (Hym.: Braconidae) feeding on current lettuce aphid, *Nasonovia ribisnigri* (Hem.: Aphididae). *BioControl in Plant Protection*, pp.113-126.
- Favret, C., 2014. Aphid species file. Available at: <http://aphid.speciesfile.org/HomePage/Aphid/HomePage.aspx> (accessed 08 October 2020).
- Fenton, B., Malloch, G., Navajas, M., Hillier, J. and Birch, A.N.E., 2003. Clonal composition of the peach-potato aphid *Myzus persicae* (Homoptera: Aphididae) in France and Scotland: Comparative analysis with IGS fingerprinting and microsatellite markers. *Annals of Applied Biology*, 142(3), pp.255-267.
- Fereres, A. and Moreno, A., 2009. Behavioural aspects influencing plant virus transmission by homopteran insects. *Virus research*, pp.158-168.
- Feyereisen, R., 2015. Insect P450 inhibitors and insecticides: challenges and opportunities. *Pest management science*, pp.793-800.
- Figuroa, C.C., Fuentes-Contreras, E., Molina-Montenegro, M.A. and Ramírez, C.C., 2018. Biological and genetic features of introduced aphid populations in agroecosystems. *Current opinion in insect science*, pp.63-68.
- Forbes, A.R. and MacKenzie, J.R., 2019. The lettuce aphid, *Nasonovia ribisnigri* (Homoptera: Aphididae) damaging lettuce crops in British Columbia. *Journal of the Entomological Society of British Columbia*, pp.28-31.
- Fournier, V. and Brodeur, J., 2000. Dose-response susceptibility of pest aphids (Homoptera: Aphididae) and their control on hydroponically grown lettuce with the entomopathogenic fungus *Verticillium lecanii*, azadirachtin, and insecticidal soap. *Environmental entomology*, pp.568-578.
- Goggin, F.L., 2007. Plant–aphid interactions: molecular and ecological perspectives. *Current opinion in plant biology*, pp.399-408.

- Gratwick, M., 1992. Lettuce aphids. In *Crop Pests in the UK*. Springer, Dordrecht. pp. 51-55
- Guillemaud, T., Mieuze, L. and Simon, J.C., 2003. Spatial and temporal genetic variability in French populations of the peach–potato aphid, *Myzus persicae*. *Heredity*, pp.143-152.
- Hagen, K.S. and Van Den Bosch, R., 1968. Impact of pathogens, parasites, and predators on aphids. *Annual Review of Entomology*, pp.325-384.
- Hagen, K.S., Bomboch, S. and McMurtry, J.A., 1976. The biology and impact of predators. *Theory and practice of biological control*, pp.93-142.
- Halkett, F., Plantegenest, M., Bonhomme, J. and Simon, J.C., 2008. Gene flow between sexual and facultatively asexual lineages of an aphid species and the maintenance of reproductive mode variation. *Molecular ecology*, pp.2998-3007.
- Hance, T., 1997. Do aphids survive over summer in maize crops in Belgium? A test of the “aphid reservoir” hypothesis. *Biological agriculture & horticulture*, pp.191-201.
- Harrington, R. (1994) Aphid layer. *Antenna*, pp.50-51.
- Hassall, K.L., Dye, A., Potamitis, I. and Bell, J.R. (2021), Resolving the identification of weak-flying insects during flight: a coupling between rigorous data processing and biology. *Agr Forest Entomol*, pp.489-505.
- Hassanali, A., Herren, H., Khan, Z.R., Pickett, J.A. and Woodcock, C.M., 2008. Integrated pest management: the push–pull approach for controlling insect pests and weeds of cereals, and its potential for other agricultural systems including animal husbandry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, pp.611-621.
- Hatchett, J.H. and Gallun, R.L., 1970. Genetics of the ability of the Hessian fly, *Mayetiola destructor*, to survive on wheats having different genes for resistance. *Annals of the Entomological Society of America*, pp.1400-1407.
- Health and Safety Executive, 2021. Regulating pesticides in the UK after Brexit. Available at: <https://www.hse.gov.uk/pesticides/brexit.htm#active>. (Accessed: 21/12/2021).
- Hesketh, H., Alderson, P.G., Pye, B.J. and Pell, J.K., 2008. The development and multiple uses of a standardised bioassay method to select hypocrealean fungi for biological control of aphids. *Biological Control*, pp.242-255.
- Hopper, J.V., Nelson, E.H., Daane, K.M. and Mills, N.J., 2011. Growth, development and consumption by four syrphid species associated with the lettuce aphid, *Nasonovia ribisnigri*, in California. *Biological Control*, pp.271-276.
- Hough, G. L., 2013. *Biology and Control of Currant Lettuce Aphid Nasonovia ribisnigri*, University of Warwick.
- Hughes, S.J., 1976. Sooty moulds. *Mycologia*, pp.693-820.
- International Aphid Genomics Consortium., 2010. Genome sequence of the pea aphid *Acythosiphon pisum*. *PLoS Biology* (online) 8, e100313.

- IRAG., 2021. Insecticide Resistance Action Group. Insecticide resistance status in the UK brassica crops.
- Ismoilov, K., Wang, M., Jalilov, A., Zhang, X., Lu, Z., Saidov, A., Sun, X. and Han, P., 2020. First report using a native lacewing species to control *Tuta absoluta*: From laboratory trials to field assessment. *Insects*, p.286.
- Jarvis, C.E., 1992. The Linnaean plant name typification project. *Botanical journal of the Linnean Society*, pp.503-513.
- Jayasinghe, W.H., Akhter, M.S., Nakahara, K. and Maruthi, M.N., 2021. Effect of aphid biology and morphology on plant virus transmission. *Pest Management Science*.
- Jiang, Y.L., Wu, Y.Q., Qiao, G.X., Duan, Y. and Miao, J., 2012. Community structure and temporal dynamics of arthropod in upper canopy of wheat field: A study with suction trap. *Chinese Journal of Ecology*, pp.2378-2384.
- Kaloshian, I., 2004. Gene-for-gene disease resistance: bridging insect pest and pathogen defense. *Journal of chemical ecology*, pp.2419-2438.
- Kandel, Y., Vulcan, J., Rodriguez, S.D., Moore, E., Chung, H.N., Mitra, S., Cordova, J.J., Martinez, K.J., Moon, A.S., Kulkarni, A. and Etestad, P., 2019. Widespread insecticide resistance in *Aedes aegypti* L. from New Mexico, USA. *PloS one*, p.e0212693.
- Kennedy, J.S., Booth, C.O. and Kershaw, W.J.S., 1961. Host finding by aphids in the field: III. Visual attraction. *Annals of applied Biology*, pp.1-21.
- Kesseli, R., Ochoa, O. and Michelmore, R., 1991. Variation at RFLP loci in *Lactuca* spp. and origin of cultivated lettuce (*L. sativa*). *Genome*, pp.430-436.
- Kift, N.B., Mead, A., Reynolds, K., Sime, S., Barber, M.D., Denholm, I. and Tatchell, G.M., 2004. The impact of insecticide resistance in the currant-lettuce aphid, *Nasonovia ribisnigri*, on pest management in lettuce. *Agricultural and Forest Entomology*, pp.295-309.
- Kindlmann, P., Hullé, M. and Stadler, B., 2007. Timing of dispersal: effect of ants on aphids. *Oecologia*, pp.625-631.
- Kogan, M. (1994). Plant resistance in pest management. In: *Introduction to Insect Pest Management*, 3rd edn, pp. 73-128. Edited by R.L. Metcalf and W.H. Luckmann. New York: John Wiley and Sons.
- Koudela, M. and Petříková, K., 2008. Nutrients content and yield in selected cultivars of leaf lettuce (*Lactuca sativa* L. var. *crispa*). *Horticultural Science*, pp.99-106.
- Leather, S.R. 2014. NIW Insect of the day: Large willow aphid. Harper Adams University. Available at: [https://www.harper-adams.ac.uk/news/202279/niw-insect-of-the-day-large-willow-aphid#:~:text=The%20large%20\(giant\)%20willow%20aphid,nymphs%20weigh%20about%200.25mg](https://www.harper-adams.ac.uk/news/202279/niw-insect-of-the-day-large-willow-aphid#:~:text=The%20large%20(giant)%20willow%20aphid,nymphs%20weigh%20about%200.25mg) (Accessed: 21/12/2021).
- Leberg, P.L., 1992. Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution*, pp.477-494.

- Leman, A., Vijverberg, R. and Messelink, G.J., 2014. Optimizing biological control of mealybugs with lacewing larvae. *IOBC-WPRS Bulletin*, pp.127-131.
- Lindqvist, K., 1960. On the origin of cultivated lettuce. *Hereditas*, pp.319-350.
- Liu, Y.B., 2004. Distribution and population development of *Nasonovia ribisnigri* (Homoptera: Aphididae) in iceberg lettuce. *Journal of Economic Entomology*, pp.883-890.
- Lombaert, E., Carletto, J., Piotte, C., Fauvergue, X., Lecoq, H., Vanlerberghe-Masutti, F. and Lapchin, L., 2009. Response of the melon aphid, *Aphis gossypii*, to host-plant resistance: evidence for high adaptive potential despite low genetic variability. *Entomologia Experimentalis et Applicata*, pp.46-56.
- Lorenz, H., 1998. Biology and generation-order of *Stomaphis quercus* (Lachnidae) living on *Betula pendula* near Berlin, Germany. *Aphids in natural and managed ecosystems*, pp.243-250.
- Loxdale, H.D. and Balog, A., 2018. Aphid specialism as an example of ecological–evolutionary divergence. *Biological Reviews*, pp.642-657.
- Loxdale, H.D., Hardie, J.I.M., Halbert, S.U.S.A.N., Footitt, R.O.B.E.R.T., Kidd, N.A. and Carter, C.I., 1993. The relative importance of short-and long-range movement of flying aphids. *Biological Reviews*, pp.291-311.
- Lv, W., Du, B., Shangguan, X., Zhao, Y., Pan, Y., Zhu, L., He, Y. and He, G., 2014. BAC and RNA sequencing reveal the brown planthopper resistance gene BPH15 in a recombination cold spot that mediates a unique defense mechanism. *BMC genomics*, pp.1-16.
- Mackenzie, J.R., Vernon, R.S. and Szeto, S.Y., 1988. Efficacy and residues of foliar sprays against the lettuce aphid, *Nasonovia ribisnigri* (Homoptera: Aphididae), on crisphead lettuce. *Journal of the Entomological Society of British Columbia*, pp.3-9.
- Martin, J.H., 1983. The identification of common aphid pests of tropical agriculture. *International journal of pest management*, pp.395-411.
- Mathers, T.C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S.T., Baa-Puyoulet, P., Bretaudeau, A., Clavijo, B., Colella, S., Collin, O. and Dalmay, T., 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome biology*, pp.1-20.
- Mathers, T.C., Wouters, R.H., Mugford, S.T., Swarbreck, D., Van Oosterhout, C. and Hogenhout, S.A., 2021. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Molecular biology and evolution*, pp.856-875.
- McCreight, J.D. and Liu, Y.B., 2012. Resistance to lettuce aphid (*Nasonovia ribisnigri*) biotype 0 in wild lettuce accessions PI 491093 and PI 274378. *HortScience*, pp.179-184.
- McCreight, J.D., 2008. Potential sources of genetic resistance in *Lactuca* spp. to the lettuce aphid, *Nasonovia ribisnigri* (Mosely)(Homoptera: Aphididae). *HortScience*, pp.1355-1358.

- Michaud, J.P. and Sloderbeck, P.E., 2005. Russian Wheat Aphid: an introduced pest of small grains in the High Plains. *Kansas State University Agricultural Experiment Station and Cooperative Extension Service, MF-2666*.
- Michelon, N., Pennisi, G., Ohn Myint, N., Orsini, F. and Gianquinto, G., 2020. Strategies for improved Water Use Efficiency (WUE) of field-grown lettuce (*Lactuca sativa* L.) under a semi-arid climate. *Agronomy*, p.668.
- Morales-Hojas, R., 2017. Molecular ecology of insect pests of agricultural importance: the case of aphids. *Ecological Entomology*, pp.18-27.
- Morales-Hojas, R., Gonzalez-Uriarte, A., Alvira Iraizoz, F., Jenkins, T., Alderson, L., Kruger, T., Hall, M.J., Greenslade, A., Shortall, C.R. and Bell, J.R., 2020. Population genetic structure and predominance of cyclical parthenogenesis in the bird cherry-oat aphid *Rhopalosiphum padi* in England. *Evolutionary applications*, pp.1009-1025.
- Morales-Hojas, R., Sun, J., Alvira Iraizoz, F., Tan, X. and Chen, J., 2020. Contrasting population structure and demographic history of cereal aphids in different environmental and agricultural landscapes. *Ecology and evolution*, pp.9647-9662.
- Moreno, A., Nebreda, M., Diaz, B.M., García, M., Salas, F. and Fereres, A., 2007. Temporal and spatial spread of Lettuce mosaic virus in lettuce crops in central Spain: factors involved in Lettuce mosaic virus epidemics. *Annals of applied biology*, pp.351-360.
- N.I. Katis, J.A. Tsitsipis, M. Stevens, G. Powell., 2007. Transmission of Plant Viruses, in: H.F. van Emden, R. Harrington (Eds.), *Aphids as Crop Pests*, CABI, London, pp. 353–377.
- Natwick, E.T. and Lopez, M.I., 2016. Insecticide Efficacy Against Aphids in Leaf Lettuce, 2015. *Arthropod Management Tests*, p.39.
- Nauen, R., Reckmann, U., Thomzik, J. and Thielert, W., 2008. Biological profile of spirotetramat (Movento®)—a new two-way systemic (ambimobile) insecticide against sucking pest species. *Bayer CropScience Journal*, pp.245-278.
- Nebreda, M., Michelena, J.M. and Fereres, A., 2005. Seasonal abundance of aphid species on lettuce crops in Central Spain and identification of their main parasitoids/Jahreszeitliche Abundanz von Blattlausarten auf zentralspanischen Kopfsalatfeldern und Bestimmung ihrer wichtigsten Parasitoiden. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz/Journal of Plant Diseases and Protection*, pp.405-415.
- Nebreda, M., Moreno, A., Pérez, N., Palacios, I., Seco-Fernández, V. and Fereres, A., 2004. Activity of aphids associated with lettuce and broccoli in Spain and their efficiency as vectors of Lettuce mosaic virus. *Virus Research*, pp.83-88.
- Nicol, D., Armstrong, K.F., Wratten, S.D., Walsh, P.J., Straw, N.A., Cameron, C.M., Lahmann, C. and Frampton, C.M., 1998. Genetic diversity of an introduced pest, the green spruce aphid *Elatobium abietinum* (Hemiptera: Aphididae) in New Zealand and the United Kingdom. *Bulletin of Entomological Research*, pp.537-543.
- Northey, T., Venthur, H., De Biasio, F., Chauviac, F.X., Cole, A., Ribeiro, K.A.L., Grossi, G., Falabella, P., Field, L.M., Keep, N.H. and Zhou, J.J., 2016. Crystal structures and binding dynamics

- of odorant-binding protein 3 from two aphid species *Megoura viciae* and *Nasonovia ribisnigri*. *Scientific reports*, pp.1-13.
- Palumbo, J. C., and T. A. Hannan., 2002. Population growth of lettuce aphid, *Nasonovia ribisnigri*, on resistant butter and head lettuce cultivar. Vegetable Report. University of Arizona, College of Agriculture and Life Sciences, Tucson, AZ. ([http:// ag.arizona.edu/pubs/crops/az1292/az1292_2f.pdf](http://ag.arizona.edu/pubs/crops/az1292/az1292_2f.pdf)).
- Palumbo, J.C., Reyes, F.J., Mullis Jr, C.H., Amaya, A., Ledesma, L. and Carey, L., 2001. Neonicotinoids and azadirachtin in lettuce: comparison of application methods for control of lettuce aphids.
- Parker, W.E., Collier, R.H., Ellis, P.R., Mead, A., Chandler, D., Smyth, J.A.B. and Tatchell, G.M. 2002. Matching control options to a pest complex: the integrated pest management of aphids in sequentially-planted crops of outdoor lettuce. *Crop Protection*, 235-248.
- Parry, H.R. 2013. Cereal aphid movement: general principles and simulation modelling. *Mov Ecol.* pp.1-14
- Pascual-Villalobos, M.J., Lacasa, A., Gonzalez, A., Varo, P. and Garcia, M.J., 2006. Effect of flowering plant strips on aphid and syrphid populations in lettuce. *European Journal of agronomy*, pp.182-185.
- Peterson, R.K., Varella, A.C. and Higley, L.G., 2017. *Tolerance: the forgotten child of plant resistance*. *PeerJ*, pp.34-39
- Polat, E., Demir, H. and Erler, F., 2010. Yield and quality criteria in organically and conventionally grown tomatoes in Turkey. *Scientia Agricola*, pp.424-429.
- Powell, W., 1974. Supercooling and the low-temperature survival of the green spruce aphid *Elatobium abietinum*. *Annals of Applied Biology*, pp.27-37.
- Prince, G. and Chandler, D., 2020. Susceptibility of *Myzus persicae*, *Brevicoryne brassicae* and *Nasonovia ribisnigri* to fungal biopesticides in laboratory and field experiments. *Insects*, p.55.
- PSD., 2003. Food and environment protection act 1985 control of pesticide regulations 1986 (S.I. 1986 NO. 1510): Approval for off-label use of an approved pesticide product. Notice of approval number: 1886 of 2003. Pesticide Safety Directorate.
- Purdy, L., 1979. *Sclerotinia sclerotiorum*: history, diseases and symptomatology, host range, geographic distribution, and impact. *Phytopathology*, pp.875-880.
- Quentin, U., Hommes, M. and Basedow, T.H., 1995. Studies ob the biological-control of aphids (Hom, Aphididae) on lettuce in greenhouses. *Journal of Applied Entomology-Seitschrift Fur Angewandte Entomologie*, pp.227-232.
- Raid, R.N., 2004. Lettuce diseases and their management. In *Diseases of Fruits and Vegetables: Volume II* (pp. 121-147). Springer, Dordrecht.
- Rawat, N., Neeraja, C.N., Nair, S. and Bentur, J.S., 2012. Differential gene expression in gall midge susceptible rice genotypes revealed by suppressive subtraction hybridization (SSH) cDNA libraries and microarray analysis. *Rice*, pp.1-15.

- Reinink, K., Dieleman, F. L., 1993. Survey of aphid species on lettuce. *IOBC/WPRS Bull*, pp.56-68.
- Reinink, K., Dieleman, F.L., Jansen, J. and Montenarie, A.M., 1989. Interactions between plant and aphid genotypes in resistance of lettuce to *Myzus persicae* and *Macrosiphum euphorbiae*. *Euphytica*, pp.215-222.
- Reuter, L.L., Toscano, N.C. and Perring, T.M., 1993. Modification of feeding behavior of *Myzus persicae* (Homoptera: Aphididae) by selected compounds. *Environmental entomology*, pp.915-919.
- Reynolds, A.M. and Reynolds, D.R., 2009. Aphid aerial density profiles are consistent with turbulent advection amplifying flight behaviours: abandoning the epithet 'passive'. *Proceedings of the Royal Society B: Biological Sciences*, pp.137-143.
- RIS., 2021. The Insect Survey, Rothamsted Research. Available at: <https://www.rothamsted.ac.uk/insect-survey> (Accessed: 26/10/2021).
- Roach, S.H. and Agee, H.R., 1972. Trap colors: preference of alate aphids. *Environmental Entomology*, pp.797-798.
- Rocca, M. and Messelink, G.J., 2017. Combining lacewings and parasitoids for biological control of foxglove aphids in sweet pepper. *Journal of Applied Entomology*, pp.402-410.
- Rodenburg, C.M. and Basse, H., 1960. *Varieties of lettuce: an international monograph* (No. 3). Instituut voor de Veredeling van Tuinbouwgewassen.
- Rubio-Meléndez, M.E., Barrios-SanMartin, J., Pina-Castro, F.E., Figueroa, C.C. and Ramirez, C.C., 2019. Asexual reproduction of a few genotypes favored the invasion of the cereal aphid *Rhopalosiphum padi* in Chile. pp.66-77.
- Rufingier, C., Schoen, L., Martin, C. and Pasteur, N., 1997. Resistance of *Nasonovia ribisnigri* (Homoptera: Aphididae) to five insecticides. *Journal of Economic Entomology*, pp.1445-1449.
- Rulkens, A.J.H., 1987. *De CGN sla-collectie: inventarisatie, paspoortgegevens en enkele richtlijnen voor de toekomst*. Centre for Genetic Resources.
- Salazar-López, N.J., Aldana-Madrid, M.L., Silveira-Gramont, M.I. and Aguiar, J.L., 2016. Spirotetramat—An alternative for the control of parasitic sucking insects and its fate in the environment. *IntechOpen*, pp.41-54.
- Saldaña, S., Lima, M. and Estay, S., 2007. Northern Atlantic Oscillation effects on the temporal and spatial dynamics of green spruce aphid populations in the UK. *Journal of Animal Ecology*, pp.782-789.
- Sánchez-Bayo, F., Goka, K. and Hayasaka, D., 2016. Contamination of the aquatic environment with neonicotinoids and its implication for ecosystems. *Frontiers in Environmental Science*, p.71.
- Sauer-Kesper CL, Lucia N, Buser H and Vogler U., 2011. Bedeutung und Verbreitung des neuen Biotyps-Nr:1 der Grünen Salatlaus in der Deutschschweiz. *Agrarforschung Schweiz* 10: 462–469.
- Schoonhoven, L.M., Jermy, T. and Van Loon, J.J.A., 1998. Host-plant selection: how to find a host plant. In *Insect-plant biology*. Springer US. pp. 121-153.

- Scorsetti, A.C., Maciá, A., Steinkraus, D.C. and Lastra, C.L., 2010. Prevalence of Pandora neoaphidis (Zygomycetes: Entomophthorales) infecting Nasonovia ribisnigri (Hemiptera: Aphididae) on lettuce crops in Argentina. *Biological Control*, pp.46-50.
- Sentis, A., Bertram, R., Dardenne, N., Ramon-Portugal, F., Louit, I., Le Trionnaire, G., Simon, J.C., Magro, A., Pujol, B., Hemptinne, J.L. and Danchin, E., 2019. Different phenotypic plastic responses to predators observed among aphid lineages specialized on different host plants. *Scientific reports*, pp.1-9.
- Shahid, S., Shi, Y., Yang, C., Li, J., Ali, M.Y., Smagghe, G. and Liu, T.X., 2021. CCHamide2-receptor regulates feeding behavior in the pea aphid, *Acyrtosiphon pisum*. *Peptides*, p.170596.
- Shatilov, M.V., Razin, A.F. and Ivanova, M.I., 2019, November. Analysis of the world lettuce market. In *IOP Conference Series: Earth and Environmental Science*, p.012053.
- Shrestha, G. and Enkegaard, A., 2013. The green lacewing, *Chrysoperla carnea*: preference between lettuce aphids, *Nasonovia ribisnigri*, and western flower thrips, *Frankliniella occidentalis*. *Journal of insect science*, p.94.
- Shrestha, G., Enkegaard, A. and Steenberg, T., 2015. Laboratory and semi-field evaluation of *Beauveria bassiana* (Ascomycota: Hypocreales) against the lettuce aphid, *Nasonovia ribisnigri* (Hemiptera: Aphididae). *Biological Control*, pp.37-45.
- Shrestha, G., Skovgård, H. and Enkegaard, A., 2014. Parasitization of commercially available parasitoid species against the lettuce aphid, *Nasonovia ribisnigri* (Hemiptera: Aphididae). *Environmental entomology*, pp.1535-1541.
- Shrestha, G., Skovgård, H., Reddy, G.V., Steenberg, T. and Enkegaard, A., 2017. Role of the aphid species and their feeding locations in parasitization behavior of *Aphelinus abdominalis*, a parasitoid of the lettuce aphid *Nasonovia ribisnigri*. *Plos one*, p.e0184080.
- Simon, J.C. and Peccoud, J., 2018. Rapid evolution of aphid pests in agricultural environments. *Current opinion in insect science*, pp.17-24.
- Simon, J.C., Baumann, S., Sunnucks, P., Hebert, P.D.N., Pierre, J.S., Le Gallic, J.F. and Dedryver, C.A., 1999. Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, pp.531-545.
- Singh, K.S., Cordeiro, E.M., Troczka, B.J., Pym, A., Mackisack, J., Mathers, T.C., Duarte, A., Legeai, F., Robin, S., Bielza, P. and Burrack, H.J., 2021. Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest *Myzus persicae*. *Communications biology*, pp.1-16.
- Smith, C.M. and Boyko, E.V., 2007. The molecular bases of plant resistance and defense responses to aphid feeding: current status. *Entomologia experimentalis et applicata*, pp.1-16.
- Smith, C.M. and Chuang, W.P., 2014. Plant resistance to aphid feeding: behavioral, physiological, genetic and molecular cues regulate aphid host selection and feeding. *Pest management science*, pp.528-540.
- Smith, H.A. and Chaney, W.E., 2007. A survey of syrphid predators of *Nasonovia ribisnigri* in organic lettuce on the central coast of California. *Journal of Economic Entomology*, pp.39-48.

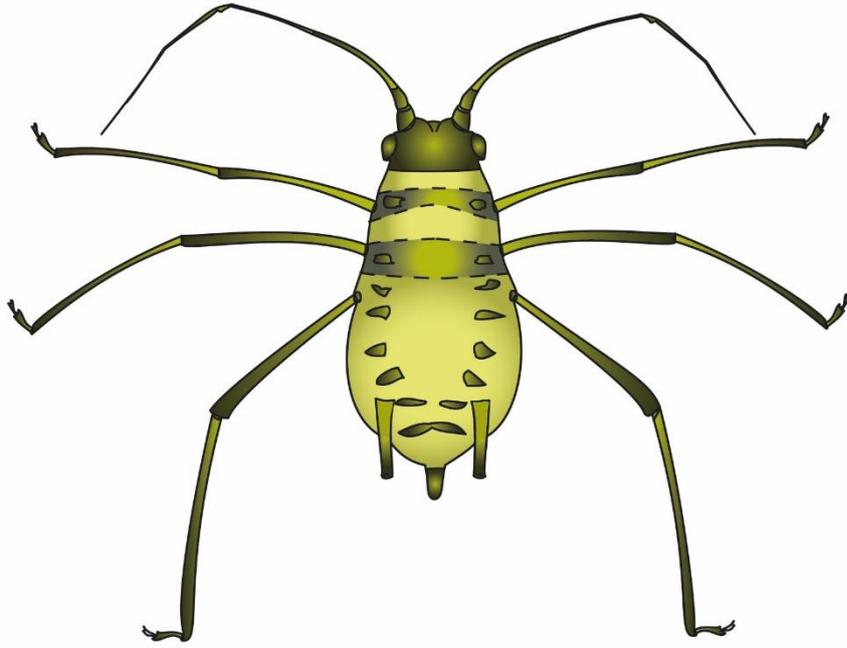
- Sparks, T.C. and Nauen, R., 2015. IRAC: Mode of action classification and insecticide resistance management. *Pesticide biochemistry and physiology*, pp.122-128.
- Srinivasan, D.G. and Brisson, J.A., 2012. Aphids: a model for polyphenism and epigenetics. *Genetics Research International*, 2012.
- Srinivasan, R. and Alvarez, J.M., 2007. Effect of mixed viral infections (Potato virus Y–Potato leafroll virus) on biology and preference of vectors *Myzus persicae* and *Macrosiphum euphorbiae* (Hemiptera: Aphididae). *Journal of economic entomology*, pp.646-655.
- Stout, M.J., 2013. Reevaluating the conceptual framework for applied research on host-plant resistance. *Insect Science*, pp.263-272.
- Stufkens, M.A.W. and Teulon, D.A.J., 2001. Aphids species on potato crops in Canterbury. *New Zealand Plant Protection*, pp.235-239.
- Stufkens, M.A.W., and D.A.J. Teulon., 2003. Distribution and host range of the lettuce aphid in New Zealand and its effect on the lettuce industry. *N. Z. Plant Prot.* pp.27-32.
- Tauber, M.J., Tauber, C.A., Daane, K.M. and Hagen, K.S., 2000. Commercialization of predators: recent lessons from green lacewings (Neuroptera: Chrysopidae: Chrosoperla). *American Entomologist*, pp.26-38.
- Teixeira, L.A. and Andaloro, J.T., 2013. Diamide insecticides: Global efforts to address insect resistance stewardship challenges. *Pesticide Biochemistry and Physiology*, pp.76-78.
- ten Broeke, C.J., Dicke, M. and van Loon, J.J., 2013. Feeding behaviour and performance of different populations of the black currant-lettuce aphid, *Nasonovia ribisnigri*, on resistant and susceptible lettuce. *Entomologia Experimentalis et Applicata*, pp.130-141.
- Thabuis, A.P.P., Teekens, K.C. and Van Herwijnen, Z.O., Rijk Zwaan Zaadteelt en Zaadhandel BV., 2014. *Lettuce that is resistant to the lettuce aphid Nasonovia ribisnigri biotype 1*. U.S. Patent 8,829,280.
- Thompson, R.C., Whitaker, T.W. and Kosar, W.F., 1941. Interspecific genetic relationships in *Lactuca*. *Journal of Agricultural Research*, pp.91-107.
- Valenzuela, I. and Hoffmann, A.A., 2015. Effects of aphid feeding and associated virus injury on grain crops in Australia. *Austral Entomology*, pp.292-305.
- van den Heuvel, J.F., Verbeek, M. and van der Wilk, F., 1994. Endosymbiotic bacteria associated with circulative transmission of potato leafroll virus by *Myzus persicae*. *Journal of General Virology*, pp.2559-2565.
- van der Arend, A.J., 2003. The possibility of *Nasonovia ribisnigri* resistance breaking biotype development due to plant host resistance: a literature study. *Eucarpia leafy vegetables*, pp.75-81.
- van Helden, M. and Tjallingii, W.F., 1993. Tissue localisation of lettuce resistance to the aphid *Nasonovia ribisnigri* using electrical penetration graphs. *Entomologia Experimentalis et Applicata*, pp.269-278.

- van Helden, M., 1995. *The resistance of lettuce to the aphid Nasonovia ribisnigri*. Wageningen University and Research.
- Van Leeuwen, T., Vanholme, B., Van Pottelberge, S., Van Nieuwenhuysse, P., Nauen, R., Tirry, L. and Denholm, I., 2008. Mitochondrial heteroplasmy and the evolution of insecticide resistance: non-Mendelian inheritance in action. *Proceedings of the National Academy of Sciences*, pp.5980-5985.
- Vargas, R.R., Troncoso, A.J., Tapia, D.H., Olivares-Donoso, R. and Niemeyer, H.M., 2005. Behavioural differences during host selection between alate virginoparae of generalist and tobacco-specialist *Myzus persicae*. *Entomologia Experimentalis et Applicata*, pp.43-53.
- Verbeek, M., Dullemans, A.M. and van der Vlugt, R.A., 2017. Aphid transmission of Lettuce necrotic leaf curl virus, a member of a tentative new subgroup within the genus *Torradovirus*. *Virus research*, pp.125-130.
- Vialatte, A., Dedryver, C.A., Simon, J.C., Galman, M. and Plantegenest, M., 2005. Limited genetic exchanges between populations of an insect pest living on uncultivated and related cultivated host plants. *Proceedings of the Royal Society B: Biological Sciences*, pp.1075-1082.
- Walley, P.G., Hough, G., Moore, J.D., Carder, J., Elliott, M., Mead, A., Jones, J., Teakle, G., Barker, G., Buchanan-Wollaston, V. and Hand, P., 2017. Towards new sources of resistance to the currant-lettuce aphid (*Nasonovia ribisnigri*). *Molecular Breeding*, p.4.
- Webster, C.G., Pichon, E., Van Munster, M., Monsion, B., Deshoux, M., Gargani, D., Calevro, F., Jimenez, J., Moreno, A., Krenz, B. and Thompson, J.R., 2018. Identification of plant virus receptor candidates in the stylets of their aphid vectors. *Journal of virology*, pp.18-32.
- Wheeler, T.R., Hadley, P., Morison, J.I.L. and Ellis, R.H., 1993. Effects of temperature on the growth of lettuce (*Lactuca sativa* L.) and the implications for assessing the impacts of potential climate change. *European Journal of Agronomy*, pp.305-311.
- Williams, I.S. and Dixon, A.F., 2007. Life cycles and polymorphism. *Aphids as crop pests*. Wallingford: CAB International, pp.69-85.
- Wiseman B. R., 1994. Plant resistance to insects in integrated pest management. *Plant Disease*, p.927.
- Wongwai, C., Mungcharoen, T. and Tongpool, R., 2014. Environmental Mitigation Possibility via Organic Farming: Lettuce Case Study. *International Journal of Environmental Science and Development*, p.426.
- Woodcock, B.A., Isaac, N.J., Bullock, J.M., Roy, D.B., Garthwaite, D.G., Crowe, A. and Pywell, R.F., 2016. Impacts of neonicotinoid use on long-term population changes in wild bees in England. *Nature Communications*, pp.1-8.
- Young, C.S., Clarkson, J.P., Smith, J.A., Watling, M., Phelps, K. and Whipps, J.M., 2004. Environmental conditions influencing *Sclerotinia sclerotiorum* infection and disease development in lettuce. *Plant Pathology*, pp.387-397.

Yu-Qing, J.Y.L.W., Zhong, D.Y.M.J.G. and Ge-Xia, J.Q., 2011. Research on the effectiveness of suction trapping insects and monitoring populations of wheat aphids. *Chinese Journal of Applied Entomology*.

Zhao, C., Escalante, L.N., Chen, H., Benatti, T.R., Qu, J., Chellapilla, S., Waterhouse, R.M., Wheeler, D., Andersson, M.N., Bao, R. and Batterton, M., 2015. A massive expansion of effector genes underlies gall-formation in the wheat pest *Mayetiola destructor*. *Current Biol*

Chapter 2



CHAPTER 2: Long-term trends in the abundance of *Nasonovia ribisnigri* in the UK; forecasting and improving trapping methodology

2.1 Introduction

2.1.1 Migration biology

The colonisation of crops by horticulturally important aphids occurs due to the seasonal migration of host-alternating aphids moving from their primary host (usually a woody species) to the secondary host (usually a herbaceous plant) during spring (Leather and Walters, 1984). Aphids can undertake both sexual (holocyclic) and asexual (anholocyclic) life cycles (Blackman and Eastop, 1984). Holocyclic reproduction involves sexual reproduction in the autumn with eggs subsequently being laid on the primary host and which hatch the following spring, whereas anholocyclic reproduction means that both nymphs and adults overwinter on a secondary host plant. An advantage to an anholocyclic lifecycle is that it enables both adults and nymphs to overwinter without an egg stage and allows development and migration onto secondary hosts (including crops) to occur earlier in the spring compared to holocyclic reproduction (Peng *et al.*, 2017). Aphids undertaking anholocyclic reproduction are more vulnerable to adverse weather conditions, with both wet and cold conditions affecting survivorship (Bale, 1996; Knight and Bale, 1986). Anholocyclic populations are more frequent in temperate parts of Europe, and this is also true for *N. ribisnigri* which is anholocyclic in Spain and the south of England (Nebreda *et al.*, 2005; Nieto Nafria, 1974). *Nasonovia ribisnigri* is a major pest of outdoor lettuce and is considered a minor pest of currant (*Ribes* spp.) (Adams and Thresh, 1987; Blackman and Eastop, 1984). In the UK during a typical growing season (March-October), outdoor lettuce is predominantly grown in the east of England, particularly in East Anglia (DEFRA, Horticulture Statistics, 2019). A Paradox exists in which *N. ribisnigri* appear infrequently in the RIS suction traps but can appear in very high numbers on outdoor lettuce, contributing to significant crop losses (Nebreda *et al.*, 2004; Reinink and Dieleman, 1993; Sauer-Kesper *et al.*, 2011). As a result, it is important to understand the migration behaviour and autecology of this elusive pest of outdoor lettuce.

Aphids disperse using two different modes; appetitive flight in which local movements of no more than 200 m are undertaken in still to light winds and within the flight boundary layer below 10 m; long distance migration in which aphids enter the convective boundary layer above the flight boundary layer and may travel several hundreds of kilometres (Parry *et al.*, 2013). These unstable air masses

develop following heating of the Earth's surface after sunrise and promote the vertical movement of insects through the atmosphere up to a height 3,000 m (Garratt, 1992; Taylor, 1960). The early work by Johnson (1957) showed how aphids exploit the atmosphere, wherein aphid densities decline with increasing altitude, but Johnson provided little insight into aphid migration behaviour whilst they are airborne. Wainwright *et al.* (2017) showed that above 10 m from the ground, aphids move with the upwardly moving air, contributing their own flight energy, albeit moving at a slower pace than the air itself. Aphids are weak fliers but move around freely during still air at speeds of <0.5 m/sec during daylight and when the winds typically do not exceed 3 m/sec (Hassall *et al.*, 2021). In more challenging weather conditions when wind speeds exceed this threshold, the horizontal (i.e. laminar flow) component, or the effect of strong turbulence, will take aphids downwind or ground them (Bell *et al.*, 2013; Irwin *et al.*, 2007; Kennedy and Thomas, 1974).

2.1.2 Sampling aphids during flight

Due to the horticulturally importance of many aphids, substantial financial investments have been made to understand their migration patterns into crops. Using standardised methods to capture aphids allows for forecasting models to be constructed which help predict the likelihood of an aphid outbreak occurring in crops (Harrington *et al.*, 2007; Klueken *et al.*, 2009; Thackray *et al.*, 2009). The main methods employed for detecting and monitoring aphids are the use of yellow sticky traps, yellow water traps and suction traps (Chu *et al.*, 2004; De Barro, 1991; Lamichhane *et al.*, 2016; Taylor, 1960). Additionally, crop monitoring and crop walking are regularly conducted to identify apterous aphids (Dassonville *et al.*, 2013; Oakley and Walters, 1994; Stufkens and Teulon, 2001) and to detect changes in population size, such as the mid-summer crash (Dixon, 1977; Karley *et al.*, 2003; Leather and Owour, 1996). As discussed in Chapter 1, alate *N. ribisnigri* appear in much lower numbers compared to other horticulturally important pests, which hinders the creation of an accurate forecasting model. Therefore, an accurate forecasting model for *N. ribisnigri* would benefit outdoor lettuce growers.

2.1.3 Forecasting migration

Aphid forecasting is well-covered in the literature and comprises a variety of sampling and modelling techniques to understand the complex pest dynamics involved to predict the likelihood of infestations (Dedryver *et al.*, 2010; Watt, 1983; Way *et al.*, 1981). Some of the metrics used to predict likelihood of infestations are host plant-aphid densities (Al-Eryan and El-Tabbkh, 2004), annual peak numbers (Lima *et al.* 2008); mortality rates (Morgan, 2000), and counting the number of eggs on the winter host (Leather, 1983; Kurppa *et al.*, 1989), have all proven useful. Unfortunately, *N. ribisnigri* is hard to

capture in meaningful numbers and therefore forecasting the likelihood of infestation on outdoor lettuce crops is compromised by sparse data. To date, it is still unclear whether *N. ribisnigri* migrate at different heights from the Rothamsted Insect Survey (RIS) 12.2 m suction traps, only migrate at a local level, or have a low population density and therefore appear infrequently in traps used currently. To address these and other issues, it is necessary to optimise the sampling to increase the number of individuals captured to help improve forecasting models. With the low numbers of *N. ribisnigri* captured in the Rothamsted Insect Survey (RIS) suction traps, other methods will be explored in this chapter.

It is well documented that many insect species are in decline. With the recent influx of scientific literature and subsequent media attention, insect declines are at the forefront of many new policy developments (Althaus *et al.*, 2021; Hall and Steiner, 2019). Indeed, insect declines were apparent and reported over 50 years ago (Ford, 1945; Taylor, 1974). However, more attention and concern are now paid to this topic, and it is most likely a combination of increased scientific studies, public awareness, social media and environmental action groups pushing the need to act now to halt and reverse insect decline. In the UK, many insects have shown declining trends of abundance; native bees (Biesmeijer *et al.*, 2006), carabid beetles (Brooks *et al.*, 2012), butterflies and moths (Conrad *et al.*, 2006; Warren *et al.*, 2001). Interestingly, moth abundance is reported to have increased in arable areas but declined in broadleaf woodlands and other semi-natural areas (Blumgart, 2021). Additionally, trends in abundance, species richness and diversity in regions (north vs south) have been shown to differ. In Great Britain, aphids have shown to be 'broadly stable' over the past 47 years (Bell *et al.*, 2020). However, it was shown that aphids display a high level of interannual volatility, in which overall abundance fluctuated yearly, producing wide confidence intervals (-32%, 22%) (Bell *et al.*, 2015). Another study showed that aphid populations in summer cereal crops have declined over 40 years (Ewald *et al.*, 2015). This result agrees with Bell *et al.* (2020), in which a significant decline in three major cereal aphid pests has been observed. It is unknown whether the abundance trends of *N. ribisnigri* are remaining at a steady state or have increased or declined in the UK. This information would provide further insight into the overall trend of *N. ribisnigri* and feed into future pest management strategies.

This study aimed to: i) understand trends in abundance of *N. ribisnigri* abundance in the UK, and between different regions, over a 56-year period, ii) cross-validate day degrees with first flight of *N. ribisnigri* to understand whether a relationship exists, and determine whether other pest aphids could be used as a proxy to improve forecasting, iii) determine whether the flight boundary layer for *N. ribisnigri* is lower than 12.2m and improve trapping methodology and, iv) use a citizen science approach to capture *N. ribisnigri* on outdoor lettuce plants grown across England.

2.2 Materials and methods

Experiments 1 and 2 draws upon extensive data held by the Rothamsted Insect Survey to produce two models, for annual abundance and first flight respectively. Experiment 3 considers the flight of aphids at or just above crop height between 1-6 m, using a variety of trapping methods. In a bid to collect additional observations throughout the UK, Experiment 4 reports on a social media campaign to generate samples of *N. ribisnigri*.

2.2.1 Experiment 1: *Nasonovia ribisnigri* – developing a model to describe annual abundance trends

2.2.1.1 *Nasonovia ribisnigri* data

The Rothamsted Insect Survey (RIS) has been operating a series of 12.2m suction traps since 1964 (Macaulay *et al.*, 1988). The suction traps collect aerial flying insects by sampling air at a volume of 45 m³ m⁻¹ and daily samples are sent to Rothamsted Research for sorting and identification during the peak aphid flight season (April-October). Trap design and the method of collection are standardised, ensuring catches are directly comparable throughout the time series (Macaulay *et al.*, 1988). Daily counts of *N. ribisnigri* were extracted from the RIS database from 1965 to 2020 (56-years) for every 12.2 m suction trap in the UK and pooled to form an annual summary of site-year counts (Fig. 2.1).

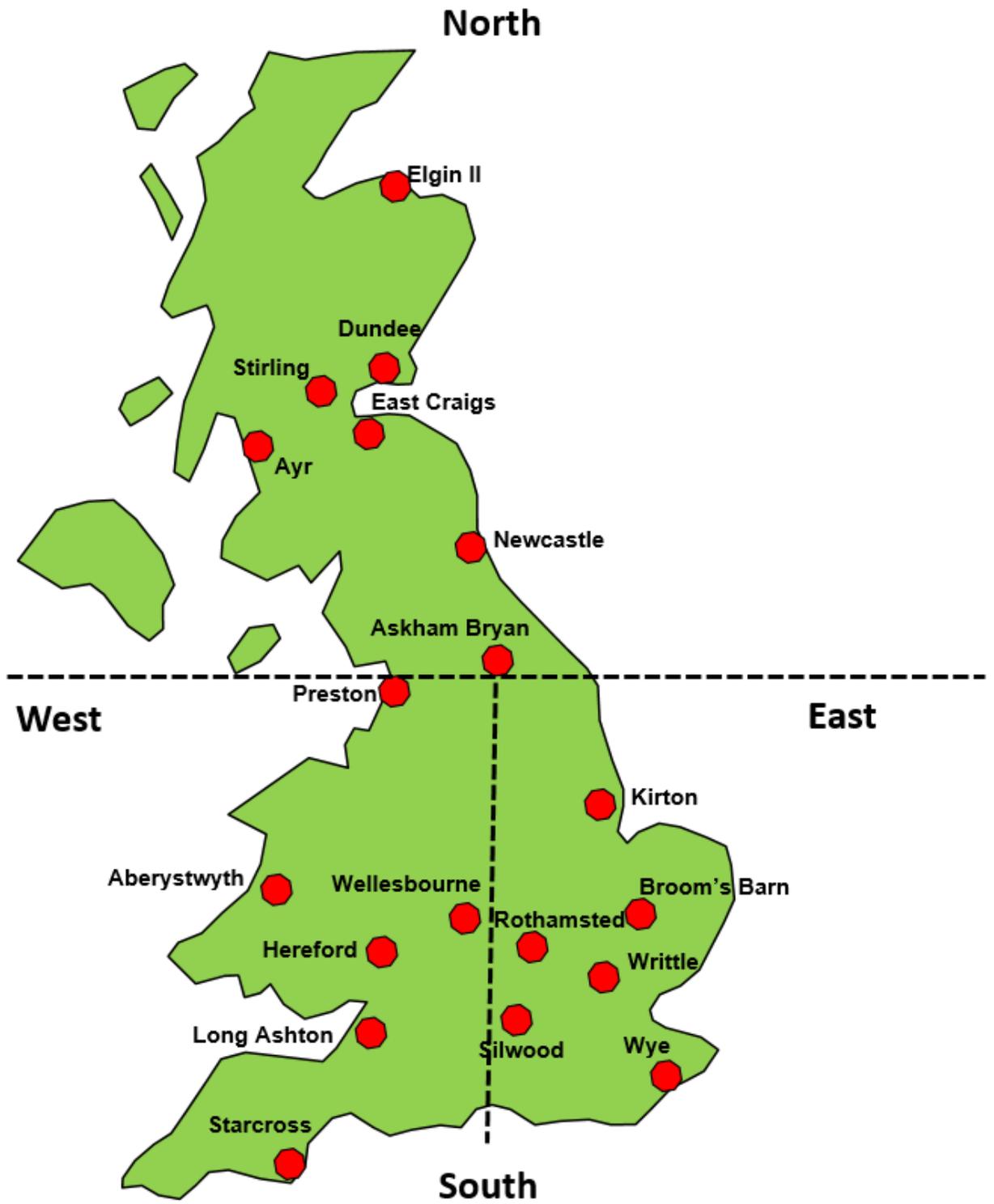


Figure 2.1: The distribution of the RIS 12.2m suction traps used in the analysis of long-term trends. For regional analysis, the sites were split into four regions as follows: North (7-sites), South (12-sites), East (6-sites) and West (6-sites).

2.2.1.2 Long-term annual trend analysis

All statistical analyses were conducted in R (version 4.0.1) using the package 'Poptrend', which was used to fit Generalized Additive Mixed Models (GAMM) of *N. ribisnigri* abundance over time (1965-2020) (Knape, 2016). The Poptrend package integrates smoothing splines and random effects structures from the mgcv package which together provide a prediction of year-to-year variation across sites to reveal an overall network trend and percentage change in abundance with 95% confidence intervals (Knape, 2016). Two model paradigms were created: log-linear GAMM with a random year effect (tempRE = "Temporal Random Effect") and, non-linear GAMM with a random year effect (tempRE = "Temporal Random Effect"). Log-linear models were used to provide estimates of percentage change, as non-linear models are not appropriate, as they are too dependent on the first and last year. Additionally, significant short-term trends were also tested within the long-term trends i.e. significant short-term declines are displayed in orange, with significant increases in green. This allows for important periods of change to be identified within a time series. Years were used as fixed effects. The model assumes a negative binomial error distribution. Site was incorporated as a random effect using the bs = "re" term, which effectively allows variance estimation in abundance to be computed within and among sites. As recommended by Fewster *et al.* (2000), a starting k value of 18 was used for non-linear degrees of freedom, as it should be one third of the length of the time series (56-years). Additional information can be found at <https://github.com/jknape/poptrend>. To check the fit of the linear models, checkFit (mgcv package) was conducted on individual models for all the UK and regions (Appendix 1). In addition, rate of yearly change was conducted on all linear models.

2.2.1.3 Effect of region on long-term abundance of *N. ribisnigri*

To understand whether differences in long-term abundance of *N. ribisnigri* changed depending on region, the UK was divided into four regions (north, south, east and west) (Fig. 2.1). Both log-linear and non-linear models using the Poptrend package (as described above) were used, in which region was a fixed effect with the same starting k value of 18 used for all sites in the UK.

2.2.2 Experiment 2: First flight model for *Nasonovia ribisnigri*

2.2.2.1 Correlation matrix of other aphid species for use as a 'proxy' for *N. ribisnigri*

A correlation matrix was constructed to understand whether other pest aphids could be used as a proxy to forecast the phenology of *N. ribisnigri*. Daily counts of all species (21) covered in the Rothamsted Insect Survey (RIS) bulletin were extracted from the data set for all suction traps running between 1965 and 2018. Due to the low numbers of *N. ribisnigri* in the suction traps, all aphid counts were converted - to 0 for absence and 1 for presence. The package dplyr was incorporated to create a correlation matrix of the daily count data for the RIS pest bulletin species (see appendix for species list) from all 12.2m suction trap sites in the UK. Daily counts were further aggregated into weeks to assess if a better correlation could be found between *N. ribisnigri* and other bulletin aphids. The dplyr package uses a Pearson's correlation coefficient of log daily counts of all RIS pest bulletin aphid species.

2.2.2.2 *Nasonovia ribisnigri* data

For all 12.2m suction traps for which data were available, daily counts of *N. ribisnigri* from 1981 to 2019 were extracted from the RIS database. To reduce false negatives, an upper threshold date (Julian day) after which first flight could not reasonably have occurred was used. Heuristically, an upper threshold of 160 Julian days removed first capture records after the 9th June. This late recording of first flight for some sites is most likely an artefact of the trap-shy species opposed to actual first capture date. Since *N. ribisnigri* appears in such low numbers in the RIS suction traps, daily counts were converted into 0 (presence) or 1 (absence).

2.2.2.3 Driver dataset: Meteorological data

Meteorological data were obtained from the Meteo France dataset at a scale of 1km x 1km grid resolution (available at: <http://www.meteofrance.com/climat>). A data frame ("meteo") was constructed using R which contained: minimum daily temperature, maximum daily temperature, mean daily temperature, average daily wind speed, daily humidity, and daily rainfall for each trap location from 1981-2019 (38 years). This large dataset was reduced to contain the 1km x 1km for suction trap location only ("small_meteo"), with years and dates converted to Julian day. Cumulative degree days were calculated using daily mean temperatures from each suction trap location with initial parameters of a low temperature threshold of 0°C and a high temperature threshold of 35°C. A sensitivity analysis was conducted by further refinement of the cumulative degree day parameters which was incorporated after the initial model to account for *N. ribisnigri* developmental parameters

with temperature. Since *N. ribisnigri* development is severely reduced at low temperatures and ceases below 4.6°C (Hough *et al.*, 2013), this was used as the low temperature threshold for day degree calculation. Development is hindered at high temperatures, with high mortality rates above 28°C (Diaz and Fereres, 2005), and therefore temperatures above this range were considered the upper threshold. Only cumulative degree days was used as a driver in the model.

2.2.2.4 Model for forecasting first flight

A forecasting model was developed in which first flight of *N. ribisnigri* was regressed against cumulative degree days at each site withing the 12.2m RIS suction trap network. The model assumes a straightforward relationship between cumulative degree days from 1st January of every year for each 12m RIS suction trap location. Due to the low numbers of *N. ribisnigri* in the suction traps, a further parameter was applied to omit occasions when first flight occurred after day 160 (early June). The appearance of *N. ribisnigri* after this date was likely to be an artefact of a low catch rate rather than a true first flight date for a given year. This additional parameter helped to reduce false negatives into the model.

A cross-validation process was applied to build the forecasting model whereby a small sample (training set) of the dataset was used to train the model, followed by use of a testing set to validate the forecasting model by estimating the prediction error. Thus, 38 years-worth of data was used to develop the model, with the remainder used to generate the predictions to test the model's performance. All forecasting model construction and analysis was conducted in R v.3.6.1 with 'tidy', 'cartography', 'ggpolt2', 'dplyr', 'mgcv' and 'sp' packages (R Development Core Team, 2020).

2.2.3 Experiment 3: Investigating the height at which *Nasonovia ribisnigri* alates fly during appetitive flight

2.2.3.1 Study site

The aim of this experiment was to understand whether the flight boundary layer for *N. ribisnigri* is lower than 12.2m by using various types of trap, and to improve *N. ribisnigri* captures. The study was conducted at an organically managed farm in Ely, Cambridgeshire, UK (52°21'N, 0°16'E) which is an area of high horticultural intensity, in a relatively homogeneous landscape. The main crops grown are celery, salad vegetables (including lettuce), potato, salad onions and sugar beet (Fig. 2.2).

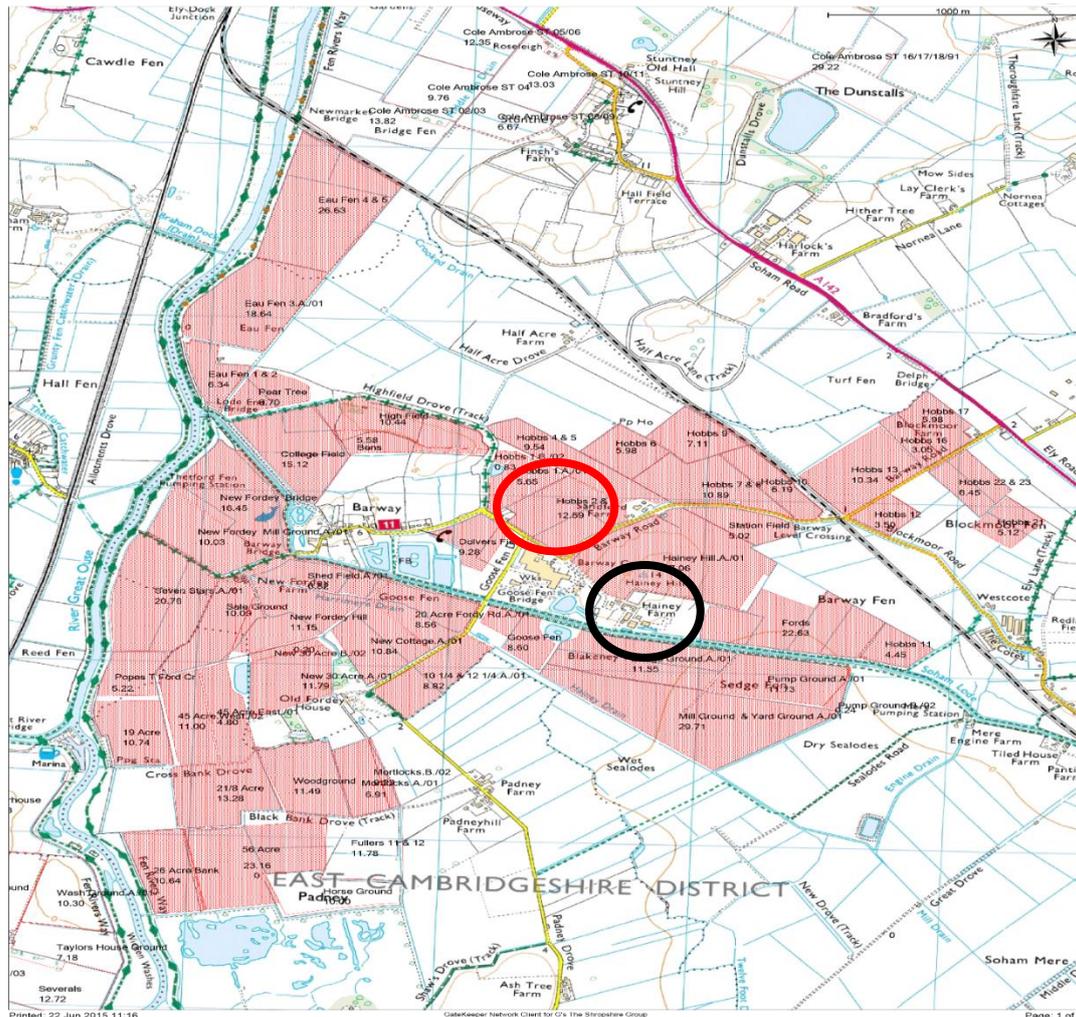


Figure 2.2: Study site map in Ely, Cambridgeshire. Fields are displayed in red with field names. Locations of pop-up suction traps, water traps, yellow sticky traps and Malaise traps are circled in black. Additional water traps, yellow sticky traps and Malaise traps were situated in an outdoor lettuce field (Hobbs field – circled in red) nearby to determine whether traps would have better success at capturing *N. ribisnigri*.

2.2.3.2 Pop-up suction traps

The pop-up suction traps were situated at the main production site which is located centrally and surrounded by fields (Fig. 2.2). Two pop-up suction traps were constructed on site at Rothamsted Research and the design replicated the existing 12.2m suction traps (Fig. 2.3). In brief, the trap consisted of a fan unit with a PVC expansion chamber and latched door, a stainless-steel inner net insert with sample pot attachment and an expander with a bird excluder mounted on top. The fan unit base was completed with a set of wheels and handles for mobility and functionality. The fan unit specifications were as follows: 315mm / 240v / 100w / IP65 Axial Fan. To remove any air flow differences between pop-up suction traps, air flow was calibrated and standardised between traps to 1200m³/hr using a 405i Smart Probe (Testo) (Appendix 1, Fig.1). The aperture diameter of the pop-up

suction traps was 250mm and is comparable to the 12.2 m RIS suction traps (250mm). A minimum height of 2 m and a maximum height of 6 m can be used for the pop-up suction trap.

To compare trap heights below the flight boundary layer (i.e. <10 m, Taylor 1958; Taylor, 1974), three trap heights were assessed: 2 m, 4 m and 6 m. Using trap heights below the 10 m flight boundary layer is more likely to capture appetitive flights of *N. ribisnigri*, with more local movements of no more than 200 m (Parry *et al.*, 2013).

Pop-up suction traps were operated continuously from 16th April – 1st August 2018 and samples were collected twice a week (15 weeks total) for the 2 m and 6 m trap comparison; and between 15th April – 1st June 2019 (7 weeks total) for the 2 m and 4 m trap comparison. Due to the limited number of pop-up suction traps available at the time of the study, 4 m and 6 m traps were not operated alongside each other. A 2 m pop-up suction trap was operating concurrently with either the 4 m or 6 m trap and acted as a control for annual variation in insect abundance.

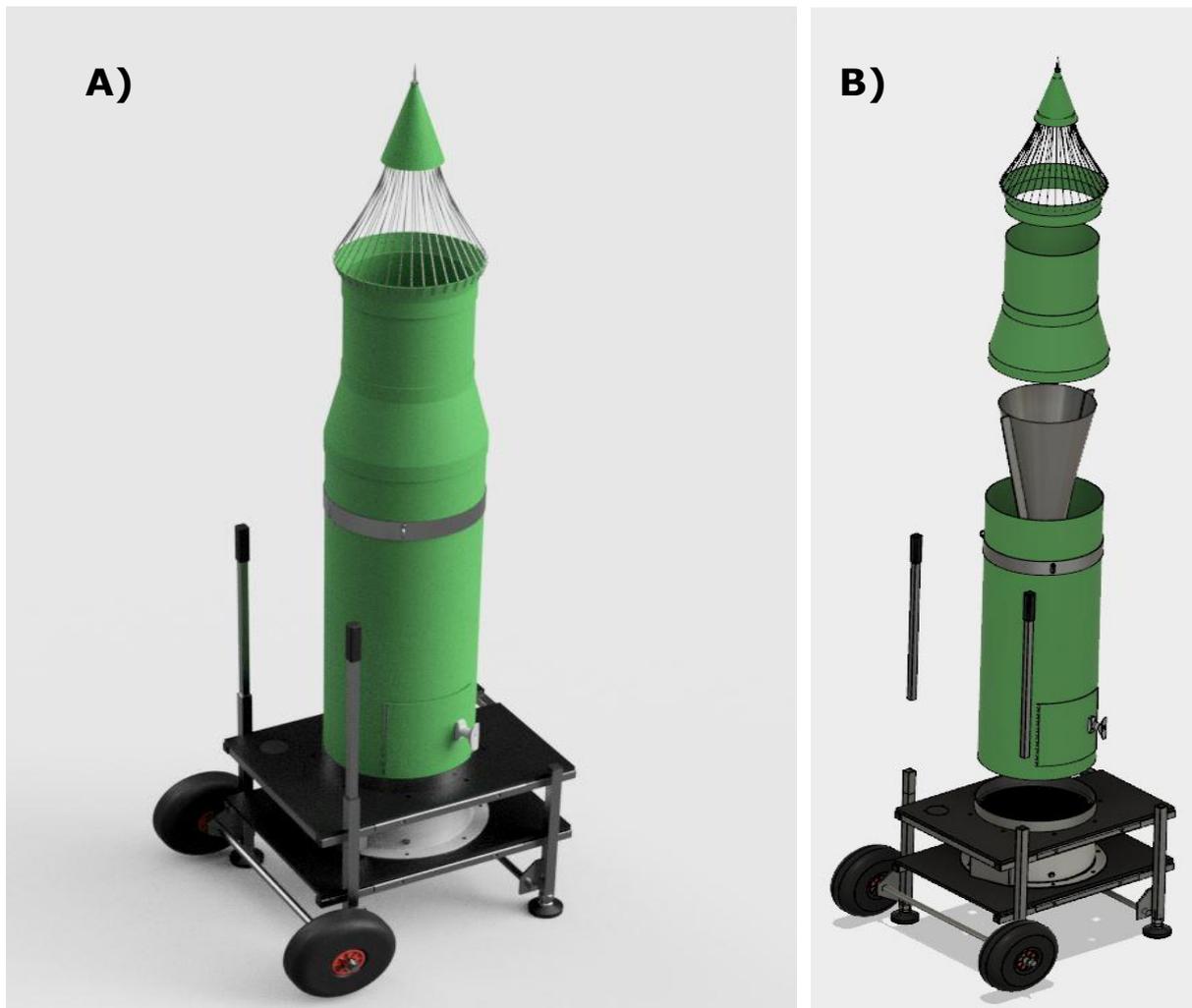


Figure 2.3: A) Pop-up suction trap schematic render showing the complete set up with handles and wheels for mobility. B) Pop-up suction trap schematic breakdown highlighting each constituent part, including the stainless-steel mesh netting insert. Both diagrams were created using Fusion 360 software, courtesy of Gemma Ford (RIS).

Sample collection consisted of emptying the collection bottles (which screw onto the inner mesh insert of the trap) into a fine insect mesh which was then placed into a sample tube with 100% ethanol and labelled accordingly (date and trap type). The collection bottles were immediately rinsed with water and refilled half way with a 70% ethanol solution (70:30 ethanol:water) and attached back onto the stainless-steel inner mesh for the next sample collection. A 70% ethanol solution was used to reduce evaporation and to keep specimens in good condition for identification.

To prevent samples from evaporating and reducing overall sample size, samples were collected twice weekly. For each week, the samples were pooled to provide a single data point, for ease of analysis and interpretation. After identification of the contents, samples were stored in a 95% ethanol solution, with 5% glycerol.

Pop-up suction traps were situated in a neighbouring field to celery and lettuce crop (> 0.5km) where a mains power supply could be reached (Fig. 2.4). Pop-up suction traps were positioned away from any buildings or tall structures, with no obstructions to aerial sampling.



Figure 2.4: The 2m pop-up suction trap (left) and the 6m pop-up suction trap *in situ* at the Cambridgeshire farm. Pop-up traps were 5m apart from one another, with no buildings or trees to obscure insect sampling or affect the boundary layer.

2.2.3.3 Other complementary aphid traps

Water traps

Water traps were used throughout the duration of the study period and consisted of coloured bowls secured onto a 1 m high platform and situated close to field margins. Since aphids are known to locate suitable host plants using visual cues, three colours were used, to cover a wide range of the visible light spectrum (De Barro, 1991; Morris, 2018). These were white, yellow ($\lambda \sim 575$ nm) and green ($\lambda \sim 525$ nm). An aqueous solution with a small amount of scentless detergent was used in all water traps. Each water trap was secured onto a wooden platform 1m above ground. One trap of each colour was situated in proximity to the pop-up suction traps. Another four water traps of each colour were

situated on each edge of Hobbs field (Fig. 2.2). To avoid any potential sampling bias, the positions of the water traps were randomised using a random number generator (Random, 2019).

Yellow sticky traps

Yellow sticky traps were used alongside the water traps. These were positioned close (~ 3 m) to the water traps to allow for comparison of trap 'effectiveness'. Samples were collected twice a week and the traps replaced (Fig. 2.5A). One sticky trap was situated in proximity to the pop-up suction traps and four were positioned on each edge of the outdoor lettuce field (Hobbs, Fig. 2.2).



Figure 2.5: A yellow sticky trap after 3-days of sample collection in an outdoor lettuce field. B) A malaise trap set up *in situ*. Flying insects are intercepted by the fine mesh in the centre of the malaise trap and instinctively move up into the collection bottles located at the top of the trap. It is possible to collect bidirectional samples, to test for differences, as visualised in the image. However, only 1 bottle was used for both directions in the study.

Malaise traps

Consisting of a 1.5 m tall mesh tent structure, Malaise traps passively collect small flying insects by intercepting their flight path (Fig. 2.5B). Insects which land on the mesh undertake phototaxis and move up towards the top of the mesh and into a collection bottle, containing water and a small amount of scentless detergent. A malaise trap was deployed in close proximity to the pop-up suction traps and different coloured water traps (Fig. 2.4). A second Malaise trap was situated at Hobbs field (Fig. 2.2) along with the remaining water traps and yellow sticky traps.

2.2.3.4 Rothamsted Research Insect Survey data

Daily captures of *N. ribisnigri* by the suction trap at Broom's Barn in Suffolk, a site ~20 km away from the study site, were extracted from the Rothamsted Insect Survey (RIS) database. The count data was extracted for the period of the experiment to allow direct comparisons with the pop-up suction traps situated at the study site.

2.2.3.5 Statistical analysis

All statistical analysis was conducted using R v.3.6.1 (R Development Core Team, 2020). Samples from each type of trap were pooled into annual totals of all aphids and of *N. ribisnigri* only. Since not all of the pop-up suction trap heights were tested together during the two-year experiment, separate paired analysis was conducted for the 2 m vs. 6 m (2018) and 2 m vs. 4 m (2019) samples, with the 2 m pop-up suction trap acting as a standard.

2.2.4 Experiment 4: Citizen science and social media

2.2.4.1 Materials and methods

Citizen science is a tool gaining traction in the scientific literature, particularly to tackle data deficiency and produce large scale datasets with minimal outlay of time and resources (Breeze *et al.*, 2020; Brouwer and Hessels, 2019).

To obtain samples of *N. ribisnigri* throughout the UK, citizen science was used to investigate whether this aphid can be detected in a non-horticultural environment. Social media, Twitter, was used to promote 'free lettuce in exchange for aphids' to advertise for volunteers to plant lettuce in their gardens or allotments and send back any aphids found (Appendix 1). The Tweet was sent out in March 2019 and contained information about the study and contact details so that potential volunteers could send an email or direct message with their name and address. The names and addresses of volunteers were stored on a secure server at Rothamsted Research meeting all GDPR requirements. Additionally, the Royal Horticulture Society (RHS) featured the appeal for volunteers in their monthly magazine (The Garden) in April 2019. Once volunteers made contact, they were sent a 'lettuce pack' and lettuce aphid monitoring instructions, which included details for sowing and growing the lettuce plants, what to look out for with regard to aphids, and how to take and send a sample to Rothamsted Research (Appendix 2). All volunteers were asked to grow their lettuce plants outdoors in their garden or allotment to encourage any aphids onto the plants.

The lettuce pack contained 30-40 lettuce seeds, 20 x peat-free compost plugs, 2 x small envelopes, 2 x first class stamps and 2 x sample pots. The sample pots were small Perspex containers (Blackman boxes) measuring 5 x 3.5 x 2 cm (H, W, D), with a 2 cm hole cut out and replaced with mesh to allow for ventilation in transit. The lettuce cultivar 'Auvona' was used and this does not contain resistance to *N. ribisnigri*. This variety is a dark green open-heart Romaine lettuce and produces plenty of leaves. A susceptible variety was used to enable both resistance-breaking and susceptible biotypes to colonise and feed.

2.3 Results

2.3.1 Experiment 1: *Nasonovia ribisnigri* – developing a model to describe annual abundance

During the 56-year study period and across 19 sites, a total of 7,377 *N. ribisnigri* were captured, comprising 409 site-years. The minimum time series length varied between 11-56 years, averaging 37 years. The total number of *N. ribisnigri* per region during the 56-year analysis was 2,181 in the North, 5,196 in the South, 2,776 in the East and 2,420 in the West.

2.3.1.1 Long-term trends in annual abundance of *N. ribisnigri*

The abundance of *N. ribisnigri* in the UK has significantly declined by -76% since 1965 ($P = <0.001$), highlighted by the linear year effect (Fig. 2.6 A). The non-linear year effect model shows significant year by year decline between 1965 to 2008 ($P = <0.001$), with no significant year by year decline in *N. ribisnigri* abundance between 2008 to 2020. The trend between 2008 to 2020 however was still in decline but at a slower rate compared to before 2008 (Fig. 2.6 B).

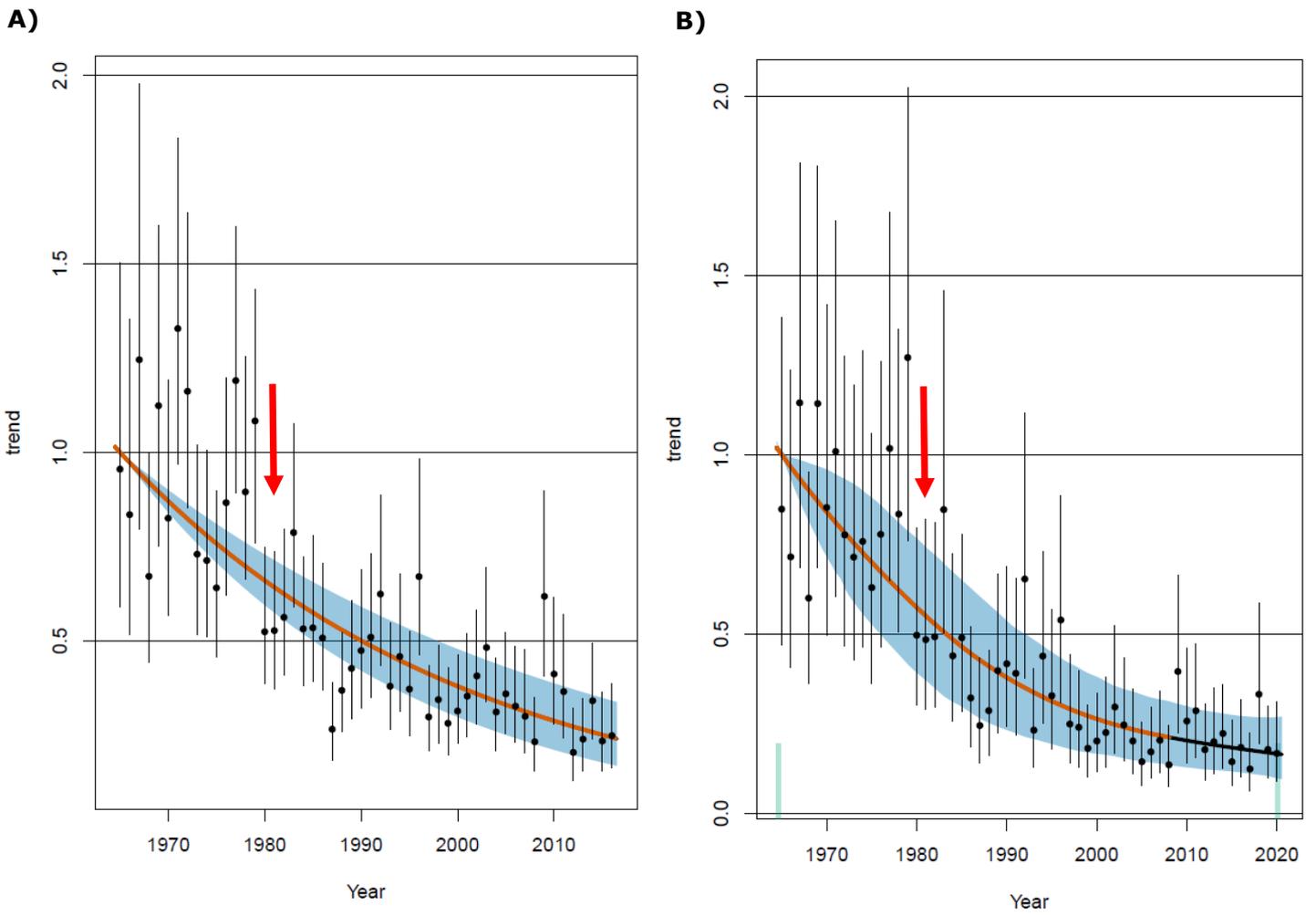


Figure 2.6: Trends of total *N. ribisnigri* abundance in the UK between 1965-2020. The black line represents the trend (i.e. spline), with significant periods of decline indicated in orange. Points and whiskers represent year random effects for each year. A) log-linear year effect of *N. ribisnigri* abundance between 1965-2020. B) non-linear year effect of *N. ribisnigri* abundance between 1965-2020. A 95% confidence interval around the spline for each model is highlighted in blue. All trends are scaled and relative to the starting year of 1965. Red arrow indicates when lettuce cultivars resistant to *N. ribisnigri* were introduced.

2.3.1.2 Total abundance of *N. ribisnigri* in each region

The abundance of *N. ribisnigri* significantly declined in all regions during 1965-2020. The smallest declines were found in the west (-62%, $P = <0.001$), followed by the north (-72%, $P = <0.001$) and south (-78%, $P = <0.001$) and the largest decline was in the east (-85%, $P = <0.001$) (Fig. 2.7). Results are reported for the log-linear year effect (non-linear year effects are reported in the appendix). All regions showed non-linear trends. In the north, *N. ribisnigri* experienced significant year-by-year declines between 1976-1985 with a relatively long period of stability (23-years) until 2008 (Fig. 2.7 B). The south experienced significant year-by-year declines in abundance of *N. ribisnigri* between 1968-2000 and stabilised after this point (Fig. 2.7 D). Abundance in the eastern region underwent significant year-by-year declines from the beginning of the study period (1965) until 2008 (43-years). A sharp drop in abundance of *N. ribisnigri* can be observed during 1980, which was followed by a significant year-by-year downward trend until 2008. The western region showed a significant decline in abundance of *N. ribisnigri* between 1972-1985 (13-years).

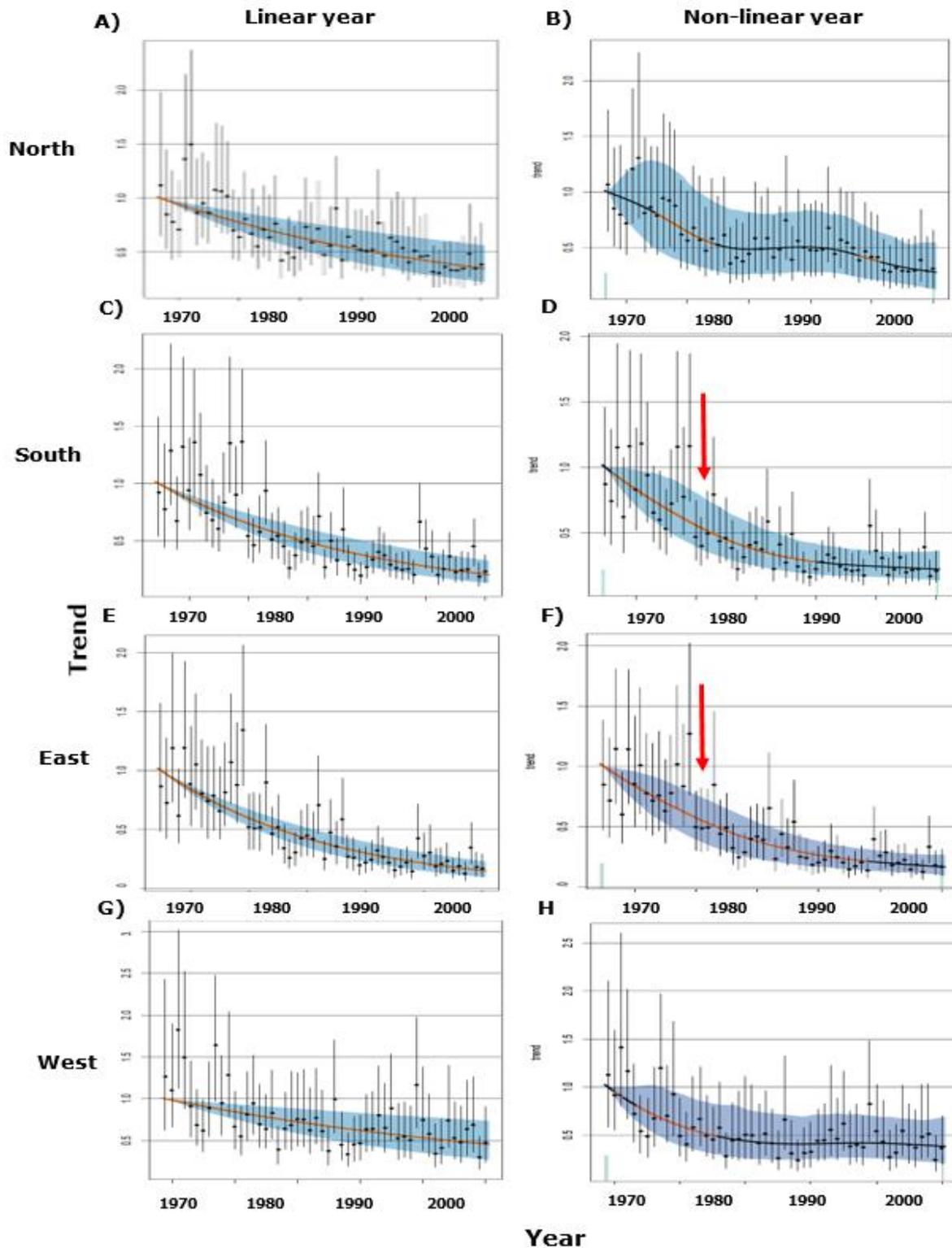


Figure 2.7: Trends in total abundance of *N. ribisnigri* in the UK between 1965-2020 for each region. The black line represents the trend, with significant periods of decline indicated in orange. Points and whiskers represent year random effects for each year. A) **north** log-linear year effect between 1965-2020. B) **north** non-linear year effect between 1965-2020. C) **south** log-linear year effect between 1965-2020. D) **south** non-linear year effect between 1965-2020. E) **east** log-linear year effect between 1965-2020. F) **east** non-linear year effect between 1965-2020. G) **west** log-linear year effect between 1965-2020. H) **west** non-linear year effect between 1965-2020. A 95% confidence interval for each model are highlighted in blue. All trends are scaled and relative to the starting year (1965) at 1. Red arrow indicates when lettuce cultivars with resistance to *N. ribisnigri* were introduced. All x-axis start at the same position and y-axis for the West is slightly larger due to the size of the error bars between 1970 – 1975.

2.3.2 Experiment 2: First flight model for *Nasonovia ribisnigri*

2.3.2.1 Phenology of *Nasonovia ribisnigri* and abundance in RIS suction traps

The currant-lettuce aphid, *N. ribisnigri*, has a typical phenology for a host-alternating (heterecious) species in which 2 distinct peaks in abundance can be observed (Fig. 2.8). Typically, the first appearance of *N. ribisnigri* occurs as early as week 15 (early April), with a large drop in active flying (alates) during August (week 31-36). This decline in abundance is known as the mid-summer crash (Karley *et al.*, 2003). There is a notable drop in alates during week 24 (early June), with a second larger spike in numbers from week 25 (mid-June).

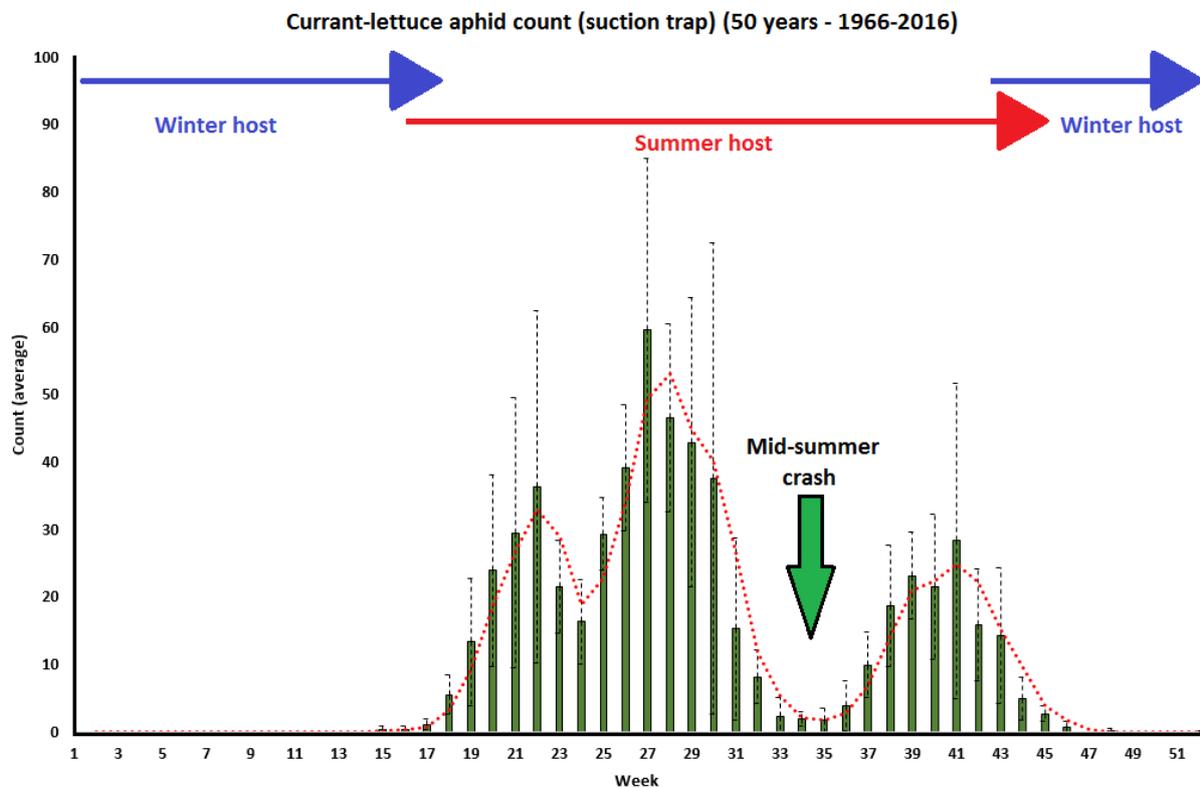


Figure 2.8: Phenology of the currant-lettuce aphid (*N. ribisnigri*) captured and recorded in all active 12.2m RIS suction traps between 1966-2016. Blue arrows indicate time spent on the winter (primary) host and red arrows indicate when *N. ribisnigri* will be on the summer (alternative) host, including lettuce. Two peaks in abundance are apparent and a mid-summer crash is evident between them, which is typical for many heterecious aphids.

2.3.2.2 Correlation matrix of other aphid species for use as a 'proxy' for *Nasonovia ribisnigri*

The Pearson's correlation coefficient on daily log counts of the aphids captured in the RIS suction traps (Appendix 1, table 1) shows that the abundance of *N. ribisnigri* does not correlate well with other aphid species captured (Fig. 2.9). The best correlation identified between aphid species was between *Cavariella aegopodii* (Scopoli, 1963) and *Brachycaudus helichrysi* (Kaltenbach, 1843); and between *Rhopalosiphum padi* and *Rhopalosiphum oxycanthes* (Schrank, 1801).

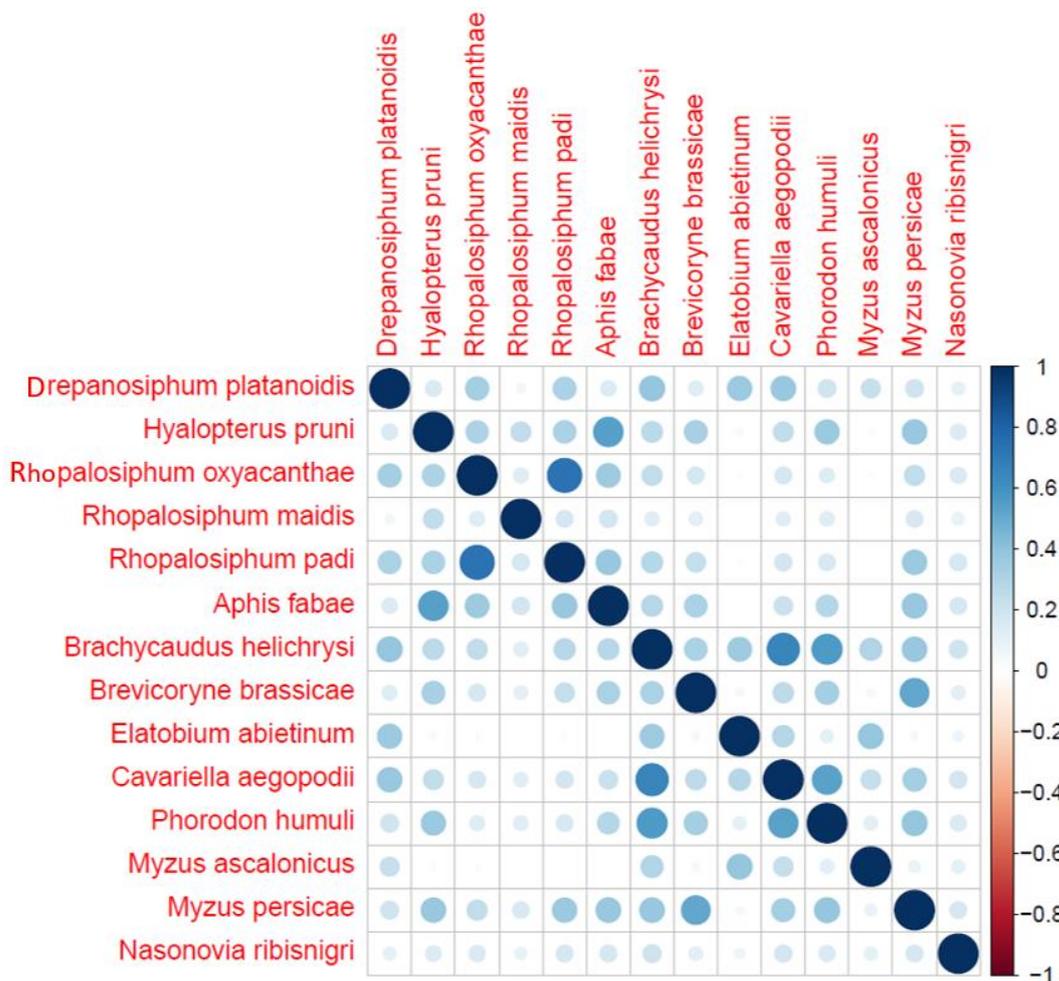


Figure 2.9: The correlation matrix plot of log daily counts of the aphid species captured in the RIS suction traps showing a low correlation with *N. ribisnigri* between 1965 and 2019.

As expected, increasing the daily log counts to weekly log counts improved overall correlations amongst aphids, including *N. ribisnigri*, but the level of correlation was still comparatively low. A weak correlation between *N. ribisnigri* and *Brachycaudus helichrysi* can be observed (Fig. 2.10). Similarly, a weak correlation can be observed between *N. ribisnigri* and *C. aegopodii*, and *Myzus persicae* and *B. helichrysi*. The level of correlation is still comparatively low, especially when compared to correlations between other aphid species such as *Rhopalosiphum padi* and *R. oxycanthae*.

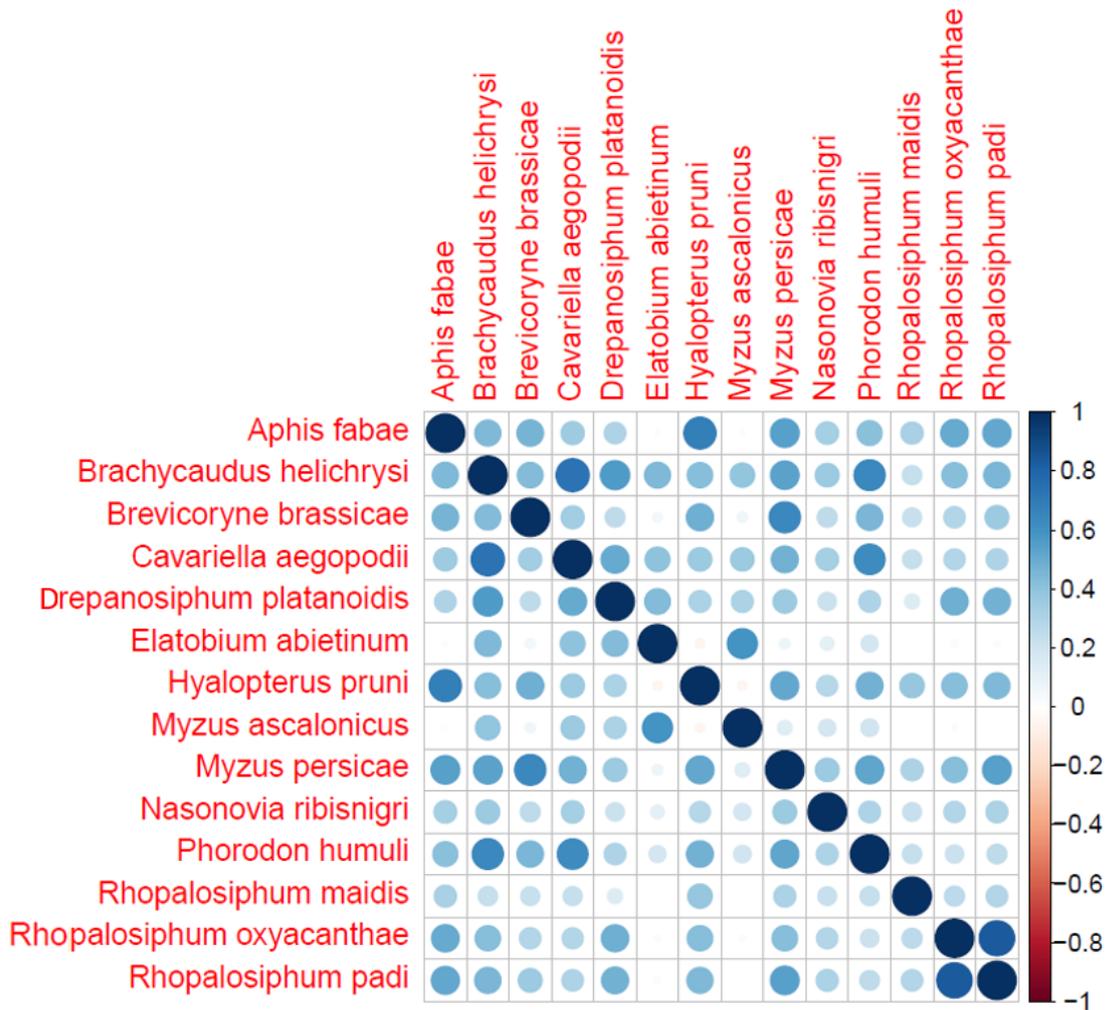


Figure 2.10: The correlation matrix plot of log weekly counts of aphid species captured in the RIS suction traps showing a low correlation with *N. ribisnigri* between 1965 and 2019.

2.3.2.3 First flight of *N. ribisnigri* in the UK

During 1965 to 2019, the traps at a total of 27 sites captured *N. ribisnigri* (Fig. 2.11). First flights of *N. ribisnigri* were recorded as early as week 14 (end of March) and as late as week 40 (end of September). Throughout the 54 years, first flights of *N. ribisnigri* fluctuated both between sites and years. This variation is evident for most sites, particularly at Newcastle, Brooms Barn and Newcastle (Fig. 2.11). As a result of this variation in first flight captures, an upper threshold of 160 days was used, as described in the methods.

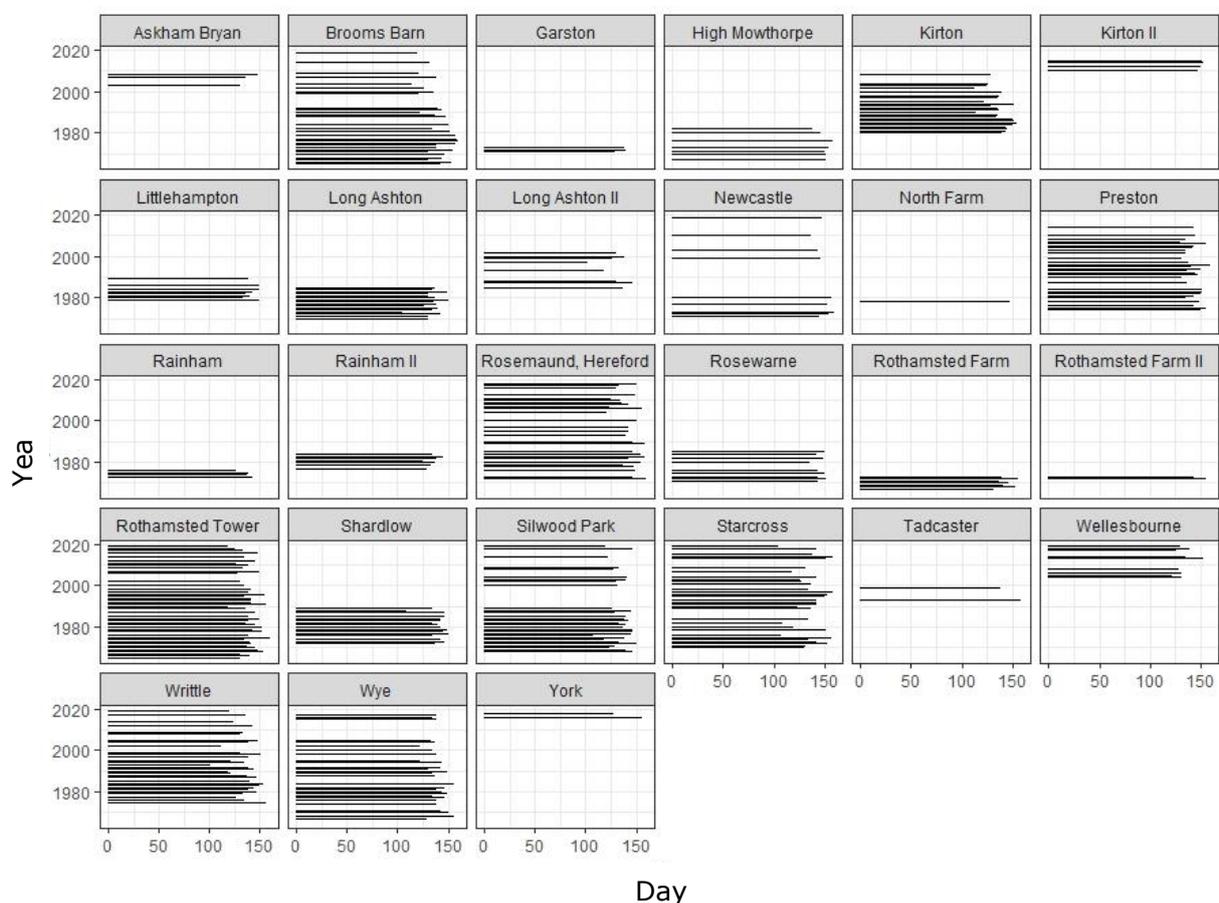


Figure 2.11: Records of the first flight of the currant-lettuce aphids captured between 1965-2019. The x-axis shows the number of days until the first *N. ribisnigri* was captured in a given year, with the y-axis denoting the year of capture. Only the data from Broom's Barn, Kirton, Preston, Hereford, Rothamsted Tower, Starcross, Writtle and Wye were used for correlating first flight with degree days.

During the 54 years that *N. ribisnigri* has been caught by suction traps in the UK, first flight has generally occurred earlier in more recent years. Records of late first flights can be seen in 1972, 1975, 1984 and 1995, when first flights occurred in late May – early June at most of the UK sites (Fig. 2.12). After this point, occurrences of late first flight do not occur for all RIS suction traps in the UK and are limited to a few traps each year. For example, in 1997 and 2014, the first *N. ribisnigri* was caught in the RIS suction trap at Starcross between mid-May and June, whereas, first captures by the other traps were much earlier.

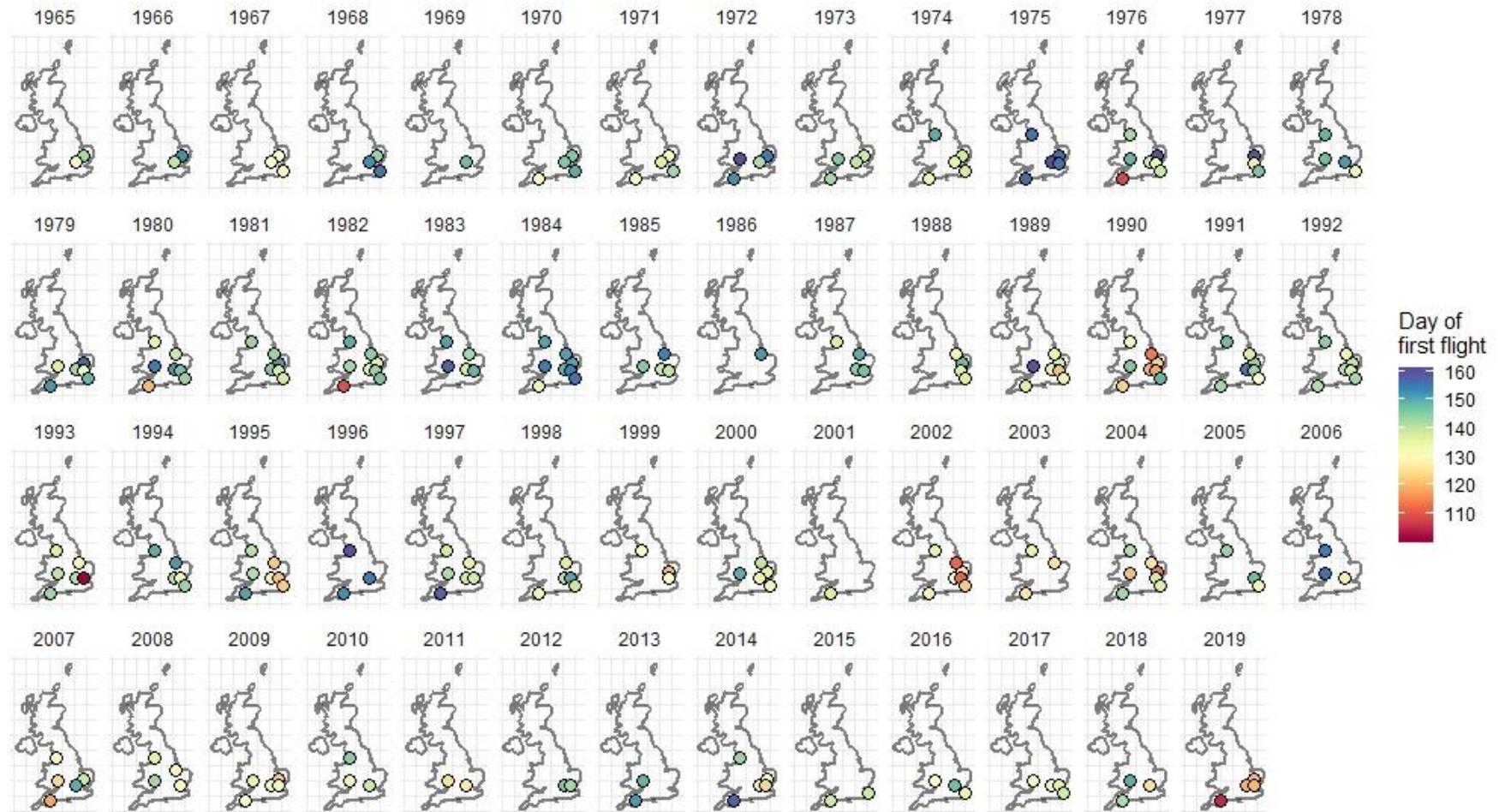


Figure 2.12: Geographical distribution of 12.2m RIS suction traps that caught *N. ribisnigri* in a given year (day of first flight) between 1965-2019 (54 years). Coloured symbols represent the day when *N. ribisnigri* was first captured at that given location.

2.3.2.4 Model for forecasting first flight

After parameter refinements, the cross-validation model accounted for 29.6% of the variation using degree day and first capture of *N. ribisnigri* in the RIS 12.2m suction trap network across 9 sites (Broom’s Barn, Kirton, Preston, Hereford, Rothamsted Tower, Starcross, Writtle and Wye) (Table 2.1). A negative correlation is evident between day of first flight and cumulative degree day, which suggests that cumulative warmer temperatures from the 1st of January enables *N. ribisnigri* to produce alates earlier.

Table 2.1: Forecasting model summary statistics: No. of sites = number of sites included in the model; Day cut-off = number of Julian days of *N. ribisnigri* first flight in a given year (model discards any ‘first flight’ data after this date; T value = T value of the intercept between Julian day and degree day. *Silwood park was removed from model 6-8 and replaced with Kirton.

Model #	No. of sites	Day cut-off	Temp. (min/max)	T value	F-statistic	Multiple R-squared	Deviance explained (%)
1	27	<300	0/35	49.71	52.85	0.144	14.4%
2	27	<180	0/35	49.71	52.85	0.144	14.4%
3	27	<160	0/35	63.49	69.41	0.222	22.2%
4	10	<160	0/35	59.58	75.39	0.284	28.4%
5	9	<160	0/35	58.41	75.59	0.288	28.8%
6	9*	<160	0/35	55.95	71.73	0.296	29.6%
7	9*	<160	0/26	55.95	71.73	0.296	29.6%
8	9*	<160	4.6/26	55.95	71.73	0.296	29.6%

2.3.3 Experiment 3: Investigating the height at which *Nasonovia ribisnigri* alates fly during appetitive flight

Throughout the study period, a total of 35352 insects were counted and identified at least to order but of these, very few *N. ribisnigri* were caught. Of these, 11469 were aphids (32.4 %), which were identified to species where possible, or down to a minimum of genus. Of the 21 aphid species reported in the RIS bulletin, 9906 aphids (28%) were identified. The remaining 1563 aphids were identified but do not currently appear in the RIS bulletin. In 2018, 3697 (27%) of the total catch for the 2 m pop-up suction trap were aphids compared with 2515 (22.6%) caught in the 6 m pop-up suction trap (Fig. 2.13 A); During the sampling period in 2019, 2748 (58.4%) of the total catch in the 2 m pop-up suction trap were aphids whereas 2317 (39.3%) were obtained the 4 m pop-up suction trap (Fig. 2.13 B). The yellow water traps captured the third highest numbers of aphids for both years, followed by white and green water traps. Malaise traps caught the least aphids in both years.

Most of the aphids caught on the yellow sticky traps were either stuck sideways or, in some cases, dorsal side down, making identification using the typical aphid characteristics (siphunculi and abdominal markings) difficult. The bi-weekly sampling of the yellow sticky traps meant that some insects desiccated in warm sunny weather which further impeded identification. Due to these difficulties in identifying aphids on yellow sticky traps, the results were excluded from further analysis.

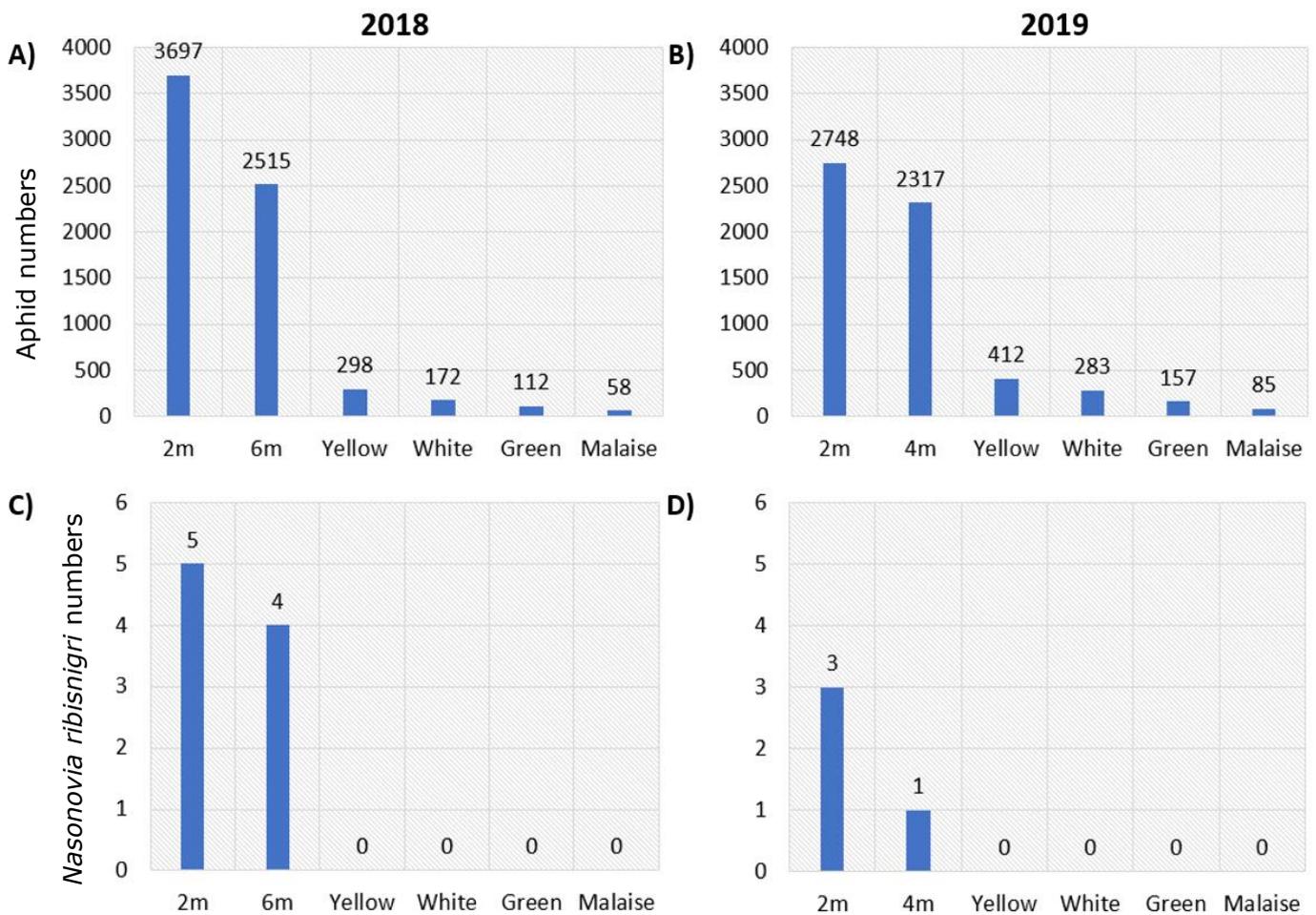


Figure 2.13: Total numbers of aphids caught in the different types of trap during 2018 and 2019 at the Cambridgeshire study site. A) Total numbers of aphids caught during 2018 in the 2 m pop-up suction trap, 6 m pop-up suction trap, yellow water trap, white water trap, green water trap and Malaise trap, B) Total numbers of aphids caught during 2019 in the 2 m pop-up suction trap, 6 m pop-up suction trap, yellow water trap, white water trap, green water trap and Malaise trap, C) Total numbers of *N. ribisnigri* captured during 2018 in all types of trap, D) Total numbers of *N. ribisnigri* caught during 2019 in all types of trap.

In 2018, a total of 9 *N. ribisnigri* were caught in the 2 m and 6 m pop-up suction traps together, with none captured in the water or Malaise traps (Fig. 2.13 C). Similarly, only 4 *N. ribisnigri* were captured in the 2 m and the 4 m pop-up suction traps combined in 2019 (Fig. 2.13 D). During the same periods in 2018 and 2019, the 12.2 m Broom's Barn suction trap caught 5 and 2, *N. ribisnigri* respectively (data not shown). However, since sampling effort was not standardised between traps, the data cannot be directly compared nor statistically analysed.

2.3.4 Experiment 4: Citizen science

In total, 135 lettuce packs were sent to volunteers across the UK. Following the initial Tweet, there was interest from over 60 potential volunteers. With further dissemination of the information, an additional 26 got in touch. In addition, 49 volunteers came from reading the article in the RHS magazine 'The Garden'. Overall, the initial Tweet for lettuce volunteers made 13,627 impressions (the number of people who saw the Tweet on Twitter), with 145 engagements (number of interactions with the Tweet), which included 43 Retweets. The distribution of volunteers was relatively evenly spread across England with low participation in the south-west (1) and south-east (2) (Fig. 2.14). A total of 131 volunteers in England, together with 3 in Wales and 1 in Scotland, were sent lettuce packs in 2019.



Figure 2.14: Distribution of volunteers throughout the UK. Map created using batch geocoding using doogal (Bell, 2021).

Of the 135 volunteers who were sent a lettuce pack in 2019, 99 successfully grew the lettuce outdoors in their garden or allotment. Out of these 99, 36 sent samples of suspected aphids to Rothamsted Research for identification (Fig. 2.15). 15 samples were unfortunately insects other than aphids and included insects from the Order Hemiptera such as planthoppers and leafhoppers (Auchenorrhyncha), adult and larval lacewings (Neuroptera) and ladybird larvae (Coleoptera). The remaining 21 samples were aphids but not *N. ribisnigri*. Identification confirmed that the main species which colonised the volunteers' lettuce plants were the peach-potato aphid (*Myzus persicae*), potato aphid (*Macrosiphum euphorbiae*) and the glasshouse-potato aphid (*Aulacorthum solani*, (Kaltenbach, 1843)). The remaining 63 volunteers saw no aphids or other insects on their lettuce. Finally, the remaining 36 volunteers provided no response after receiving the lettuce pack, despite a follow-up email at the end of the season (October 2019).

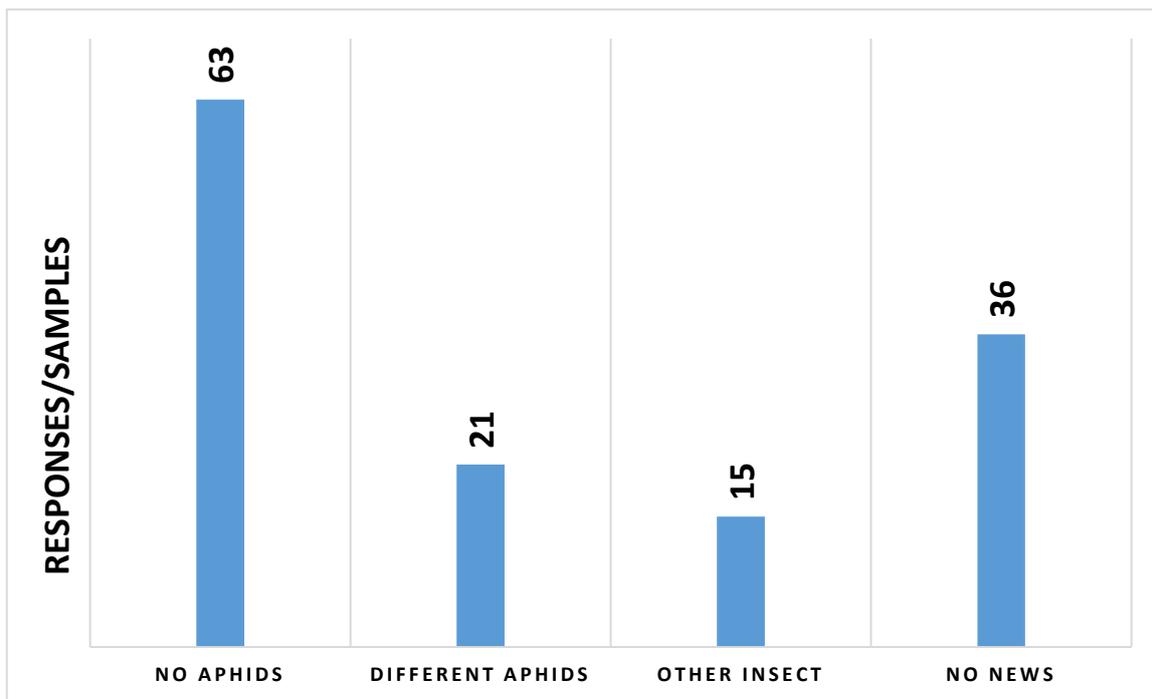


Figure 2.15: Number of responses from volunteers and samples obtained. Aphids found included: the peach-potato aphid *Myzus persicae*, potato aphid *Macrosiphum euphorbiae* and the glasshouse-potato aphid *Aulacorthum solani*. Other insects found included: planthoppers and leafhoppers (Auchenorrhyncha), adult and larvae of lacewings (Neuroptera) and ladybird larvae (Coleoptera).

2.4 Discussion

The decline in the abundance of *Nasonovia ribisnigri* in the UK between 1965 and 2020

Log-linear models showed that the annual count of *N. ribisnigri* has significantly declined by 76% between 1965 and 2020 throughout the UK. The non-linear trend suggested that *N. ribisnigri* abundance has been declining significantly from 1965 to 2010, with non-significant declines thereafter. These results are in contrast to the recent findings of Bell *et al.* (2020) in which aphid numbers in the UK were broadly stable between 1969 and 2016. However, it was shown by that the three most common horticultural pest aphids that appear in the RIS suction traps (*R. padi*, *R. oxtcanthae* and *Sitobion avenae*) are significantly declining in abundance, with the remaining common species showing a stable trend in abundance over time (Bell *et al.*, 2020). It is likely that *R. padi*, *R. oxtcanthae* and *S. avenae* have been affected by the various control measures implemented by growers because of their association with cereals. In the study by Bell *et al.* (2015) aphids were shown to have ‘interannual volatility’, in which aphid abundance fluctuates on a yearly basis, with reports of a non-significant decline in abundance of -7.6%. Interestingly, the interannual volatility is non-existent in *N. ribisnigri* abundance in the UK, with only a few year-to-year fluctuations present, demonstrated by the relatively narrow non-linear year effect 95% upper and lower confidence interval (-84%, -65%). In more recent years, *N. ribisnigri* have been generally very low in abundance when monitored using current trapping methods, compared with other horticulturally important aphids, and either this observed annual volatility in other aphids is not detectable; or does not occur in this species of aphid. However, volatility is bounded by the upper and lower values. With the lower numbers of *N. ribisnigri* in the suction traps, especially in more recent years, the effects of volatility would have low chance to impact the analysis. The present study highlights the benefits of examining individual aphid species separately to understand autecological trends, rather than inference from pooled abundances that do not account for species-specific behaviours (Jowett *et al.*, 2021).

There was a notable difference in the abundance of *N. ribisnigri* between regions. The regional changes in abundance during the 56-year study period show that the non-linear year effects in abundance consistently declined year-by-year in the East, but not in the North or West. The South

showed a significant decline in abundance of *N. ribisnigri* from 1965 to 2008, but this was non-significant thereafter. However, this result is the combination of both the east and west region which are contrastingly different in the abundance trend declines. Both the North and West showed a similar trend; in which a significant decline was observed between 1975 to 1985, and then abundance generally stabilised after this point. Unsurprisingly, the decline in *N. ribisnigri* was greatest in the east (-85%), where most of the outdoor lettuce is cultivated, and therefore where more control measures are applied to reduce aphid numbers. The additional pressures of insecticides and usage of resistant cultivars are most likely to be the cause of this steep decline. However, since no data exists to test this hypothesis, it is merely conjecture. Interestingly, a sharp decline in annual abundance of *N. ribisnigri* can be observed in the east after 1980, which coincides with the introduction of lettuce cultivars containing the Nr-gene. This single gene resistance conferred complete resistance against *N. ribisnigri* until 2007, when resistance breakdown was first reported (ten Broeke *et al.*, 2013). Additionally, neonicotinoid seed treatment was first introduced in the late 1980s which could have further reduced *N. ribisnigri* populations on outdoor lettuce. Since most outdoor lettuce cultivars grown in the UK contain the Nr-gene (~80-90%), is it highly likely that this strong selection pressure further reduced the overall size of populations of *N. ribisnigri*. In the north and west regions, the extra pressure from outdoor lettuce cultivation is low. The decline in abundance of *N. ribisnigri* in these regions is likely to be a result of the changing landscape, and potentially the loss of either winter or alternative summer hosts, or a combination of both. Both the breakdown of resistance in once-resistant lettuce cultivars, and the banning of neonicotinoid seed treatments, could lead to an increase in the abundance of *N. ribisnigri* in future, particularly in the east.

Model for forecasting first flight

By utilising the detailed long-term data of the RIS suction trap network, it is possible to study past spatial and temporal phenological patterns of aphids throughout the UK. This network of traps is a vital source of information for growers and the data can be used to provide a pre-season forecast for some pest species such as *M. persicae*, *R. padi* and *Phorodon humuli* (Schrank, 1801) (Fabre *et al.*, 2010; Heathcote *et al.*, 1969; Tatchell, 1990). However, the precision and accuracy of those relationships may very much depend on within field landscape and environmental factors in addition to aphid biology. The RIS can forecast likely aphid pest infestations by regressing first flight against weather conditions to predict migration events, particularly mean air temperature in January and February (Bell *et al.*, 2015; Harrington *et al.*, 1991; Zhou *et al.*, 1995), or be driven by first flights and numbers to the 17th June to predict virus incidence in sugar beet using a semi-mechanistic model (Werker *et al.*, 1999). Unfortunately, *N. ribisnigri* appears in very low numbers in the RIS suction trap network which makes predicting migration events into lettuce difficult and inaccurate.

Due to the low numbers of *N. ribisnigri*, introduces uncertainty as to whether records of first flight are accurate. However, with parameter refinement and model adjustments to eliminate likely false negatives, the cross-validation between *N. ribisnigri* first flight and degree days is apparent but weak. The first flight of *N. ribisnigri* generally appeared sooner in more recent years compared to before 1990. This finding conforms to previous literature, highlighting that increases in degree days expects to advance the appearance of aphids earlier in the year (Harrington *et al.*, 2007). The earlier appearance of *N. ribisnigri* in the RIS suction traps could increase the potential for them to colonise their secondary host plants (including outdoor lettuce) sooner. Regression between cumulative degree day and alate production is well documented throughout the literature (Harrington *et al.*, 2007; Howling *et al.*, 1993; Walters and Dewar. 1986). Such relationships have shown to exist between other pest aphid species (peach-potato aphid *Myzus persicae*, English grain aphid *Sitobion avenae* and the bird cherry-oat aphid *Rhopalosiphum padi*) (Hansen, 2006; Holloway *et al.*, 2018). However, these particular aphids are frequently caught in the 12.2 m suction traps, and in very high numbers during peak flying periods (>500 individuals in a single trapping event); whereas *N. ribisnigri* does not occur

at these levels (<10 individuals in a single trapping event). This severely hinders the creation of an accurate forecasting model, as false negatives of first flight are apparent throughout the dataset. Although the cross-validation accounts for $\approx 30\%$ (with parameter refinement) of the deviance explained between date of first capture and degree day, the outlook in increasing this level of explanation is promising. Alternative models, such as random forests or artificial neural network, both machine learning approaches, may well improve the prediction of complex relationships between aphid species and their meteorological drivers (Holloway *et al.*, 2018). Incorporating additional environmental variables could help improve the accuracy of predicting first flight.

There is an evident gap in the knowledge between the appearance of *N. ribisnigri* in the RIS suction traps and the first appearance of *N. ribisnigri* on outdoor lettuce. With the low numbers of *N. ribisnigri* in current trapping methodology, and the delay in of outdoor lettuce colonisation, it could be beneficial to investigate other environmental variables – such as monitoring secondary alternative host plants of *N. ribisnigri*. Instead of using the low counts of *N. ribisnigri* in current trapping methodology, using numbers of *N. ribisnigri* on outdoor lettuce could provide more useful information. It has been shown that low counts of *N. ribisnigri* have been observed in water traps, with high numbers found on outdoor lettuce (Collier *et al.*, 2002). These additional variables could be used to predict the likelihood of infestation on outdoor lettuce as opposed to first flight, as first flight is not as important to a grower. Interestingly, in the first peak of *N. ribisnigri* activity, before the mid-summer crash, there is a notable drop in alates during week 24 (early June), with a second larger spike in numbers from week 25 (mid-June). Despite their comparatively low abundance, *N. ribisnigri* followed the same trend regarding the mid-summer crash, in which a notable decline in alates can be observed during August. This is most likely a result of a reduction in host plant quality (Awmack and Leather, 2002; Dixon, 1977; Leather and Owuor, 1996) and a combination of natural enemies such as coccinellids, chrysopids and entomopathic fungi (Karley *et al.*, 2003; Karley *et al.*, 2004; Southwood & Comins, 1976). This knowledge highlights the need to understand the cryptic autecology of *N. ribisnigri* to create an improved forecasting model.

Low dispersal of *Nasonovia ribisnigri*

The present study confirmed that pop-up suction trap at heights considerably lower than 12.2 m did not increase captures of *N. ribisnigri*. Similarly, low numbers were found in all the suction traps, including the RIS suction trap situated at Broom's Barn (~20 km away), supporting this conclusion. The results of the current study suggest that *N. ribisnigri* do not migrate at lower heights and instead, that their overall population density is much lower than other aphid species. Consequently, *N. ribisnigri* is not using a particular dispersal mode (appetitive or long distance) below the flight boundary layer and hence, suction traps are a fair reflection of the migration densities of this aphid. It is unlikely that they migrate in higher densities at heights exceeding 12.2 m, as it has been shown that insect numbers generally decrease with increasing height above the ground (Taylor, 1974). It can be concluded that *N. ribisnigri* is truly a rare aphid.

The pop-up suction traps appear to be more effective at capturing flying insects than the other types of trap used in this study. However, since sampling effort was not standardised the trap captures cannot be directly compared. All the pop-up suction traps at the study site (2 m, 4 m and 6 m) collected far greater numbers of aphids during both the study periods of 2018 and 2019. After the pop-up suction traps, the yellow water traps caught the most aphids during both years. However, this was less than 10% of the total catch of aphids compared to the 2 m pop-up suction trap in 2018 and 15% in 2019. This large reduction in the total number of aphids caught highlights the effectiveness of the pop-up suction traps. The comparatively limited samples collected by water traps may have serious consequences when drawing inferences about the population density of aphids and potential outbreaks, if this technique is used in isolation.

Previous work has shown that few alate *N. ribisnigri* were caught using water traps despite the high numbers of both apterous and alates present on insecticide-free outdoor lettuce (Collier *et al.*, 2002). Interestingly, overall numbers of *N. ribisnigri* caught in the RIS traps between 1965 and 2020 in the East and West region are comparatively similar, with 2,776 and 2,420, respectively. It would be reasonable to conclude that, since most of the outdoor lettuce is currently cultivated in the East, this provides more abundant secondary hosts and resources resulting in *N. ribisnigri* population densities

exceeding those in the West. This large increase of *N. ribisnigri* populations during outdoor lettuce outbreaks, however, is not detected in the RIS suction trap networks nor in pop-up suction traps or other traps in proximity to the outbreaks. Additionally, there were two separate outbreaks of *N. ribisnigri* at the Hainey (organic) farm during the pop-up trap height experiment (Experiment 3). Firstly during 2018 in August, and secondly in 2019 during early September. The numbers of *N. ribisnigri* in the field were very high, including numerous alates, but very few *N. ribisnigri* were captured in any of the traps (situated <500 m away) during the outbreaks. Additionally, these outbreaks are usually isolated to one or two fields (Norman, personal communication) and would suggest dispersal is relatively isolated when on outdoor lettuce. Although there were no traps situated closer to these fields, the alates produced during an outbreak would appear to be displaying appetitive flight behaviour and moving within the flight boundary layer below 10 m and migrating less than 200 m (Parry *et al.*, 2013). These results reinforce the conclusion that *N. ribisnigri* exists in low population densities during most of their life cycle and, when large outbreaks do occur on outdoor lettuce, they are short-lived and disperse at low levels. However, further semi-field experiments would need to be conducted to test this hypothesis.

A distinct advantage of the pop-up suction traps compared to yellow sticky traps and yellow water traps is that they passively sample the air column at a given height and operate 24 hours. Both the sticky and water traps are activity traps and attract aphids using colour. As a result, during overcast days, the lower light levels will affect trap effectiveness (Kennedy *et al.*, 1961). With the use of a passive trap such as the pop-up suction traps, light levels do not affect catch rate as they do not attract aphids, but moreover sample the air column in an area. The pop-up suction traps, however, are more of a specialist item and therefore the initial cost far exceeds the cost of sticky or water traps and they require a mains power supply. Additionally, having multiple pop-up suction traps in the same locality would not be practical nor feasible. The aphid monitoring scheme of yellow water traps that are incorporated throughout parts of the UK to monitor for potential virus-vectoring aphids, would not be practicable if using pop-up suction traps instead (Fera, 2021). Taylor (1974) estimated that the RIS 12.2 m suction traps provide a reasonable representation of insects within an 80 km radius. Since appetitive flight occurs within the flight boundary layer below 10 m, the pop-up suction traps are likely

to be monitoring more localised aphid movement (Parry *et al.*, 2013). Therefore, the lower pop-up suction traps likely represent a smaller radius and consequently provide a finer scale. From an IPM perspective, this may facilitate more targeted control and reduce the need for unnecessary pesticide applications (Klueken *et al.*, 2009). For example, reducing chemical usage could aid in dampening the speed at which insecticide resistance develops and increase the number of non-target beneficial insects and natural enemies that contribute to biological control (Lewis, 1981; Merrill *et al.*, 2010). Furthermore, the 2 m pop-up suction traps are transportable, allowing for easy deployment. Additional research into the effective range of pop-up traps would need to be conducted to determine the representative radius. With more restrictions in horticulture on the repertoire of chemicals permitted and their applications, understanding current and future pest pressures is of paramount importance.

Citizen Science

The engagement of citizen scientists can advance scientific understanding through data collection across large areas, improve public awareness of important issues such as environmental impacts and conservation and connects members of the general public to science (Frensley, 2017; Gaba and Bretagolle, 2020; Gregory *et al.*, 2004). No *N. ribisnigri* were found on any outdoor grown lettuce by the volunteers engaged in the study. This suggests that *N. ribisnigri* is not abundant in non-horticultural areas and thus further supports the suggestion that suction trap captures are a true representation of their low migration densities. Since lettuce is a reasonably fast-growing crop, many volunteers were advised to sequentially sow seeds from April to provide ample opportunity for aphid colonisation and provide lettuce for consumption throughout the summer. Additionally, this provided the best chance to obtain samples of *N. ribisnigri* as they usually infest lettuce mid to late summer. Despite this effort to have lettuce growing throughout April-September, no *N. ribisnigri* were found on the lettuce samples sent in for formal identification. Interestingly, the infestations of *N. ribisnigri* reported in 2019 were very low, with only a few isolated outbreaks in late August and early September in the Cambridgeshire region. It can be assumed that 2019 was generally a low-pressure year for *N.*

ribisnigri, so it is likely that sparse numbers of *N. ribisnigri* limited dispersion, which could have reduced the numbers seen on lettuce throughout England. This could be a result of either a deleterious spring affecting aphid numbers (holocyclic) or a low presence of alternative secondary hosts reducing populations; or a combination of both. Lettuce is predominantly grown in the east of England, with a large majority grown in East Anglia. Despite the comparatively small areas of lettuce grown by volunteers, it is interesting to note that the volunteers within these 'high risk' areas (East of England) did not get infestations of *N. ribisnigri* on lettuce nor other aphid outbreaks in general. Future citizen science work could improve our understanding of pest reservoirs and movements within high risk areas relative to low risk areas.

The overall response rate (73%) from the citizen science experiment was extremely high, given the investment in time and effort from the volunteers (Bryman, 2013; Ruane, 2015). A similar study involving volunteers who had agreed to participate in a scientific survey with considerable time investment resulted in a response rate of 39% (Land-Zandtra *et al.*, 2016). For web surveys, 39% is considered a reasonable response rate (Bryman 2013; Cook *et al.*, 2000). Therefore, the method of advertisement and inducement to participate was successful in engaging citizen scientists. Citizen scientists have been shown to be more willing to participate and keep up participation if they have an interest in the topic, and feel their contribution has value (Eveleigh *et al.*, 2014; Frensley, 2017; Fischer *et al.*, 2021). Twitter, through which the first appeal for volunteers was advertised, has a large scientific community. Additionally, the Royal Horticultural Society (RHS), has a large community of passionate gardeners. Many volunteers sent regular updates throughout the study period, including photos of their progress, without any input from the researcher (Fig. 2.16). The overall enthusiasm and interest of the volunteers was high, with a lot of positive feedback received.



Figure 2.16: Number of responses from volunteers and samples obtained. Aphids found included: the peach-potato aphid *Myzus persicae*, potato aphid *Macrosiphum euphorbiae* and the glasshouse-potato aphid *Aulacorthum solani*. Other insects found included: planthoppers and leafhoppers (Auchenorrhyncha), adult and larvae of lacewings (Neuroptera) and ladybird larvae (Coleoptera).

It has been shown that both the length of a questionnaire, and associated information, heavily influences response rate (Dillman 1991). Additional time taken for participants is an important factor in response rate, with more time taken resulting in fewer responses (Cui, 2003; Bryman, 2015). The lettuce pack was designed to be easy and quick to set up and the accompanying material was minimal and succinct. To help with the overall uptake and ease of the project, lettuce packs were put together to include everything required to grow lettuce, including sample pots, envelopes and stamps or easy sample collection and sending. The feedback received on both the lettuce packs and information guides was positive, with many volunteers enquiring whether the study will continue the following year. It is likely the combination of the appeal of free lettuce, participation in a relatively straightforward scientific experiment and lack of a follow-up survey which led to the higher than expected response rate.

In conclusion, the lettuce packs were well received by volunteers and uptake was positive. The response rate from volunteers was surprisingly high compared to other similar studies. Continued engagement throughout the season was also encouraging, with a good number of participants sending at least one sample of suspected aphids on their lettuce for formal identification. Though no *N. ribisnigri* were identified on any samples received, this was likely to be a result of a low-pressure year and low dispersal of *N. ribisnigri*, and as such, demonstrates that citizen science is a useful tool to explore aphid pest distributions in future studies.

Acknowledgements

I am grateful to collaborators and staff who have contributed data from the Rothamsted Insect Survey, a National Capability, is funded by the Biotechnology and Biological Sciences Research Council under the Core Capability Grant BBS/E/C/000J0200. I would also like to thank all the volunteers who made this study possible and demonstrated enthusiasm from the start. Additionally, I would like to thank Gemma Ford for producing the pop-up suction trap schematic diagram using Fusion 360 software.

2.5 References

- Adams, A.N. and Thresh, J.M., 1987. Reversion of black currant. *Virus Diseases of Small Fruits.*, pp.133-136.
- Al-Eryan, M.A.S. and El-Tabbakh, S.S., 2004. Forecasting yield of corn, *Zea mays* infested with corn leaf aphid, *Rhopalosiphum maidis*. *Journal of applied entomology*, pp.312-315.
- Althaus, S.L., Berenbaum, M.R., Jordan, J. and Shalmon, D.A., 2021. No buzz for bees: Media coverage of pollinator decline. *Proceedings of the National Academy of Sciences*.
- Awmack, C.S. & Leather, S.R. (2002) Host plant quality and fecundity in herbivorous insects. *Annual Review of Entomology*, 817-844.
- Bale, J.S., 1996. Insect cold hardiness: a matter of life and death. *European Journal of Entomology*, pp.369-382.
- Bell, C. (2021) *Batch geocoding*. Available at: <https://www.doogal.co.uk/BatchGeocoding.php> (Accessed: 11 November 2021).
- Bell, J.R., Alderson, L., Izer, D., Kruger, T., Parker, S., Pickup, J., Shortall, C.R., Taylor, M.S., Verrier, P. and Harrington, R., 2015. Long-term phenological trends, species accumulation rates, aphid traits and climate: Five decades of change in migrating aphids. *Journal of Animal Ecology*, pp.21-34.
- Bell, J.R., Aralimarad, P., Lim, K.S. and Chapman, J.W., 2013. Predicting insect migration density and speed in the daytime convective boundary layer. *PLoS one*, p.e54202.
- Bell, J.R., Blumgart, D. and Shortall, C.R., 2020. Are insects declining and at what rate? An analysis of standardised, systematic catches of aphid and moth abundances across Great Britain. *Insect conservation and diversity*, pp.115-126.
- Biesmeijer, J.C., Roberts, S.P., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., Schaffers, A.P., Potts, S.G., Kleukers, R.J.M.C., Thomas, C.D. and Settele, J., 2006. Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science*, pp.351-354.
- Blackman, R.L. and Eastop, V.F., 1984. *Aphids on the World's Crops, An Identification Guide*. First edn. Chichester: John Wiley & Sons.
- Blumgart, D., 2021. *Investigating the Mechanisms Behind Moth Declines: Plants, Land-use and Climate* (Doctoral dissertation, Lancaster University (United Kingdom)).
- Breeze, T.D., Bailey, A.P., Balcombe, K.G., Brereton, T., Comont, R., Edwards, M., Garratt, M.P., Harvey, M., Hawes, C., Isaac, N. and Jitlal, M., 2020. Pollinator monitoring more than pays for itself. *Journal of Applied Ecology*.

- Brooks, D.R., Bater, J.E., Clark, S.J., Monteith, D.T., Andrews, C., Corbett, S.J., Beaumont, D.A. and Chapman, J.W., 2012. Large carabid beetle declines in a United Kingdom monitoring network increases evidence for a widespread loss in insect biodiversity. *Journal of applied Ecology*, pp.1009-1019.
- Brouwer, S. and Hessels, L.K., 2019. Increasing research impact with citizen science: The influence of recruitment strategies on sample diversity. *Public Understanding of Science*, pp.606-621.
- Bryman, A., 2013. Quality issues in mixed methods research. In *Talk to The White Rose Social Science DTC Second Annual Spring Conference*. University of Leicester.
- Chu, C.C., Simmons, A.M., Chen, T.Y., Alexander, P.J. and Henneberry, T.J., 2004. Lime green light-emitting diode equipped yellow sticky card traps for monitoring whiteflies, aphids and fungus gnats in greenhouses. *Insect Science*, pp.125-133.
- Collier, R.H., Tatchell, G.M., Ellis, P.R. and Parker, W.E., 1999. Strategies for the control of aphid pests of lettuce. *IOBC WPRS BULLETIN*, pp.25-36.
- Conrad, K.F., Warren, M.S., Fox, R., Parsons, M.S. and Woiwod, I.P., 2006. Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis. *Biological conservation*, pp.279-291.
- Cook, C., Heath, F. and Thompson, R.L., 2000. A meta-analysis of response rates in web-or internet-based surveys. *Educational and psychological measurement*, pp.821-836.
- Cui, W.W., 2003. Reducing error in mail surveys. Practical assessment, research & evaluation, pp.1-7.
- Dassonville, N., Thiellemans, T. and Gosset, V., 2013. FresaProtect and BerryProtect: mixes of parasitoids to control all common aphid species on protected soft fruit crops: Product development and case studies from three years of experience. *Aspects of Applied Biology*, pp.79-81.
- De Barro, P.J., 1991. Attractiveness of four colours of traps to cereal aphids (Hemiptera: Aphididae) in South Australia. *Australian Journal of Entomology*, pp.263-264.
- Dedryver, C.A., Le Ralec, A. and Fabre, F., 2010. The conflicting relationships between aphids and men: a review of aphid damage and control strategies. *Comptes rendus biologies*, pp.539-553.
- DEFRA (Department for Environment, Food and Rural Affairs)., 2021. DEFRA Statistics: Agriculture Facts England Regional Profiles March 2021.
- Diaz, B.M. and Fereres, A., 2005. Life table and population parameters of *Nasonovia ribisnigri* (Homoptera: Aphididae) at different constant temperatures. *Environmental entomology*, pp.527-534.
- Dillman, D.A., 1991. The design and administration of mail surveys. *Annual review of sociology*, pp.225-249.
- Dixon, A.F.G. 1977. Aphid Ecology: Life cycles, polymorphism, and population regulation. *Annual Review of Ecology & Systematics*, pp.329-353.

- Eveleigh, A., Jennett, C., Blandford, A., Brohan, P. and Cox, A.L., 2014, April. Designing for dabblers and deterring drop-outs in citizen science. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2985-2994.
- Ewald, J.A., Wheatley, C.J., Aebischer, N.J., Moreby, S.J., Duffield, S.J., Crick, H.Q. and Morecroft, M.B., 2015. Influences of extreme weather, climate and pesticide use on invertebrates in cereal fields over 42 years. *Global change biology*, pp.3931-3950.
- Fabre, F., Dedryver, C.A., Plantegenest, M., Hullé, M. and Rivot, E., 2010. Hierarchical Bayesian Modelling of plant colonisation by winged aphids: Inferring dispersal processes by linking aerial and field count data. *Ecological Modelling*, pp.1770-1778.
- Fera. 2021. Aphid monitoring with water traps. Available at: https://aphmon.fera.co.uk/scheme_info.cfm (accessed: 12 November 2021).
- Fewster, R.M., Buckland, S.T., Siriwardena, G.M., Baillie, S.R. and Wilson, J.D., 2000. Analysis of population trends for farmland birds using generalized additive models. *Ecology*, pp.1970-1984.
- Fisher, W.R., 2021. *Human communication as narration: Toward a philosophy of reason, value, and action*. Univ of South Carolina Press.
- Ford, E.B. (1945) *Butterflies*. Collins, London, UK.
- Frensley, T., 2017. Bridging the benefits of online and community supported citizen science: A case study on motivation and retention with conservation-oriented volunteers.
- Gaba, S., and Bretagnolle, V., 2020 Designing multifunctional and resilient agricultural landscapes: lessons from long-term monitoring of biodiversity and land use. In: *The Changing Status of Arable Habitats in Europe* (Eds. Hurford C, Wilson, P. & Storkey J.). Springer Nature Publishing. In press
- Garratt, J. R. 1992. *The Atmospheric Boundary Layer*, Cambridge, Univ. Press, pp.89-134.
- Gregory, R.D., Noble, D. and Custance, J., 2004. The state of play of farmland birds: population trends and conservation status of lowland farmland birds in the United Kingdom. *Ibis*, pp.1-13.
- Hall, D.M. and Steiner, R., 2019. Insect pollinator conservation policy innovations at subnational levels: Lessons for lawmakers. *Environmental science & policy*, pp.118-128.
- Hansen, L.M., 2006. Models for spring migration of two aphid species *Sitobion avenae* (F.) and *Rhopalosiphum padi* (L.) infesting cereals in areas where they are entirely holocyclic. *Agricultural and Forest Entomology*, pp.83-88.
- Harrington, R., Clark, S.J., Welham, S.J., Verrier, P.J., Denholm, C.H., Hulle, M., Maurice, D., Rounsevell, M.D., Cocu, N. and European Union Examine Consortium, 2007. Environmental change and the phenology of European aphids. *Global change biology*, pp.1550-1564.
- Harrington, R., Howling, G.G., Bale, J.S. and Clark, S., 1991 1. *EPPO Bulletin*, pp.499-505.

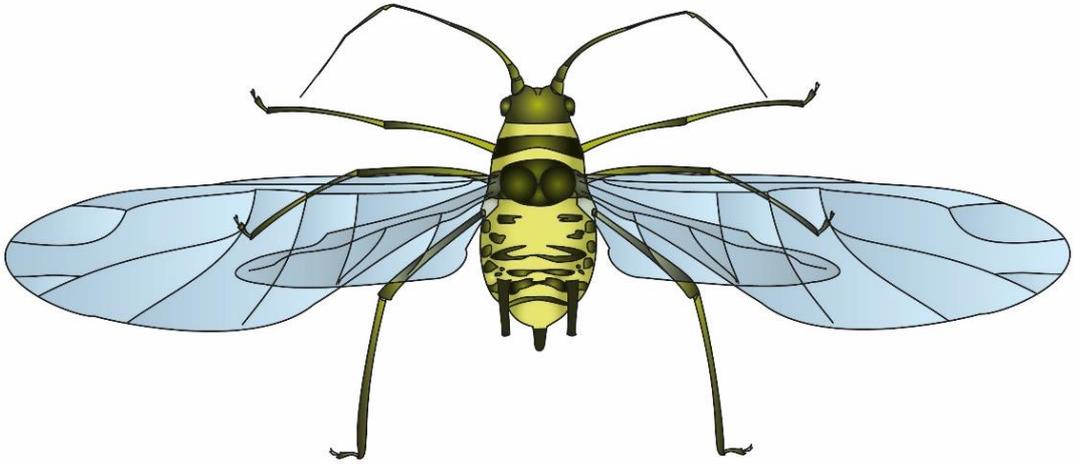
- Hassall, K.L., Dye, A., Potamitis, I. and Bell, J.R. 2021. Resolving the identification of weak-flying insects during flight: a coupling between rigorous data processing and biology. *Agr Forest Entomol*, 23: 489-505. <https://doi.org/10.1111/afe.12453>
- Heathcote, G.D., Palmer, J.M.P. and Taylor, L.R., 1969. Sampling for aphids by traps and by crop inspection. *Annals of Applied Biology*, pp.155-166.
- Holloway, P., Kudenko, D. and Bell, J.R., 2018. Dynamic selection of environmental variables to improve the prediction of aphid phenology: A machine learning approach. *Ecological Indicators*, pp.512-521.
- Hough, G. L., 2013. *Biology and Control of Currant Lettuce Aphid Nasonovia ribisnigri*, University of Warwick.
- Howling, G.G., Harrington, R., Clark, S.J. and Bale, J.S., 1993. The use of multiple regression via principal components in forecasting early season aphid (Homoptera: Aphididae) flight. *Bulletin of entomological research*, pp.377-381.
- Irwin, M.E., Kampmeier, G.E. and Weisser, W.W., 2007. Aphid movement: process and consequences. *Aphids as crop pests*, pp.153-186.
- Johnson, C. G. 1957. The vertical density of aphids in the air and the temperature lapse rate. *Quart. J. Roy. Meteor. Soc.*, pp.194–201.
- Jowett, K., Milne, A.E., Garrett, D., Potts, S.G., Senapathi, D. and Storkey, J., 2021. Above-and below-ground assessment of carabid community responses to crop type and tillage. *Agricultural and Forest Entomology*, pp.1-12.
- Karley, A.J., Parker, W.E., Pitchford, J.W., & Douglas, A.E. (2004) The mid-season crash in aphid populations: why and how does it occur? *Ecological Entomology*, 383-388.
- Karley, A.J., Pitchford, J.W., Douglas, A.E., Parker, W.E., & Howard, J.J. (2003) The causes and processes of the mid-summer population crash of the potato aphids *Macrosiphum euphorbiae* and *Myzus persicae* (Hemiptera: Aphididae). *Bulletin of Entomological Research*, pp.425-437.
- Kennedy, J.S. and Thomas, A.A.G., 1974. Behaviour of some low-flying aphids in wind. *Annals of Applied Biology*, pp.143-159.
- Kennedy, J.S., Booth, C.O. and Kershaw, W.J.S., 1961. Host finding by aphids in the field: III. Visual attraction. *Annals of applied Biology*, pp.1-21.
- Clueken, A.M., Hau, B., Ulber, B. and Poehling, H.M., 2009. Forecasting migration of cereal aphids (Hemiptera: Aphididae) in autumn and spring. *Journal of Applied Entomology*, pp.328-344.
- Knape, J., 2016. Decomposing trends in Swedish bird populations using generalized additive mixed models. *Journal of Applied Ecology*, pp.1852-1861.

- Knight, J.D. and Bale, J.S., 1986. Cold hardiness and overwintering of the grain aphid *Sitobion avenae*. *Ecological Entomology*, pp.189-197.
- Kruskal, J.B., 1964. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, pp.115-129.
- Kurppa, A., Kurppa, S. and Hassi, A., 1989. Importance of perennial grasses, and winter cereals as hosts of barley yellow dwarf virus (BYDV) related to fluctuations of vector aphid population. *Ann Agr Fenn*, pp.309-315.
- Lamichhane, J.R., Aubertot, J.N., Begg, G., Birch, A.N.E., Boonekamp, P., Dachbrodt-Saaydeh, S., Hansen, J.G., Hovmøller, M.S., Jensen, J.E., Jørgensen, L.N. and Kiss, J., 2016. Networking of integrated pest management: A powerful approach to address common challenges in agriculture. *Crop protection*, pp.139-151.
- Land-Zandstra, A.M., Devilee, J.L., Snik, F., Buurmeijer, F. and van den Broek, J.M., 2016. Citizen science on a smartphone: Participants' motivations and learning. *Public Understanding of Science*, pp.45-60.
- Leather, S.R. & Owuor, A. 1996. The influence of natural enemies and migration on spring populations of the green spruce aphid, *Elatobium abietinum* Walker (Hom., Aphididae). *Journal of Applied Entomology*, 529-536.
- Leather, S.R. and Walters, K.F.A., 1984. Spring migration of cereal aphids. *Zeitschrift für Angewandte Entomologie*, pp.431-437.
- Leather, S.R., 1983. Forecasting aphid outbreaks using winter egg counts: An assessment of its feasibility and an example of its application in Finland. *Zeitschrift für Angewandte Entomologie*, pp.282-287.
- Lewis, T., 1981. Pest monitoring to aid insecticide use. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, pp.153-162.
- Lima, M., Harrington, R., Saldaña, S. and Estay, S., 2008. Non-linear feedback processes and a latitudinal gradient in the climatic effects determine green spruce aphid outbreaks in the UK. *Oikos*, pp.951-959.
- Macaulay, E.D.M., Tatchell, G.M. and Taylor, L.R., 1988. The Rothamsted Insect Survey '12-metre'suction trap. *Bulletin of Entomological Research*, pp.121-128.
- Merril, S.C., Gebre-Amlak, A., Armstrong, J.S. and Pearirs, F.B., 2010. Nonlinear degree-day models of the sunflower weevil (Curculionidae: Coleoptera). *Journal of Economic Entomology*, pp.303-307.
- Morgan, D., 2000. Population dynamics of the bird cherry-oat aphid, *Rhopalosiphum padi* (L.), during the autumn and winter: a modelling approach. *Agricultural and Forest Entomology*, pp.297-304.
- Morris, R., 2018. First experiences with water traps. *Methodology*.

- Nebreda, M., Michelena, J.M. and Fereres, A., 2005. Seasonal abundance of aphid species on lettuce crops in Central Spain and identification of their main parasitoids/Jahreszeitliche Abundanz von Blattlausarten auf zentralspanischen Kopfsalatfeldern und Bestimmung ihrer wichtigsten Parasitoiden. *Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz/Journal of Plant Diseases and Protection*, pp.405-415.
- Nebreda, M., Moreno, A., Pérez, N., Palacios, I., Seco-Fernández, V. and Fereres, A., 2004. Activity of aphids associated with lettuce and broccoli in Spain and their efficiency as vectors of Lettuce mosaic virus. *Virus Research*, pp.83-88.
- Nieto Nafria, J.M., 1974. Aphidinea de la Cordillera Central y provincia de Salamanca. INIA Monographs, p.168.
- Oakley, J.N. and Walters, K.F.A., 1994. A field evaluation of different criteria for determining the need to treat winter wheat against the grain aphid *Sitobion avenae* and the rose-grain aphid *Metopolophium dirhodum*. *Annals of Applied Biology*, pp.195-211.
- Parry, H.R., 2013. Cereal aphid movement: general principles and simulation modelling. *Movement Ecology*, pp.1-15.
- Peng, X., Qiao, X. and Chen, M., 2017. Responses of holocyclic and anholocyclic *Rhopalosiphum padi* populations to low-temperature and short-photoperiod induction. *Ecology and evolution*, pp.1030-1042.
- R CORE TEAM 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Random, 2019. Random.org. Available at: www.random.org/list (Accessed: 10/01/2020).
- Reinink, K., Dieleman, F. L., 1993. Survey of aphid species on lettuce. IOBC/WPRS Bull. pp.56-68.
- RIS. (2021) Aphid bulletin | Insect Survey. Available at: www.insectsurvey.com/aphid-bulletin (accessed: 11 November 2021).
- Sauer-Kesper CL, Lucia N, Buser H and Vogler U., 2011. Bedeutung und Verbreitung des neuen Biotyps-Nr:1 der Grünen Salatlaus in der Deutschschweiz. *Agrarforschung Schweiz*. pp.462–469.
- Southwood, T.R.E. & Comins, H.N. 1976. A synoptic population model. *Journal of Animal Ecology*, 949-965.
- Stufkens, M.A.W. and Teulon, D.A.J., 2001. Aphids species on potato crops in Canterbury. *New Zealand Plant Protection*, pp.235-239.
- Tatchell, G.M. and Woiwod, I.P., 1990. Aphid migration and forecasting. In *Meeting on Euraphid network: trapping and aphid prognosis, proceedings*, pp. 15-28.
- Taylor LR. Aphid dispersal and diurnal periodicity, *Proc Linn Soc Lond*, 1958, pp.67-73.

- Taylor, L. R. 1974. Insect Migration, Flight Periodicity and the Boundary Layer. *Journal of Animal Ecology*, 43(1), 225–238. <https://doi.org/10.2307/3169>.
- Taylor, L.R., 1960. The distribution of insects at low levels in the air. *The Journal of Animal Ecology*, pp.45-63.
- ten Broeke, C.J., Dicke, M. and van Loon, J.J., 2013. Feeding behaviour and performance of different populations of the black currant-lettuce aphid, *Nasonovia ribisnigri*, on resistant and susceptible lettuce. *Entomologia Experimentalis et Applicata*, pp.130-141.
- Thackray, D.J., Diggle, A.J. and Jones, R.A.C., 2009. BYDV PREDICTOR: a simulation model to predict aphid arrival, epidemics of Barley yellow dwarf virus and yield losses in wheat crops in a Mediterranean-type environment. *Plant Pathology*, pp.186-202.
- Wainwright, C.E., Stepanian, P.M., Reynolds, D.R. and Reynolds, A.M., 2017. The movement of small insects in the convective boundary layer: linking patterns to processes. *Scientific Reports*, pp.1-8.
- Walters, K.F.A. and Dewar, A.M., 1986. Overwintering strategy and the timing of the spring migration of the cereal aphids *Sitobion avenae* and *Sitobion fragariae*. *Journal of Applied Ecology*, pp.905-915.
- Warren, M.S., Hill, J.K., Thomas, J.A., Asher, J., Fox, R., Huntley, B., Roy, D.B., Telfer, M.G., Jeffcoate, S., Harding, P. and Jeffcoate, G., 2001. Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature*, pp.65-69.
- Watt, A.D., 1983. The influence of forecasting on cereal aphid control strategies. *Crop Protection*, pp.417-429.
- Way, M.J., Cammell, M.E., Taylor, L.R. and Woiwod, I.P., 1981. The use of egg counts and suction trap samples to forecast the infestation of spring-sown field beans, *Vicia faba*, by the black bean aphid, *Aphis fabae*. *Annals of applied biology*, pp.21-34.
- Werker, A.R., Jaggard, K.W. and Allison, M.F., 1999. Modelling partitioning between structure and storage in sugar beet: effects of drought and soil nitrogen. *Plant and soil*, pp.97-106.
- Zhou, X., Harrington, R., Woiwod, I.P., Perry, J.N., Bale, J.S. and Clark, S.J., 1995. Effects of temperature on aphid phenology. *Global Change Biology*, pp.303-313.

Chapter 3



CHAPTER 3: Alternative summer hosts, green bridges, overwintering biology and using a novel molecular method to detect previous host plants of *N. ribisnigri*

3.1 Introduction

The currant lettuce aphid (*Nasonovia ribisnigri*) usually colonises outdoor lettuce (*Lactuca sativa*) as a secondary host plant in late summer. The subsequent reproduction in the crop is exponential, and since *N. ribisnigri* preferentially feeds in the centre of lettuce plants, they are difficult to detect until the crop quality is compromised. Due to this feeding behaviour, foliar applications of non-systemic insecticides are ineffective, making control particularly difficult.

The difficulties of capturing *N. ribisnigri* using different trapping methodology were previously detailed (Chapter 2), highlighting the associated problems of constructing an accurate forecasting model. Despite the evident 50-year decline in abundance throughout the UK, *N. ribisnigri* is still the most commercially damaging aphid on outdoor lettuce (Reinink and Dieleman, 1993; RIS, 2021; Sauer-Kesper *et al.*, 2011; Stufkens and Teulon, 2003). *Nasonovia ribisnigri* does not usually become a problem on outdoor lettuce until late summer (July-October) but appears in low numbers in the RIS suction traps from April. It is unclear why this large delay between detection in suction traps and colonisation of outdoor lettuce occurs, despite the availability of outdoor lettuce crops during this period. As a result, to elucidate when *N. ribisnigri* will become a problem in outdoor lettuce crops to enable earlier control measures and mitigative approaches to be implemented, it is desirable to understand their autecology.

Four experiments were established to address the evident scientific gaps in understanding where *N. ribisnigri* resides prior to migration onto outdoor lettuce. The first two experiments were carried out in parallel on field sites belonging to Gs Fresh (industry partner) to explore the use of different alternative summer hosts in the farm landscape. The third experiment investigated whether host plant barcodes (ITS2) can be detected in *N. ribisnigri* using molecular tools, and the fourth investigated the

overwintering phenology of *N. ribisnigri* on a range of primary hosts. A detailed diagram depicting the following experiments and how they address the aim of this chapter is located in figure 3.1.

3.2 Experiment 1 & 2: Field margins, alternative hosts, and green bridges

Nasonovia ribisnigri is known to overwinter on a variety of woody currant plants (*Ribes* spp.) (primary host) and migrate in the spring to several secondary host plants within the Brassicaceae, Scrophulariaceae, Solanaceae and Asteraceae (which includes lettuce) families (Diaz and Fereres, 2005). Despite the prevalence of *N. ribisnigri* on outdoor lettuce in late summer, current monitoring strategies fail to capture sufficient data to predict when this pest is likely to become a problem in the crop. A number of trapping approaches have been investigated to improve capture of *N. ribisnigri*, but none have been shown to capture consistently large numbers (see Chapter 2). In recent years, the RIS suction trap network captures have been between 50-100 *N. ribisnigri* per annum across the whole network, whereas numbers of *Myzus persicae* and *Rhopalosiphum padi* during the same time have been much higher, 13,088 and 31,184, respectively (RIS, 2021). The sparse recoveries of *N. ribisnigri* hinder the construction of accurate forecasting models and mean that effective 'routine' field monitoring strategies based on trapping data are severely compromised. Although the first flights usually occur during April-May, *N. ribisnigri* does not become a problem in outdoor lettuce crops until late summer and early autumn (August – October), despite the availability of outdoor lettuce crops between April – November. Hence, there is an absence of biological insight between April and August when populations of *N. ribisnigri* are rarely found.

Within a typical farmland environment, there are many potential alternative summer host plants which may act as a reservoir for pests (Maling *et al.*, 2008). Pest populations could build up on these alternative summer hosts and then move into crops in large numbers, seriously impacting crop quality and yield. Such alternative hosts are often termed to be providing a 'green bridge', which refers to a host plant or set of host plants which provide resources to pests and facilitate later movement into

the crop (Favetti *et al.*, 2017). Without understanding the host plant relationships of *N. ribisnigri*, growers may unknowingly be providing a green bridge for aphids to move to the crop.

The first two experiments aim to i) identify the potential alternative hosts within different types of field margin within an organic and conventional setting, ii) determine whether *N. ribisnigri* resides within these different types of margin during the summer and, iii) whether known alternative host plants can encourage *N. ribisnigri* colonisation, and help understand if these could be used as a 'trap crop' (push-pull strategy) (Ben-Issa *et al.*, 2007; Cook *et al.*, 2007).

These objectives were achieved using two methodologies which were run in parallel at the field sites: experiment 1 involved field margin surveys of both grass and broadleaf species for i), and invertebrate (aphid) sweep net sampling in field margins for ii); and experiment 2 establishing *in situ* monitoring alternative plant host plots to study *N. ribisnigri* colonisation for iii).

3.3 Experiment 3: Identifying host plants in *Nasonovia ribisnigri*

Aphids evolved as a polyphagous group, with specialisation emerging some time later, probably as a result of the diversification of flowering plants in the early Cretaceous (Dixon, 1994). In today's aphid fauna, 90% of species are monophagous and do not host alternate between primary and secondary hosts, although most horticultural pests are predisposed to host-alternation (heteroecious) and have a wide range of plant hosts (Blackman & Eastop, 1984; Dixon 1977; Moran 1988). The host plants of most species of aphid are well documented and, combined with the use of data from the Rothamsted Research Insect Survey (RIS) long-term standardised monitoring of aphids, their phenology and migration to and from the host can be inferred (Macaulay and Taylor, 1988). This information is particularly important for monitoring the migration of horticulturally important aphids between their primary to secondary hosts in the spring. This information would be especially valuable for a cryptic aphid species such as *N. ribisnigri*, for which there are limited sets of trapping data to analyse and many knowledge gaps concerning their movement between their winter and summer host plants.

Predator gut analyses determine what animals have consumed, typically incorporate primers to amplify the mitochondrial cytochrome *c* oxidase gene, which has been shown to enable researchers to discern species identity in metabarcoding studies (Batovska *et al.*, 2021; Utzeri *et al.*, 2018; Toju *et al.*, 2018). In plants, the mitochondrial genome evolves more slowly than in animals and this gene is not a good candidate to use to determine species identity (Coissac *et al.*, 2007). Several similar barcodes have been discovered in plants, including *rbcL*, *matK* and *trnL*, but factors such as amplicon size and relatively low taxonomic resolution make them unsuitable for analysis of low quantities or degraded plant DNA (Dong *et al.*, 2012; Gielly *et al.*, 1994; Liu *et al.*, 2011). The internal transcribed spacer (ITS2) is an alternative barcode to these genes. It is a DNA region of 160-320 bp in length and has been shown to be highly specific to species level (Chen *et al.*, 2010). There are two ITSs in eukaryotic organisms, ITS1 which is located between 18S and 5.8S ribosomal RNA (rRNA), whereas ITS2 is situated between 5.8S and 28S in fungi (or 25S in plants) (Lafontaine and Tollervey, 2001). A recent study on the gut contents of herbivorous reptiles and birds successfully amplified the ITS2 plant barcode in highly DNA degraded faecal samples with a success rate of 86.1%, 99.4% and 99.9% for species, genus and family, respectively (Moorhouse-Gann *et al.*, 2018). Although these methods have not been applied to study the gut contents of small herbivorous insects, in particular aphids, the principal method remains the same.

For studies on aphids, allozyme electrophoresis has been used to detect the presence of hymenopteran parasitoid larvae in individual aphids (Castañera *et al.*, 1983; Walton *et al.*, 1990). Alternatively, the use of the polymerase chain reaction (PCR) technique and specific primers was sensitive enough to detect a single parasitoid wasp egg in the English grain aphid, *Sitobion avenae* (Traugott *et al.*, 2008). In a different study, 16S rDNA markers have been used to detect the presence of facultative endosymbiotic bacteria, especially in relation to their role in host-plant adaptation (Simon *et al.*, 2003). The methodology of Moorhouse-Gann *et al.* (2018) could be informed by these allied methods and used to elucidate general patterns of aphid host-plant ecology. Specifically, this approach could be applied to elucidate the cryptic autecology of *N. ribisnigri*, by identifying previous host plants and thus the 'green bridge' into lettuce.

To date, an aphid host-plant bioassay has yet to be used and thus a method to identify the feeding history of an individual is also absent. Obtaining conclusive evidence about the identity of both the primary overwintering host plant(s) and alternative secondary summer hosts would be highly advantageous in supporting monitoring and mitigative approaches as part of integrated pest management (IPM).

Experiment 3 aims to i) determine whether the ITS barcode from a potential host plant could be amplified in individual aphids using the UniPlant primers, ii) see how long the plant ITS barcode persists within an aphid, at detectable levels, iii) explore if the plant ITS barcode can be identified within the stylets/salivary glands or midgut and, iv) apply this method to field-caught specimens to identify the recent host plants of migrant aphids.

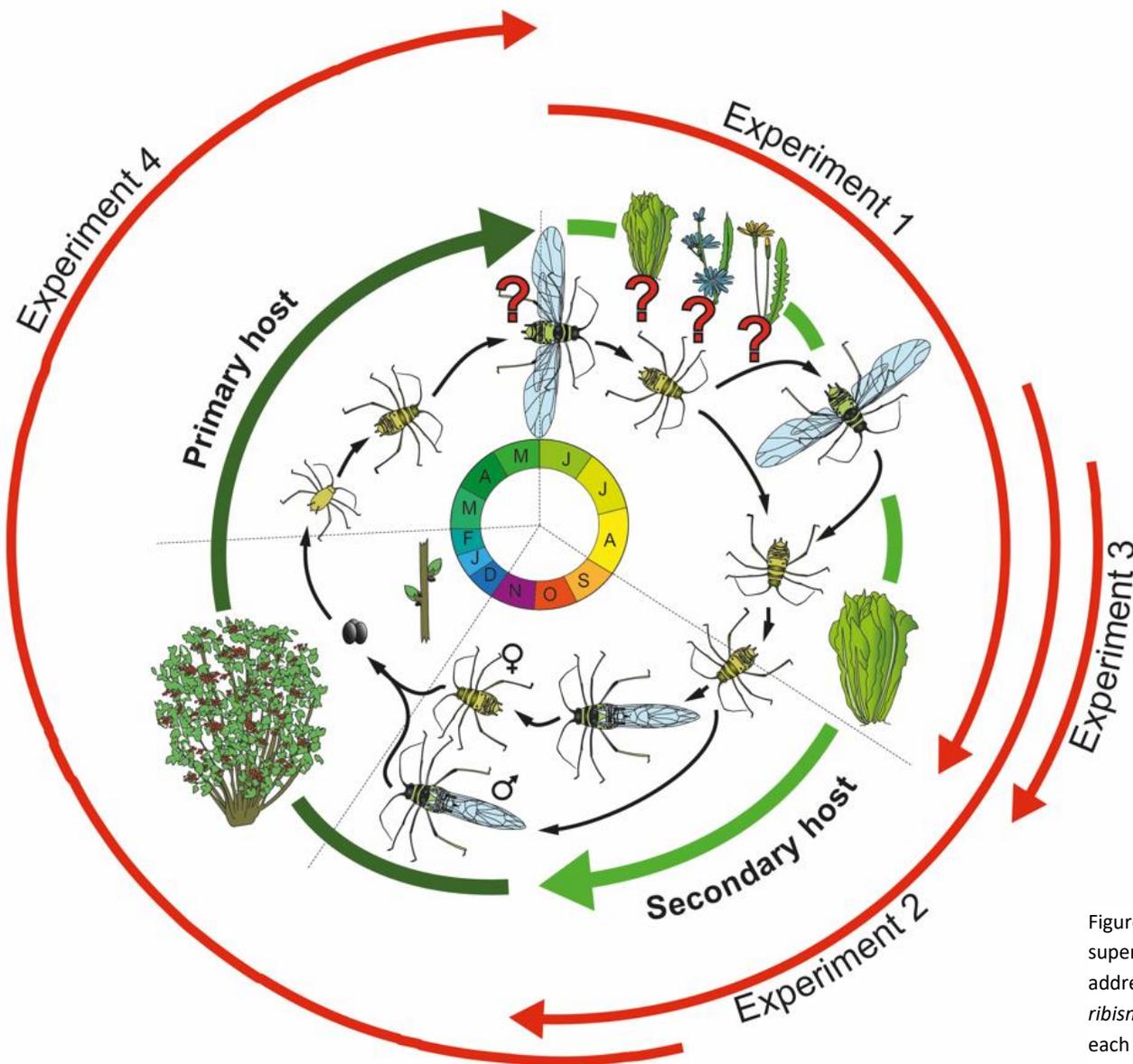
3.4 Experiment 4: Overwintering biology

Understanding the sequential pattern of host alternation in pest aphids can help to predict timing of migration and subsequent arrival in crops. *N. ribisnigri* is heteroecious and moves between a primary host (in the currant family (*Ribes* spp.)) and a secondary host, which includes members of the Asteraceae (including lettuce *Lactuca sativa*) and Solanaceae (Blackman and Eastops, 1984; Diaz and Fereres, 2005). *Nasonovia ribisnigri* usually has a holocyclic life cycle (sexual, producing eggs in winter), but has also been observed overwintering as a nymph or adult on secondary host plants in Southern England whereby sexual reproduction is 'suppressed' (anholocyclic) (van der Arend, 2003).

Previous research on host alternation and the overwintering behaviour of *N. ribisnigri* has been conducted in a laboratory setting. Hough (2013) established that *N. ribisnigri* adults and nymphs cease to develop at temperatures below 4.6°, with upper threshold temperatures above 26.4°C being detrimental to development. There is an evident gap in the literature linking overwinter biology and host alternation in a semi-natural or field setting. Small temperature fluctuations, changes in humidity, atmospheric pressures, host plant and host plant quality, photoperiod (L:D) and precipitation could be influencing and altering their development (Hand, 1983; Siddiqui *et al.* 1973). Furthermore, by

providing a secondary host in the spring (lettuce), this experiment will inform on whether *N. ribisnigri* will accept lettuce in the spring.

Experiment 4 aimed to i) better understand how *N. ribisnigri* colonises the primary host in a semi-field environment, ii) determine when alates are first produced in the following spring and, iii) understand if *N. ribisnigri* would accept and colonise lettuce in the following spring.



Experiment 1: Identifying potential alternative summer hosts of *N. ribisnigri* in different field margins types in both organic and conventional farms in Cambridgeshire and Norfolk. Using sweep netting to capture *N. ribisnigri* in the field margins.

Experiment 2: Incorporating known alternative summer host plant plots to encourage *N. ribisnigri* colonisation and study feeding behaviour in both organic and conventional farms in Cambridgeshire and Norfolk.

Experiment 3: Using molecular techniques to identify the host plant barcode (ITS2) of *N. ribisnigri* to identify potential green bridge reservoirs.

Experiment 4: Semi-field overwintering experiment to observe colonisation of known winter hosts (*Ribes* spp.) and understand whether a lettuce host plant will be accepted the following spring during alate migration (dispersal).

Figure 3.1: The lifecycle of *N. ribisnigri* with each experiment in this chapter superimposed around it. This demonstrates how each experimental component addresses the current unknown scientific knowledge gaps throughout the *N. ribisnigri* holocyclic lifecycle. Red arrows indicate the duration and time frame of each experiment.

3.5 Materials and methods

3.5.1 Experiment 1: Field margins surveys to identify potential *N. ribisnigri* host plants

3.5.1.1 Study site

The study was conducted at two sites within eastern England (near Ely, Cambridgeshire), on peat Fenland farmland. The study sites were managed and operated by the same company (G's Fresh), with similar crops being cultivated during the year at each site (celery, lettuce, potato, salad onions and beet). The primary difference between the two study sites was the management practices. The first site was situated in Ely, Cambridgeshire, UK (52°21'N, 0°16'E) and is an organically managed farm (Hainey) (Fig. 3.2). The second study site was situated in Wisington, Norfolk, UK (52°32'N, 0°27'E) and is managed conventionally, with a standardised protocol for insecticide and fungicide applications (Wisington) (Fig. 3.3).

3.5.1.2 Habitat type

Agri-environment schemes provide funding to growers and land managers to farm in a manner that supports and encourages biodiversity, enhances the landscape, and improves the quality of water, air and soil (Smallshire *et al.*, 2004). Both farms used as study sites are enrolled in agri-environment schemes (AES) and, as a result, grow a selection of seeded plant mixes which are used to encourage pollinating insects, provide refuges for birds, enhance overall biodiversity and improve biological control within vegetable crops. These plant mixes are distributed throughout the study sites and are within field margins or situated close to a field margin.

To understand the variation in the flora present in the field margins and the surrounding area, three categories of vegetation were selected to represent each margin type. The vegetation field margin composition surveys were conducted using quadrats (1 m²) for the three margin types:

Seeded plant mix (AES) – predominantly consisting of a flower-rich seed mix for pollinators, an 'organic' bird seed mix and winter bird seed mixes. Growing these seeded mixes are part of the AES.

Grass – These assemblages were grass mixes provided to enrich and provide a refuge for various wildlife.

Semi-natural – An ecological assemblage that has been modified by anthropogenic activity. In this instance, the semi-natural habitat consists of a variety of weed species, grasses and small shrubs.

The study was conducted between June-August 2019, during the peak season for growing outdoor lettuce and before *N. ribisnigri* usually becomes a problem in the crop. To select survey locations and reduce bias, a numbered grid-map overlay was applied to both sites on pre-determined habitat types (AES, Grass, Semi-natural) and selected using a random number generator (Random, 2019) for each characterised and assigned habitat type (Fig. 3.2 and 3.3).

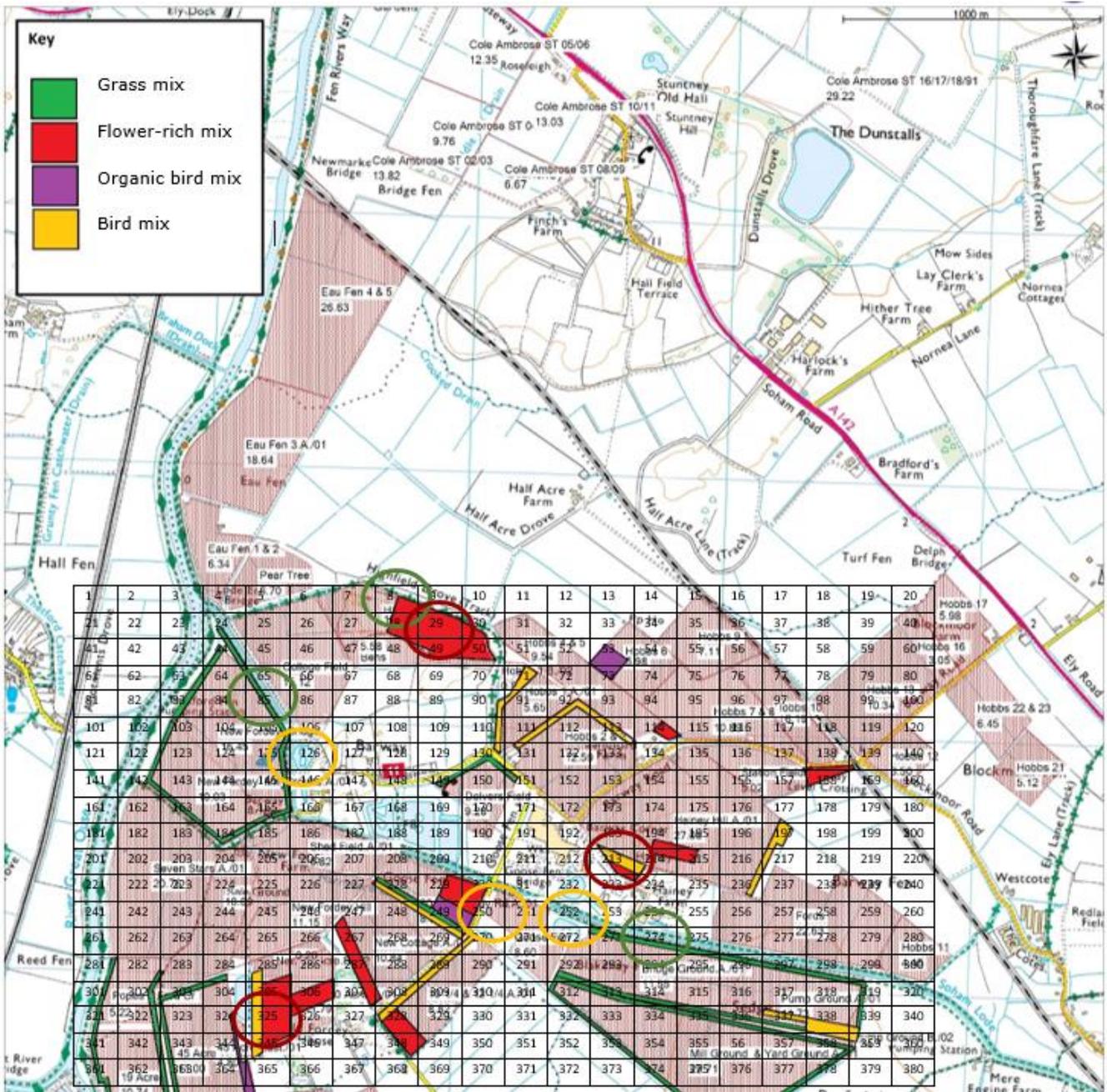


Figure 3.2: A map of the organically managed farm near Ely (Hainey), Cambridgeshire, UK (52°21'N, 0°16'E). Grid overlay highlighting how sampling sites (circled) were identified and chosen. Red circles – AES planted; yellow circles – Semi-natural; green circles – grass. Different margin types on the site are shown in different colours (Green – grass, Red – Flower rich margins (AES), Purple – Organic bird seed mix, Yellow – Bird mix). Only flower-rich margins (Red) were used for the AES margin type. Source: Gs Fresh.

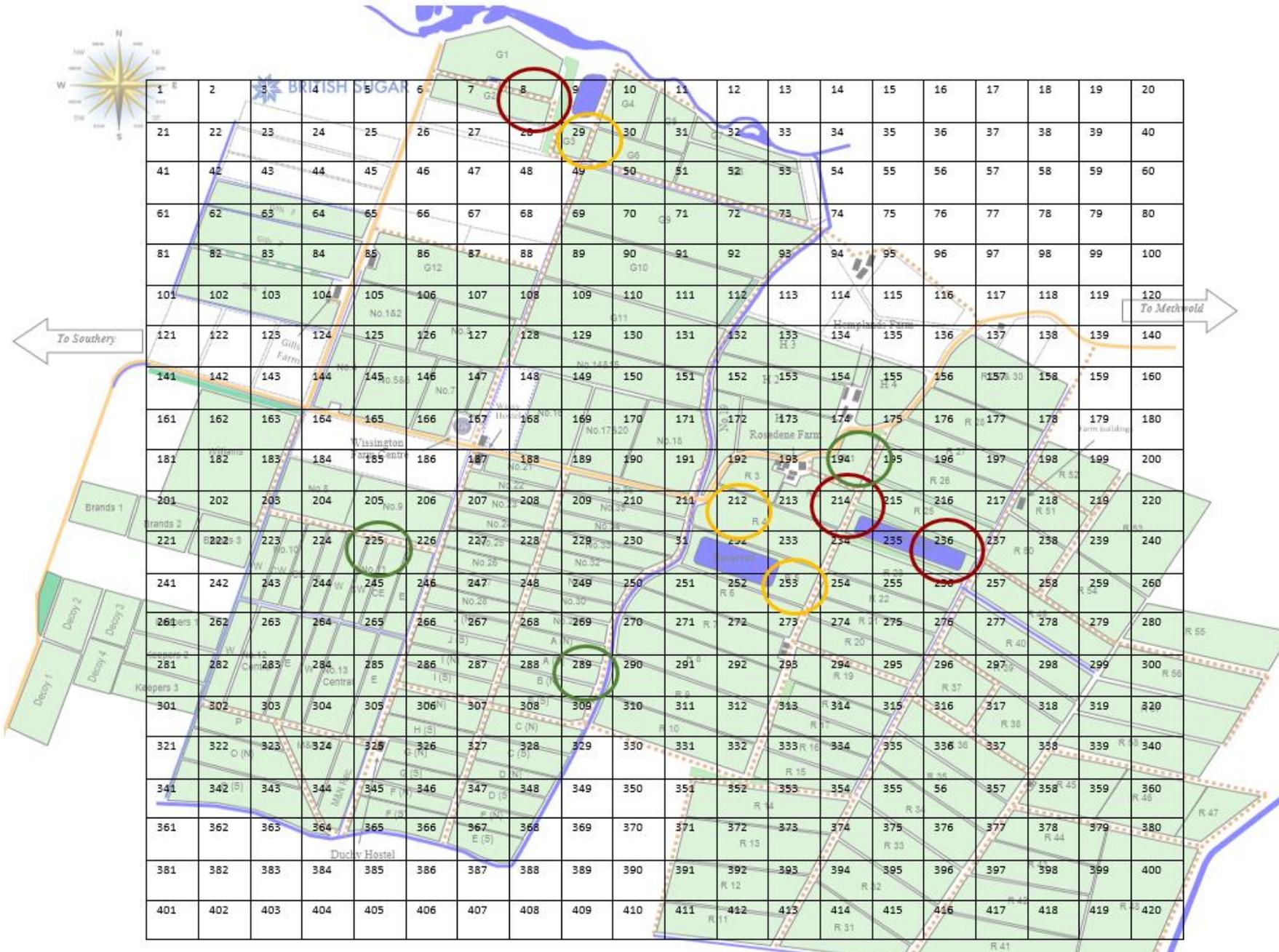


Figure 3.3: A map of the conventionally managed farm near Wissington, Norfolk, UK (52°32'N, 0°27'E). Grid overlay highlighting how sampling sites (circled) were identified and chosen. Red circles – seeded mixes; yellow circles – Semi- natural; green circles – grass. Source – Gs Fresh.

3.5.1.3 Vegetation field margin composition

Field margin surveys were only conducted during calm days (<10mph wind speed; temperature >15°C; cloud cover <70%; zero precipitation) to minimise the variation in insect activity due to weather conditions. Field margin survey locations on both the organic (Hainey) and conventional (Wissington) farms were situated at least 200m apart. At each site, a 20m transect, was used throughout the duration of the study. Transect start points were selected at random and measured and marked out for subsequent surveys. Along each transect, two 1 m² quadrats were randomly placed (at least 5m apart); all plant species within each quadrat was identified, counted, with the total percentage of each species recorded (by area covered) within the quadrat (Fig. 3.4). For ease of interpretation, species were grouped into either grasses or broadleaved (Forbs) species. The vegetation field margin



Figure 3.4: Examples of quadrats in A) a flower-rich margin (AES); B) a grass margin; C) Semi-natural habitat.

composition surveys were repeated three times across both organic and conventional sites (once a month). In total, three transects of each margin type were selected in triplicate for both sites (18 transects in total).

3.5.1.4 Invertebrate assemblages within field margins

Invertebrate sampling was carried out in each transect using a sweep net, prior to surveying the vegetation. Use of a sweep net has been shown to capture a diverse range of invertebrates, including aphids, and provides a good indication of the invertebrate assemblage within a given area (Beall, 1935; Dixon *et al.*, 2011). For each sample, the net was swept from side to side, across a width of 3m, in a smooth figure of eight motion, whilst the operator walked slowly along the predetermined 5m transect (Fig. 3.5). Two sweep net samples were taken within each 20m transect. Sweeping was done



Figure 3.5: Using a sweep net to sample for invertebrates along one of the 20m transects at the conventional farm in Norfolk (Wissington).

at the top-mid level of the vegetation to maximise the number of invertebrates captured and to standardise the sampling method. Each sample was immediately transferred into a sample pot and labelled (study site, field margin type, transect number (1 or 2) and date). Samples were processed, with all aphids identified to species.

3.6 Experiment 2: Alternative host plant plots to encourage *N. ribisnigri* colonisation

3.6.1 Study site

For experiment 2, the study was conducted at same sites as previously described (section 3.5.1.1). For site 1 (Hainey), four plots were identified to be used as alternative host plant plots (Fig. 3.6). For site 2 (Wissington), four plots were also identified to be used as alternative host plant plots (Fig. 3.7). For each site, four plots were selected to provide good site coverage and situated away from horticultural machinery and access. Each plot was 20 m long and 4 m wide and positioned adjacent to grass field margins within the field.

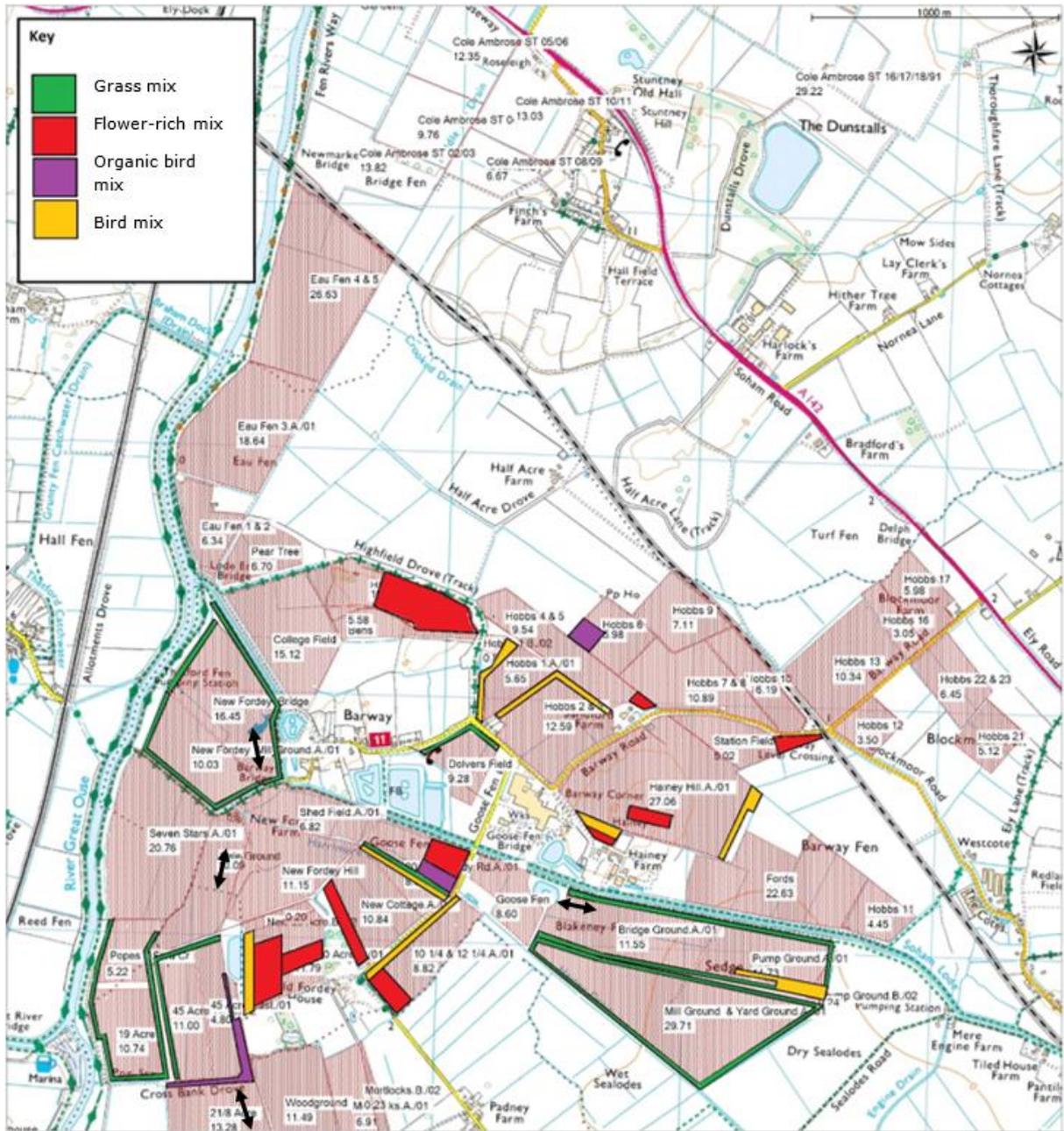


Figure 3.6: The organically managed farm in Ely, Cambridgeshire, UK (52°21'N, 0°16'E). Red area – AES planted; yellow areas – Semi-natural; green areas – grass – expand these descriptions. Black arrows show where the alternative host plant plots were located throughout Experiment 2 in 2020. Field locations: Mill Ground, Sale Ground, Bridge Ground and 21/8 Acre.

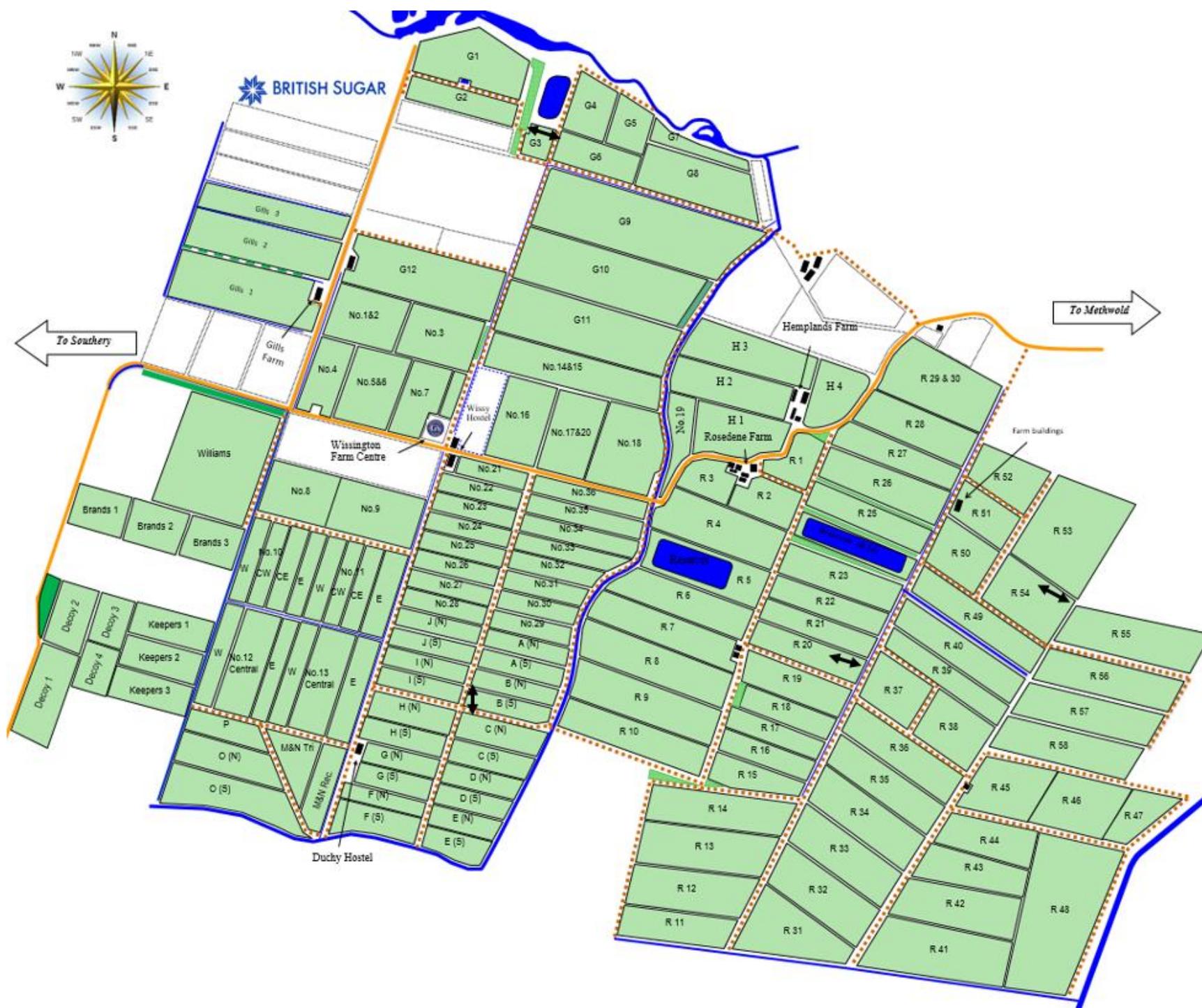


Figure 3.7: The conventionally managed farm near Wissington, Norfolk, UK ($52^{\circ}32'N$, $0^{\circ}27'E$). Black arrows show where the alternative host plant plots were located throughout Experiment 2 in 2020. Field locations: G3, R54, R20 and B(S).

3.6.2 Alternative plant host plot experiment design

A selection of alternative summer host plants shown to be accepted by *N. ribisnigri* were selected based on information from the literature. According to Hough (2013), common chicory (*Cichorium intybus*), smooth hawks-beard (*Crepis capillaris*), common nipplewort (*Lapsana communis*), fox-and-cubs (*Hieracium aurantiacum*), mouse-ear-hawkweed (*Hieracium pilosella*), wall speedwell (*Veronica arvensis*), common speedwell (*Veronica officinalis*) and birdseye speedwell (*Veronica chamaedrys*) were successful alternative hosts for *N. ribisnigri*. Based on their commercial availability, four host plants were selected: chicory (*C. intybus*), mouse-ear-hawkweed (*H. pilosella*), birdseye speedwell (*V. chamaedrys*) and rough hawkbit (*Leontodon hispidus*). Batches of 150 host plant plugs (4-weeks old) of each species were sourced from Boston Seeds, allowing for 15 plants to be transplanted in each of the eight plots (Fig. 3.8). Transplants of an open-heart romaine lettuce (*Lactuca sativa* cv. Auvona) cultivar were provided by G's Fresh. This *L. sativa* variety is susceptible to *N. ribisnigri* (does not possess the Nr-gene resistance) and therefore aphids of both the susceptible and resistance-breaking biotypes would be able to colonise the lettuce. A randomised block design was incorporated to remove potential bias, with the lettuce (cv. Auvona) acting as a pseudo control. The alternative host plant plots were 24m long and 4m wide and divided into 5 sub-plots (4m each) to accommodate each host plant (four blocks for both site 1 and site 2, eight in total). The plots in which each host plant species was planted was randomised (Random, 2019).

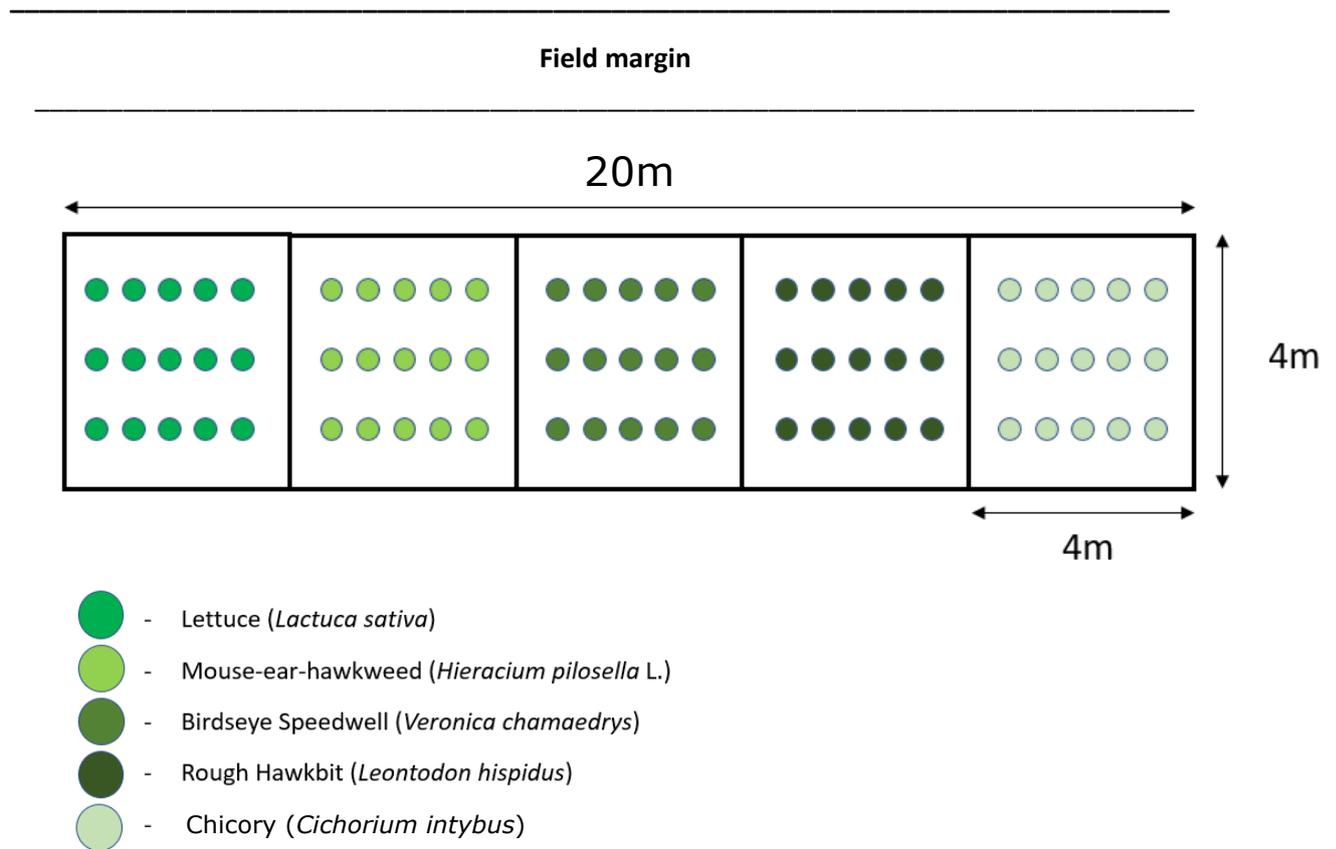


Figure 3.8: Plot layout for the alternative plant host plot experiment. Fifteen host plant plugs were directly planted into the prepared soil as 3 rows of five plants. The alternative host plant plots were 24m long and 4m wide and divided into 5 sub-plots (4m each) to accommodate each host plant (four blocks for both site 1 and site 2, eight in total). The plots in which each host plant species was planted was randomised (Random, 2020).

The plots were prepared during February and March 2020. This consisted of marking out each plot, weeding them by hand, and using a hoe and gardening fork to prepare the soil ready for the plant plugs. The plug plants were transplanted in the first week of April 2020, illustrated in figure 3.9. Wooden posts (2m long, 2 x 4 inch) were situated in each corner of the plot and along the length of the plot with signage to prevent accidental damage by agricultural machinery (Fig. 3.9 A). Due to the relatively short growth period of lettuce, new lettuce plants were transplanted during the middle of the study period (June). These new plants replaced the old lettuce plants and any aphids found were transplanted onto the new plants.

Regular maintenance of the plots was conducted to prevent unwanted species (e.g. groundsel) encroaching on the area. Additionally, for the first 4-weeks after transplanting, water was applied using a watering can to allow the plants to establish successfully (Fig. 3.9 B). After this initial bedding-in period, water was applied only during any prolonged period of warm and dry weather.

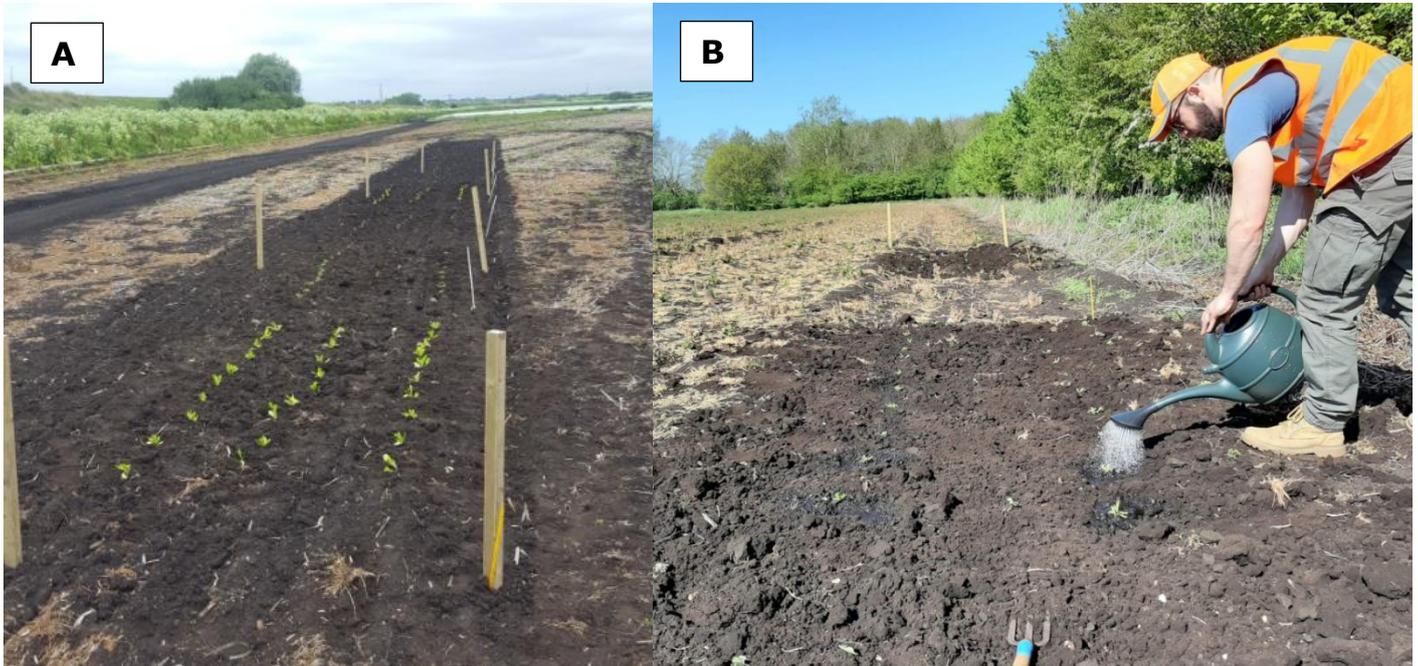


Figure 3.9: A) newly planted plot on site 1 near Hailey in Cambridgeshire (organic). B) the plants were watered for the first 4-weeks to allow them to establish (site 2, near Wissington site in Norfolk (conventional)).

3.6.2.1 Data collection

Weekly monitoring of the plots began at the end of April 2020 and continued until October 2020 - state how many weeks. Monitoring consisted of non-destructive visual checks of each plant for the presence of *N. ribisnigri* or other aphids. The presence of *N. ribisnigri* or other species of aphid was recorded together with the location, time, date, and host plant type.

3.6.3 Experiment 3: Using molecular techniques to identify the host plant barcode (ITS2) of *Nasonovia ribisnigri* to identify potential green bridge reservoirs

In this experiment, the feeding history of *N. ribisnigri* was studied *in vivo* using molecular tools to determine the levels and location of host plant material. The first part of the experiment was to understand whether the lettuce plant barcode could be detected in the first instance. The second part was to determine how long the plant barcode remained at detectable levels after periods of starvation. Thirdly, understand if trace elements of the plant barcode reside within the head/stylets or within the aphid fore/mid gut. Finally, using the newly developed method, applying it to field-caught *N. ribisnigri* to determine its potential field applications.

3.6.3.1 Study organism

A resistance-breaking biotype culture of *N. ribisnigri* (Nr 8) was used, originally collected from a lettuce crop in Yorkshire in 1999. The culture was maintained under laboratory conditions (16:8 (Light:Dark)) at a constant room temperature of $21\pm 2^{\circ}\text{C}$ on a susceptible cultivar of lettuce (*Lactuca sativa* cv. Pinokkio) at Warwick Crop Centre (University of Warwick).

Prior to the study, aphids were transported to Rothamsted Research, where a new culture was established and kept under similar laboratory conditions to those described above. For each new culture, a single apterous aphid (founding mother) was transferred onto five new lettuce plants (cv. Pinokkio) and reared for 4-weeks prior to the experiment. Five founding mothers were used for each of the five new cultures (one per culture). Both winged (alates) and non-winged (apterous) adults were used in the experiment.

3.6.3.2 Experimental design

An experiment was set up to determine, **i)** whether the lettuce 'barcode' can be detected in either alate or apterous *N. ribisnigri* morphs, **ii)** if so, for how long could the lettuce barcode remain detectable during periods of starvation (up to 48 hours), **iii)** whether detectable traces of the lettuce host plant are present within the head/stylet or the fore/mid gut of the aphid and **iv)** applying the new method to field-caught *N. ribisnigri* to determine its potential field applications.

10-20 alate and apterous *N. ribisnigri* were removed from each of the five culture plants and transferred into five separate 1.5ml Eppendorf tubes containing 1ml of 100% ethanol (0-hour time point) (**i**). These aphids were used to determine whether the lettuce plant barcode could be detected in the first instance. A further 30-50 alates and apterae were removed from each of the five culture plants and placed into a mesh enclosure, with no access to a lettuce host plant. These aphids were subjected to a period of starvation prior to transferring them to 100% ethanol as above. Starvation periods were as follows: 1, 3, 6, 24 and 48 hours (**ii**). All samples were transferred to a -20°C freezer for long-term storage prior to DNA extraction and analysis.

To determine whether plant material could be detected in the stylets or the midgut of the aphids, a separate experiment was conducted using excised stylets/heads and abdomens only on samples collected in the 0-hour time point (**iii**). Due to the small sizes of the stylets and abdomens, three stylets/heads and three abdomens from the same culture were used for DNA extraction.

Finally, a pilot experiment was conducted using whole aphids (*N. ribisnigri*) obtained from the Rothamsted Research Insect Survey (RIS) suction trap sample archive to determine whether any trace amounts of host plant DNA could be amplified from stored aphids (**iv**). Once collected and sent to Rothamsted Research, the aphids were identified to species and stored in 95% ethanol with 5% glycerol. Four alate *N. ribisnigri* captured by the suction trap at Broom's Barn in Suffolk in July-August 2018, were used. An additional four *N. ribisnigri* alates that had been captured in a 2m high suction trap situated on the organic farm (near Hainey) in Cambridgeshire were also tested. These had been captured in the same period as the aphids from Broom's Barn (see figure 3.1 for details).

3.6.3.3 DNA extraction

DNA extractions were conducted on whole individual aphids using the Qiagen QIAmp DNA Micro Kit (Qiagen, Manchester, UK) following manufacturers recommendations, with the following adjustments: individuals were homogenised using a pestle in liquid nitrogen in a 1.5ml Eppendorf, 180ul of ATL buffer was added along with 20ul proteinase K, mixed with gentle flicking, briefly spun down using a microcentrifuge and incubated overnight on an orbital shaker at 200-220 rpm at 56°C. 200ul of AL buffer (with 1µg/µl RNA carrier) was added to the samples the following morning, inverted 10 times, then 200ul of 100% ethanol added and incubated at room temperature for 5 minutes. The entire sample was transferred to a QIAmp DNA Micro Kit spin column and spun at 8000 rpm (6010 g) for 1 minute at room temperature (Eppendorf centrifuge 5425 R). The flow-through was discarded and the column placed into a new collection tube with 500ul of AW1 buffer added and centrifuged at 8000 rpm. The flow-through again was discarded and placed in a new collection tube with 500ul of AW2 buffer added into the column and centrifuged at 8000 rpm for 1 minute at room temperature. Finally, the flow-through was again discarded, the column was added into a new collection tube and spun at 13000 rpm (15871 g) for 3 minutes at room temperature. The column was placed into a 1.5ml Eppendorf and 40ul of AE buffer was added directly onto the column membrane, incubated at room temperature for 10 minutes and spun down for 1 minute at 14000 rpm (18440 g). To improve DNA yield, the recovered AE buffer was place again onto the column membrane and repeated following the same procedure. All DNA samples were quantified using the Qubit dsDNA HS Assay Kit (ThermoFisher Scientific) following manufacturers guidelines and DNA was stored in a -20°C freezer after quantification, prior to PCR amplification and downstream analysis. A negative control of molecular biology grade H₂O was included in the DNA extraction step.

3.6.3.4 PCR amplification, primers and sequencing

The universal plant ITS2 primer pair UniPlantF (5'-TGTGAATTGCARRATYCMG-3') and UniPlantR (5'-CCCGHYTGAYYTGRGGTDCDC-3'), have been used in this experiment. They have previously been shown to be the most successful at amplifying fragments of 187-237bp (Fig. 3.10) (Moorhouse-Gann *et al.*, 2018). In addition, this primer pair has been shown to co-amplify plant species with both short and

long amplicon lengths in the same PCR mix, preventing size bias. To confirm the results, three biological replicates and three technical replicates were used for each culture.

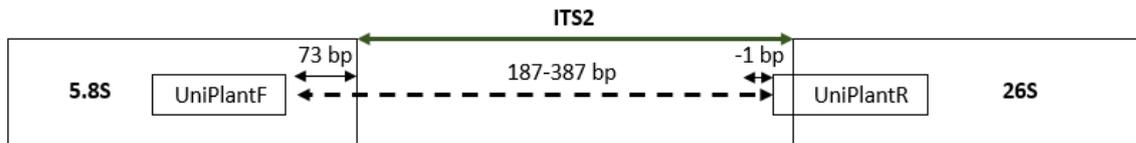


Figure 3.10: A schematic diagram of the priming sites within the second internal transcribed spacer 2 (ITS2), and respective flanking regions (5.8S and 26S). UniPlantF primer is within 5.8S and UniPlantR begins on the last bp of ITS2 and continues into 26S. Schematic adapted from Moorhouse-Gann *et al.*, 2018).

PCR amplification were carried out in 30 μ l reaction volumes containing 3 μ l DNA template (0.3 ng/ μ l), 1 x PCR buffer (x2) (15 μ l), 10.65 μ l of molecular biology grade H₂O and 0.4 μ l (0.2 μ M) of each primer (10 mM). The PCR thermocycler reaction was as follows: denaturation at 95°C for 15 min, 40 cycles of 95°C for 30 secs, 56°C for 30 secs and 72°C for 1 min with a final extension of 72°C for 10 min. A positive control of *L. sativa* host plant DNA and a negative control containing dH₂O as template were used in each PCR plate. PCR products were visualised on a 1% agarose gel to confirm successful amplification prior to being purified using the QIAquick PCR purification Kit. Purified plasmids were sent to Eurofins UK for forward and reverse Sanger sequencing (using the UniPlantF and UniPlantR primers), using their TubeSeq service using an ABI 3730XL (Applied Biosystems).

3.6.3.5 Plant barcode identification

For each returned sequence, the sequences were quality trimmed (with 25 bp window segments in which more than 2 bp with a quality value <20). Quality scores and summary statistics were calculated for each barcode marker. Each contig was manually checked for any base call disparities and edited when appropriate. Sequence assembly was completed using Geneious (v. 10.2.6). For each sequence, quality scores were generated and included mean sequence length, mean sequence quality value (QV), percentage of high-quality bases and mean percentage of low-quality bases (QV < 20) per

sequence. Sequences obtained for each sample were identified to species by comparing them to the available records on the NCBI GenBank nucleotide collection database (nr/nt) using BLAST.

3.6.4 Experiment 4: Overwintering biology

This experiment aimed to observe the colonisation of known winter hosts (*Ribes* spp.) by *N. ribisnigri*. A semi-field overwintering cage was used to observe the dispersal of *N. ribisnigri* to known winter hosts during Autumn from lettuce host plants. Weekly winter observations were conducted to monitor feeding behaviour, egg laying and nymph hatching the follow spring. In addition, lettuce was added into the semi-field cage to observe *N. ribisnigri* dispersal from the overwintering host (*Ribes* spp.) to lettuce during the following spring.

3.6.4.1 Aphid collection

Current-lettuce aphids (*Nasonovia ribisnigri*) were collected in August 2018 from lettuce plants (*Lactuca sativa* cv. Maureen) grown on the organically managed farm near Ely, Cambridgeshire, UK (52°21'N, 0°16'E). Ten whole lettuce plants (with roots) were dug up and these contained ~500 aphids comprising nymphs and apterous and alate morphs (Fig. 3.11). The plants were taken to Rothamsted Research, replanted and placed inside a field cage situated in Great Field 4 (51°80'N, 0°36') (Fig. 3.12).



Figure 3.10: Lettuce plants infested with high numbers of *N. ribisnigri* nymphs (A), and alates (B) found in a lettuce field in August 2018, during an outbreak at Hainey farm (Organic) (Site 1).

3.6.4.2 Field cage for overwintering host plants and lettuce

The field cage consisted of a stainless steel structure (2.5 x 3.5 x 3.5 (h x w x d)) with a fine heavy-duty polyester mesh (<1mm particle size) covering all sides and which was secured into the ground with metal pegs (Fig. 3.12). The lettuce plants containing the *N. ribisnigri* populations were placed in the cage together with different types of currant plant (*Ribes* spp.). Four species of currant were selected for the experiment: blackcurrant (*Ribes nigrum*), white currant (*Ribes rubrum*), red currant (*Ribes rubrum*) and gooseberry (*Ribes uva-crispa*). These were sourced from The Hertfordshire Garden Centre, Hertfordshire. All *Ribes* spp. were 2 years old, ~ 70 cm tall and in separate planters (10 litres).



Figure 3.11: Field cage enclosing primary hosts (*Ribes* spp.) and secondary host (*Lactuca sativa*).

3.6.4.3 Aphid monitoring

To observe colonisation of the *Ribes* plants (primary host), 10 leaves per plant were selected at random from five branches. Each branch was marked with a piece of string and two leaves from each branch were selected to monitor on each monitoring time point (five branches per *Ribes* spp., two leaves per branch (10 monitoring points per *Ribes* spp.)). Monitoring (non-destructive) was carried out and, aphids present (nymphs, apterae, alates and oviparae), were identified and counted with a hand

lens, each week from 2nd October 2018 until 30th April (30-weeks). Recorded *N. ribisnigri* recorded were left *in situ* to allow monitoring of colony development. To minimise disturbance, the cage was only accessed to record aphid numbers and maintain the plants only.

Five temperature loggers (Tinytag (Talk 2)) were placed in different locations; ground (inside cage), ground (outside cage), on a *Ribes* plant, on the outside side of a cage and on the inside of the cage. The hourly mean, minimum and maximum temperatures were recorded from 21st October 2018 until 30th April 2019. These temperature loggers were used to compare both inside and outside ambient temperatures to confirm that no large differences were observed during the experiment.

In the following spring, new lettuce plants from the same cultivar (cv. Maureen) were placed in the overwintering cage to observe whether *N. ribisnigri* would migrate onto this secondary host the following spring. The lettuce plants were planted in small planters and irrigated. Non-destructive monitoring of lettuce plants was conducted twice a week once the first alate *N. ribisnigri* were observed on the winter host (*Ribes* spp.). Monitoring of the lettuce plants involved carefully examining each plant using a hand lens to search for alates, apterae and nymphs.

3.7 Results

3.7.1 Experiment 1: Field margins surveys to identify potential *N. ribisnigri* host plants

3.7.1.1 Vegetation field margin surveys

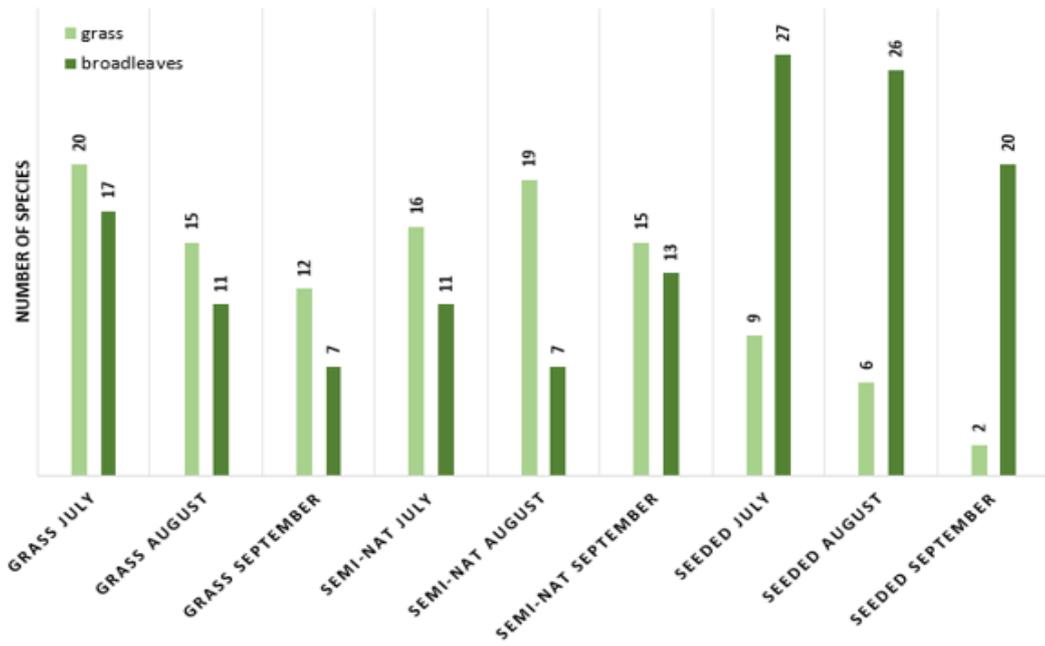
In total, 58 different plant species were identified in the transects on the organically managed farm (Site 1, Hainey). There were 13 species of grass and 45 broadleaved (Forbs) species from 18 families. In the transects on the conventionally managed farm (Site 2, Wissington), there was a total of 47 species; 11 grasses and 36 broadleaved species (Appendix 2).

During the 3-month sampling period, the number of species was highest in June and started to decrease as some plants started to senesce. This was observed in all margin types at both locations (Fig. 3.13 and 3.14). The number of grass species was higher in both the grass and semi-natural

habitats at the Hainey (organic) site whereas; the number of broadleaf species was much higher in the grass and seeded margins at the Wissington (conventional) site. The grass and broadleaf species in the semi-natural margins at Wissington were similar in number and remained relatively unchanged throughout the study period.

The percentage of grass and broadleaf total cover was dominant in the grass and semi-natural margins at Site 1 (Hainey, organic) whereas; the opposite is true for the grass margin at Site 2 (Wissington, conventional). The percentage cover for the seeded mixes at both sites were very similar, with broadleaf species dominating the margins (62%-87%).

A)



B)

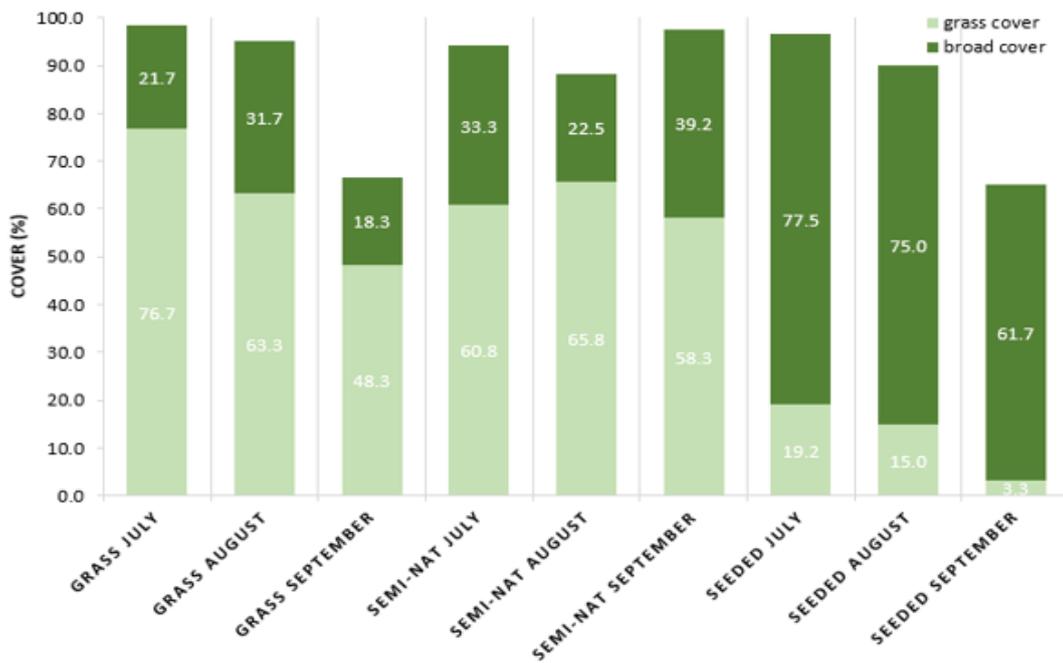


Figure 3.12: A) number of grass (light green) and broadleaf (dark green) species identified in each type of margin (grass, semi-natural, seeded) during the 3-month study at the Hainey (organic) site. Data shown for each margin type are the combined total from two quadrats for each of the three survey locations (six quadrats in total). B) Percentage cover of both grass (light green) and broad leaf (dark green) species in each margin type (grass, semi-natural, seeded) at the Hainey (organic) site.

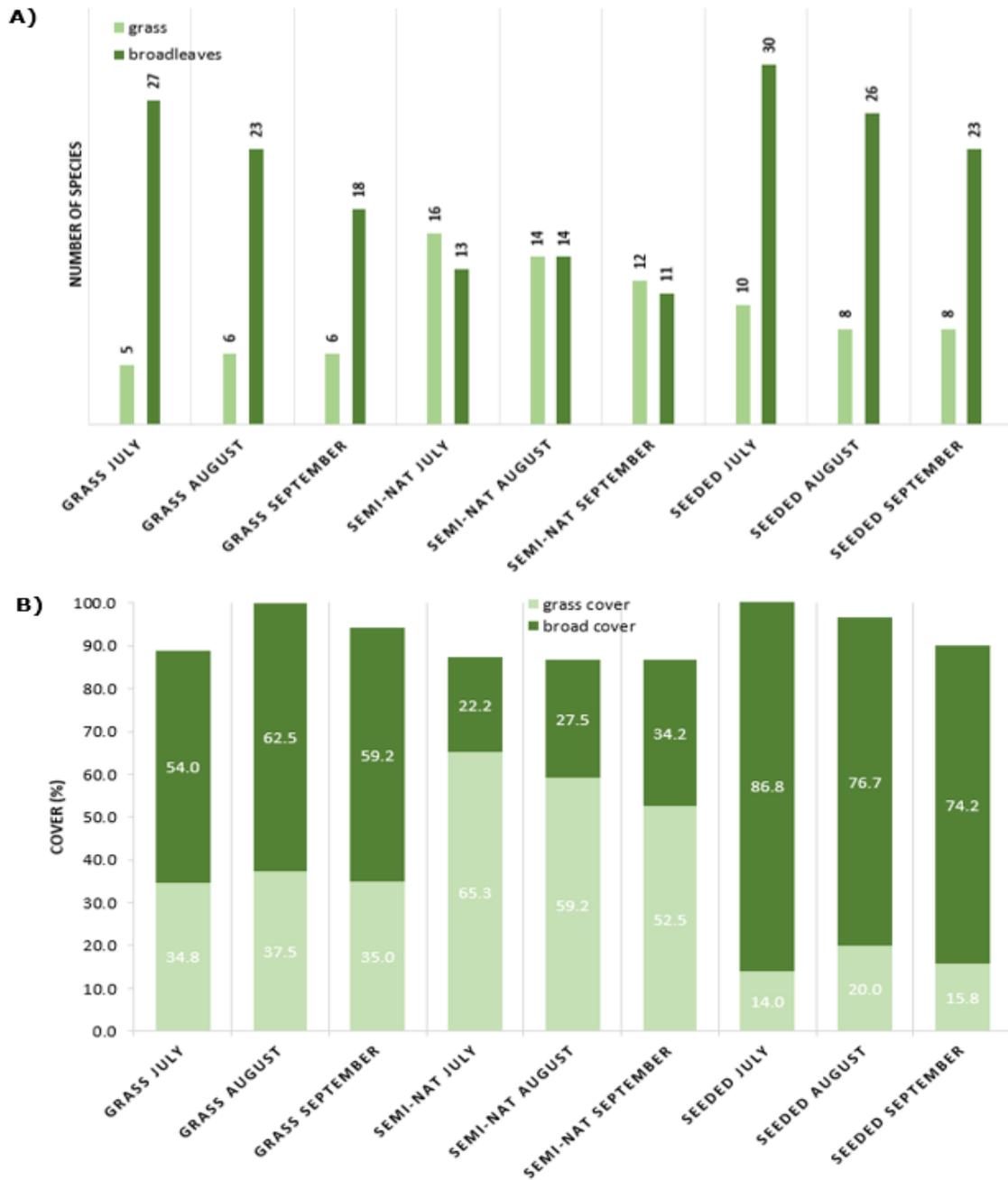


Figure 3.13: A) number of grasses (light green) and broadleaf (dark green) species identified in each margin type (grass, semi-natural, seeded) during the 3-month study at the Wissington (conventional) site. Data shown for each margin type is the combined total of two quadrats for each of the three survey locations (six quadrats in total). B) Percentage cover of both grass (light green) and broad leaf (dark green) species identified in each margin type (grass, semi-natural, seeded) at the Wissington (conventional) site.

The main species of grass to undergo senescence during the 3-month period were crested dogtail (*Cynosurus cristatus*), annual bluegrass (*Poa annua*), fox tail (*Alopecurus pratensis*) and Italian rye grass (*Festuca perennis*). These species were recorded during June only and were absent during the surveys in July and August. For the broadleaf species, prickly sowthistle (*Sonchus asper*), tansy (*Tanacetum vulgare*), white goosegrass (*Chenopodium album*), shepherd's purse (*Capsella bursa-pastoris*) and corn chamomile (*Anthemis arvensis*) were present during June only and were absent from the surveys in July and August. A full list of the species identified at both sites can be found in Appendix 2.

3.7.1.2 Invertebrate assemblages within field margins

Throughout the 3-month study, no *N. ribisnigri* were caught in the sweep net samples taken at either site. A total of 582 aphids were captured from all types of margin during the study.

3.7.2 Experiment 2: Alternative host plant plots to encourage *N. ribisnigri* colonisation

A total of six currant-lettuce aphids (*N. ribisnigri*), were first observed on chicory (*C. intybus*) in one of the plots at the Wissington (conventional) site in Norfolk on 22nd July 2020 (Fig. 3.15). Numbers of both nymphal and adult *N. ribisnigri* started to increase over the next 4 weeks but declined from mid-August. Of the four plots of host plants grown at the site near Wissington (conventional), only two (transects G3 and R54) had presence of *N. ribisnigri*, both on the chicory host plant only. After the mid-summer crash (discussed in Chapter 2), only four *N. ribisnigri* were seen, in one location, on 9th September. During the 27 weeks of monitoring (April-October), no *N. ribisnigri* were seen on any of the other host plants, including the lettuce (*L. sativa*) plants (Table 3.1).

Table 3.1: Total number of *N. ribisnigri* identified on chicory (*C. intybus*) alternative host plant plots.

Site 1 Hainey (organic)					Site 2 Wissington (conventional)						
Date (w/c)	Week	Field				Date (w/c)	Week	Field			
		G3	R54	R20	B(S)			MG	SG	BG	21/8
27-Apr-20	18	0	0	0	0	27-Apr-20	18	0	0	0	0
04-May-20	19	0	0	0	0	04-May-20	19	0	0	0	0
11-May-20	20	0	0	0	0	11-May-20	20	0	0	0	0
18-May-20	21	0	0	0	0	18-May-20	21	0	0	0	0
25-May-20	22	0	0	0	0	25-May-20	22	0	0	0	0
01-Jun-20	23	0	0	0	0	01-Jun-20	23	0	0	0	0
08-Jun-20	24	0	0	0	0	08-Jun-20	24	0	0	0	0
15-Jun-20	25	0	0	0	0	15-Jun-20	25	0	0	0	0
22-Jun-20	26	0	0	0	0	22-Jun-20	26	0	0	0	0
29-Jun-20	27	0	0	0	0	29-Jun-20	27	0	0	0	0
06-Jul-20	28	0	0	0	0	06-Jul-20	28	0	0	0	0
13-Jul-20	29	0	0	0	0	13-Jul-20	29	0	0	0	0
20-Jul-20	30	0	0	0	0	20-Jul-20	30	6	0	0	0
27-Jul-20	31	3	0	0	0	27-Jul-20	31	17	2	0	0
03-Aug-20	32	26	0	0	0	03-Aug-20	32	31	27	0	0
10-Aug-20	33	31	0	0	0	10-Aug-20	33	38	19	0	0
17-Aug-20	34	15	0	0	0	17-Aug-20	34	9	5	0	0
24-Aug-20	35	9	0	0	0	24-Aug-20	35	1	0	0	0
31-Aug-20	36	0	0	0	0	31-Aug-20	36	0	0	0	0
07-Sep-20	37	0	0	0	0	07-Sep-20	37	4	0	0	0
14-Sep-20	38	0	0	0	0	14-Sep-20	38	0	0	0	0
21-Sep-20	39	0	0	0	0	21-Sep-20	39	0	0	0	0
28-Sep-20	40	0	0	0	0	28-Sep-20	40	0	0	0	0
05-Oct-20	41	0	0	0	0	05-Oct-20	41	0	0	0	0
12-Oct-20	42	0	0	0	0	12-Oct-20	42	0	0	0	0
19-Oct-20	43	0	0	0	0	19-Oct-20	43	0	0	0	0
26-Oct-20	44	0	0	0	0	26-Oct-20	44	0	0	0	0

The first sightings of *N. ribisnigri* in the plots on the farm near Hainey (organic) occurred on 27th July 2020, in one transect at Mill Ground (Table 3.1). This transect only had presence of *N. ribisnigri* on chicory (*C. intybus*), with the remaining four alternative plants were absent of *N. ribisnigri*, including lettuce (*L. sativa*).

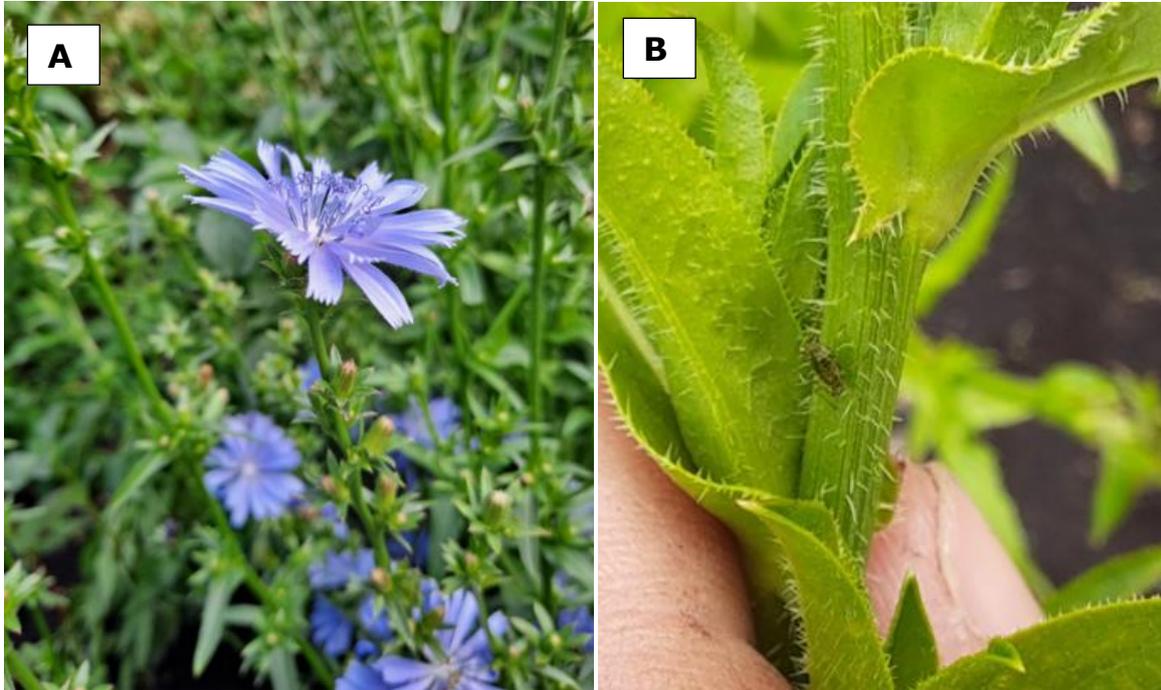


Figure 3.15: A) Transect R54 at the site near Wisington (conventional) with chicory (*C. intybus*) and B) close-up view of an apterous *N. ribisnigri* on *C. intybus* in transect R54 at the same location.

3.7.3 Experiment 3: Using molecular techniques to identify the host plant barcode (ITS2) of *N. ribisnigri* to identify potential green bridge reservoirs

3.7.3.1 Detecting the ITS2 plant barcode in *N. ribisnigri* and the longevity of ITS2 plant barcode within aphids starved up to 48 hours

The sequences obtained in the present study ranged from 200-400bp in length. Megablast search results indicated that the hits with highest similarity were the ITS2 sequences of *Lactuca sativa*, confirming that the lettuce barcode was present and detectable in individual alates up to 2 days post-feeding (Fig. 3.16). All sequences, except the negative control and one sample in the 3hr replicate successfully amplified the ITS2 region in the host plant (*L. sativa*) (Appendix 2, Table 4). Percentage identity varied between 84-92.6% for all samples, with most samples >90% identity.

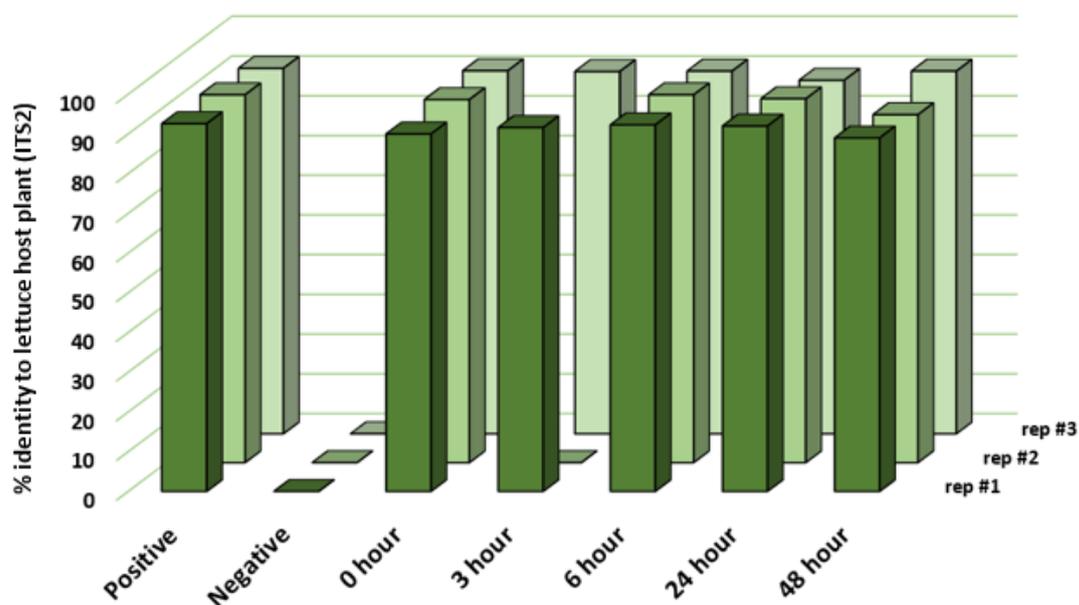


Figure 3.17: Sequence percent identity of the ITS2 lettuce plant barcode barcodes in *N. ribisnigri* after periods of starvation. In this experiment, a lettuce positive control was used throughout DNA extract, PCR and sequencing to verify results. Number of replicates in indicated on the z axis (e.g. rep #1).

Additional parameter refinement was required for one sample (24hr1), in which Asterids (taxid:71274) had to be specified in the NCBI blastn collection to yield meaningful results. Quality score was particularly low for this sequence and is most likely due to low expression and nucleotide discrepancies in the sequence between 321-332bp and 392-400bp (chromatogram data not shown).

3.7.3.2 ITS1 and ITS2 region only detected in aphid midgut

The Uniplant primers failed to amplify the lettuce barcode in the excised heads or stylets of *N. ribisnigri* and successfully amplified the lettuce barcode in all *N. ribisnigri* midgut samples (Table 3.2). This confirms that the ITS2 lettuce barcode is only at detectable levels in the midgut using this method. Sequence identities were comparable to previous sequence results.

Table 3.2: Blastn results from NCBI nucleotide collection database. Sequence length is unclipped length from Sanger sequencing results. No samples out of the 12 sequenced contained any unsure nucleotides (N). ID: ITS1 and 2 *L. sativa* has been abbreviated from figure 3.2. ALB – alate body only; AB – apterous body only; ALH – alate head/stylet only; AH – apterous head/stylet only.

Sample type	Sample	ID	Identity %
Midgut	ALB1	ITS2 <i>L. sativa</i>	88.2
Midgut	ALB2	ITS2 <i>L. sativa</i>	90.0
Midgut	ALB3	ITS2 <i>L. sativa</i>	87.9
Midgut	AB1	ITS2 <i>L. sativa</i>	90.2
Midgut	AB2	ITS2 <i>L. sativa</i>	91.1
Midgut	AB3	ITS2 <i>L. sativa</i>	90.3
Head/stylet	ALH1	-	0
Head/stylet	ALH2	-	0
Head/stylet	ALH3	-	0
Head/stylet	AH1	-	0
Head/stylet	AH2	-	0
Head/stylet	AH3	-	0

3.7.3.3 Applications to field-caught aphid specimens

Of the four *N. ribisnigri* alates caught in the Rothamsted Insect Survey suction trap at Broom's Barn in 2018, three samples failed to amplify any ITS1 and ITS2 plant barcode. One sample amplified the ITS1 and ITS2 region for *Cynosurus cristatus* (crested dogstail), belonging to the Poaceae, with 46% query coverage and 90% identity (Table 3.3). Three out of four *N. ribisnigri* alates caught in the 2m suction trap at Hailey Farm in Cambridgeshire amplified the ITS1 and ITS2 region. One sample provided a positive result for plants in the Geraniaceae (African violet family) but, due to similar query and identity results, genus or species identification was not possible. Two samples returned positive results for the field forget-me-not (*Myosotis arvensis*), with 36% query coverage and 89.3% identity. The fourth sample failed to return any blast results.

Table 3.3: Blastn results from NCBI nucleotide collection database. Sequence length is the unclipped length from the Sanger sequencing results. None of 8 sequenced samples contained any unsure nucleotides (N).

Sample	ID	Identity %
Broom's Barn1	-	0
Broom's Barn2	-	0
Broom's Barn3	<i>Cynosurus cristatus</i>	90
Broom's Barn4	-	0
Hailey1	<i>Myosotis arvensis</i>	89.3
Hailey2	<i>Myosotis arvensis</i>	89.3
Hailey3	-	0
Hailey4	Generiaceae	86-90

3.7.4 Experiment 4: Overwintering biology

Nasonovia ribisnigri alates were first observed to have migrated from the secondary host (lettuce) to the primary host (*Ribes spp.*) on 10th October 2018 (Fig. 3.17). Host acceptance was determined to be when nymphs were observed feeding on the primary host. Blackcurrant (*R. nigrum*) and gooseberry (*R. uva-crispa*) were first plants to be colonised by *N. ribisnigri*, followed by red currant (*R. rubrum*) and white currant (*R. rubrum*). By the end of October (25th October 2018), populations of *N. ribisnigri*

could be found on all the primary hosts, with little variation in total numbers between plants ($F = 4.72$, $P = 0.13$), with the exception of white currant (*R. rubrum*) which had significantly higher numbers of nymphs ($F = 18.34$, $P = 0.01$). Populations of *N. ribisnigri* (nymph, alates and apterae) had declined by early November and numbers of oviparae increased around mid-November (14th). The numbers of oviparae had reduced considerably by the end of November and, by 28th November 2018, no *N. ribisnigri* were observed. Eggs were laid by the oviparae, were black and shiny, and were typically situated either around a bud or in between a bud and a branch. Late bud development of some of the currant host plants likely affected the successful acceptance and colonisation from newly hatched nymphs.

In the following spring, newly hatched nymphs were first observed on 21st March 2019 both on blackcurrant (*R. nigrum*) and red currant (*R. rubrum*) (Fig. 3.18). Their numbers did not start to increase until mid-April (16th), when adults (apterous) were first observed. During the spring monitoring period, *N. ribisnigri* were observed on *R. nigrum* and *R. rubrum* only, which may have been because the buds on the remaining primary hosts developed much later than on these two plants. In the absence of suitable feeding locations for newly hatched nymphs, it is likely that they desiccated rapidly or moved off the unsuitable host plant.

Alates were first observed towards the end of April (25th) which compares well with the first captures of alates in the 12.2m RIS suction trap situated ~0.5km away (29th April 2019). Monitoring of the new lettuce plants commenced during the week that the *N. ribisnigri* alates appeared and continued for 8-weeks. However, no *N. ribisnigri* were observed on any of the lettuce plants, despite their close proximity to the winter host plants.

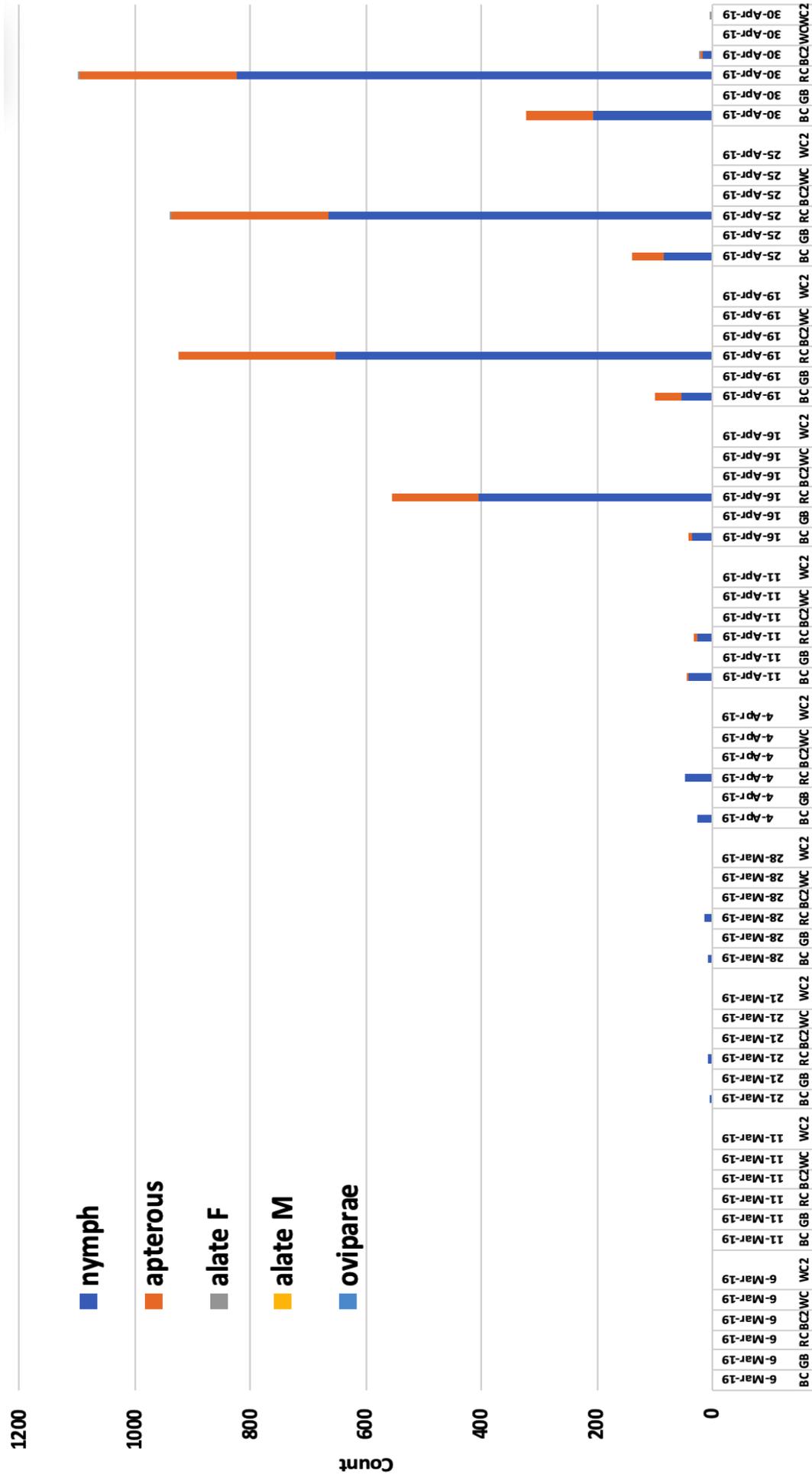


Figure 3.19: Overwintering phenology and colonisation of *N. ribisnigri* on winter host plants (currant – *Ribes* spp.) from summer host (lettuce – *Lactuca sativa* cv. Maureen) between 06MAR19-30APR19. Data from 01JAN19-05MAR19 not shown due to no activity from *N. ribisnigri*. BC - Black currant (*Ribes nigrum*); GB – Gooseberry (*Ribes uva-crispa*); RC – Red currant (*Ribes rubrum*); WC – White currant (*Ribes rubrum*). Alates first observed on 25APR19 on Black currant.

3.8 Discussion

Identification of host plant in individual aphids

Understanding the origins of a migrating insect with regard to its previous host plant is of prime importance for both science and the delivery of effective IPM. This is the first study of its kind to successfully isolate and amplify the host plant barcode (ITS2) from individual aphids. Additionally, it has been shown that the host plant barcode was still detectable 48-hours post-ingestion without the aphid having further access to a host plant. The technique did not amplify the lettuce ITS region in any of the head/stylet samples, both from winged (alate) and wingless (apterous) individuals, while amplification was successful from the bodies of both alate and apterous individuals, confirming that the midgut is the most likely source of the host plant DNA.

The UniPlant primers used in this study have previously been shown to amplify the internal transcribed spacer 2 (ITS2) of 1790 species of plant across 174 families (Moorhouse-Gann *et al.* 2018). The study by Moorhouse-Gann *et al.* was conducted on highly degraded faecal DNA samples originating from reptiles and birds. Of particular interest was the high taxonomic resolution obtained in this study, with 86.1%, 99.4% and 99.9% at the species, genus, and family level, respectively. In the present study, the UniPlant primers were shown to amplify the ITS2 region of the lettuce host plant from *N. ribisnigri* up to 48-hours post-ingestion. This indicates that trace amounts of the host plant were still present and detectable within the aphids up to at least 48 hours. In field-caught specimens, however, it is possible that trace amounts of host plant DNA might not be detectable 48-hours post-feeding due to the energy demands of flight. In the laboratory experiment, aphid movement would have been severely limited and therefore energy expenditure reduced. To determine the longevity of detection of the ITS2 plant barcode in migrating aphids, semi-field experiments would need to be conducted.

Aphids will feed using their stylets but have been shown to alight onto potential host plants and 'probe' them to test for their suitability – prior to feeding (Powell *et al.*, 2006). It is possible that trace amounts of lettuce DNA could persist and be detectable in the aphid stylets and/or salivary glands. To test for this, the head/stylet was separated from the aphid body prior to DNA extraction and

amplification of the barcode. The results from the present study suggest the midgut is the most likely source of the host plant DNA. It is likely that whilst the phloem sap flows through the aphid, it accumulates only within the midgut and little to no plant material is retained within the stylets or salivary glands. This result should be interpreted with caution as viral *Potato virus Y* non-persistent RNA is detectable on individual stylets (Khelifa, 2019). If traces of DNA were detectable within the stylets or salivary glands, this could potentially provide false positive results when applied to field-caught specimens since aphids might probe host plants but not feed on them.

From the field-caught specimens of *N. ribisnigri*, the sequence chromatograms retrieved from the Sanger sequencing were clear, with good quality scores between nucleotide bases visualised in the chromatograms. This provides evidence that only a single host plant was detectable within the midgut at the time of analysis. This supports the hypothesis that if any 'probing' behaviour had happened prior to capture and analysis, the levels of DNA from the plant were insufficient to be detected. Therefore, it is likely that only the ITS plant barcodes from aphids which have alighted, probed and accepted the host plant will be at quantifiable levels. However, these results should be interpreted with caution, as the sample size of the field-caught specimens were very low (8) due to their low catch rate. By testing this method on a more abundant and polyphagous aphid species caught in the suction traps (such as *M. persicae*) would elucidate whether multiple plant barcodes are detected.

The disadvantage of Sanger sequencing is that if multiple sequences were present in the sample, this could lead to low quality scores between bases and therefore erroneous results. DNA metabarcoding and next-generation sequencing (NGS) technologies would allow amplification of more than one ITS plant barcode if present (Coissac *et al.*, 2012; Taberlet *et al.*, 2012). However, the present study was a proof of concept, and only one host plant was used in the laboratory experiment; therefore, Sanger sequencing was more than adequate to confirm if the UniPlant primers were viable. For samples of aphids which could contain material from multiple species of plants, the UniPlant primers would be able to amplify these, with little bias for larger or smaller DNA fragments, and this approach would suit studies where material from multiple plant species would be present (such as dietary gut profiling) (Moorhouse-Gann *et al.*, 2018).

Using the UniPlant method that utilises the ITS barcode, the small pilot study on *N. ribisnigri* captured in 2018 in the RIS 12.2m suction trap at Broom's Barn (Suffolk) and the 2m pop-up suction trap at Hailey (Cambridgeshire) identified two species of secondary host plant species within tissue samples. Crested dogstail, *Cynosurus cristatus* (Poaceae) was identified in one of the aphids from Broom's Barn. Members of the Poaceae have been identified as potential host plants (Diaz and Fereres, 2005) for *N. ribisnigri* but the specific identity of the hosts is still not known. Two of the aphid samples collected at Hailey contained evidence of field forget-me-not (*Myosotis arvensis*), which is a relatively common short-lived herbaceous annual (BSBI, 2021). This plant has not been previously identified as a host plant for *N. ribisnigri* and provides additional insight into another potential intermediate secondary host prior to dispersal onto lettuce. Interestingly, in the survey of field margins *C. cristatus* was observed in both locations (Wissington and Hailey) in June but was absent in subsequent months. An overall decline in a number of grass and broadleaf species was evident in July and August, which is most likely to be because the plants had senesced.

Is *Nasonovia ribisnigri* using an intermediate secondary host prior to dispersal onto outdoor lettuce?

An intermediate host plant is a likely explanation for the sporadic nature in which *N. ribisnigri* can quickly colonise outdoor lettuce crop in late summer but is absent in spring, despite the presence of alates in the RIS suction traps (RIS, 2021). Interestingly, *N. ribisnigri* were only found on chicory (*C. intybus*) throughout the duration Experiment 3, despite the constant availability of lettuce plants in all the plots at both sites. The first signs of the presence of *N. ribisnigri* on *C. intybus* was towards the end of July, with numbers slowly increasing until the 'mid-summer crash' in aphid populations (Karley *et al.*, 2004). It is unclear why the lettuce plants situated in the same plots remained free of *N. ribisnigri*, but it is likely to be because chicory was a preferred host (Powell *et al.*, 2006). Further work would benefit from understanding whether *C. intybus* could be used as a trap crop in a push-pull IPM approach.

Similar behaviour with regard to the use of different host plants has been observed in other sap-feeding hemipterans. For example, *Orthops campestris* (Linnaeus, 1758) (Miridae), is a small mirid bug which feeds on umbellifers (Apiaceae) and can be a severe pest on outdoor celery crops (Anon, 2021; AHDB, 2017). Despite celery being sown as early as March in the UK, *O. campestris* does not become a problem on outdoor celery until mid-late summer (July-September) (Brittain, 1919). The bugs feed on cow parsley (*Anthriscus slyvestris*), wild carrot (*Daucus carota*) and hemlock (*Conium maculatum*), amongst others, during the spring and early summer (Frohlich, 1960). These plants can be used to monitor populations of *O. campestris*, prior to their migration onto celery. Monitoring these plants provides vital information on population development and provides an indication as to when migration onto outdoor celery is likely to occur. A very good indicator is when both *A. slyvestris* and *C. maculatum* start to undergo senescence and adult *O. campestris* are caught in pop-up suction traps and on orange sticky traps (AHDB, 2017; Garrett, unpublished). The biology and life history of this pest of celery is still unclear. However, in the case of *O. campestris*, the transition from Apiaceae to celery occurs mid-late summer which is most likely facilitated by the shift in host plant quality and availability. The present study suggests that *N. ribisnigri* could be using an intermediate secondary host plant prior to subsequent migration onto outdoor lettuce later in the year. This movement from an intermediate secondary host to outdoor lettuce is likely due to lack of other favourable secondary hosts.

Overwinter host and spring migration onto a lettuce alternative host

It is well known that different aphid biotypes preferentially colonise and feed on different host plants, and usually exhibit a preference for a selection of species within their host range (Ma *et al.*, 2019). For example, different biotypes of the English grain aphid, *Sitobion avenae*, have been shown to possess differentially expressed defensive genes directly related to fecundity on alternative host plants (Wang *et al.*, 2020). These differentially expressed genes are likely to have significant effects on the adaptive potential of different biotypes of *S. avenae* to different host plants. In addition, it has been demonstrated that *S. avenae* has a high level of genetic differentiation between populations collected on both cultivated and uncultivated field margins (Vialatte *et al.*, 2005). A much closer genetic

similarity was observed between populations of aphids on host plants belonging to the same tribe and indicated two largely independent biotypes of aphids, despite their proximity. It is unknown whether different biotypes of *N. ribisnigri* exhibit different preferences for host plants. In this instance, the lettuce cultivar used as the secondary host plant in the spring in Experiment 4 was the same cultivar used in the previous autumn, from which these *N. ribisnigri* originated. Application of the UniPlant methodology (Experiment 3) would show whether alates flying in late spring and early summer were likely to have originated from *Ribes* or of from a non-crop secondary host plant. Additionally, incorporating a new molecular tool, Macro Extraction and Determination from Invertebrates (MEDI), could help differentiate between individuals which have undergone long or short migrations by analysing their carbohydrate levels (Cuff *et al.*, 2021). By combining both the MEDI with UniPlant methodologies, it might be possible to elucidate long or short (appetitive) distance migration and identify which host plant an aphid has fed on previously. From integrated pest management (IPM) perspective, this would provide valuable insights to inform future control and mitigative strategies.

In Experiment 4, an alternative secondary host (lettuce) was present from the first week that alate *N. ribisnigri* were seen on the primary hosts (*Ribes* plants) (date), to monitor migration from the primary winter host to the secondary summer host. Despite the proximity of the lettuce plants, no *N. ribisnigri* were observed on any of the lettuce plants during the monitoring period. After 8-weeks (date), neither the currant plants nor the lettuce plants showed any signs of infestation by *N. ribisnigri* and the experiment was terminated. Since *N. ribisnigri* generally does not become a problem on outdoor lettuce crops until late summer (July onwards) and first flight occurs in April, it is possible that *N. ribisnigri* migrates from a winter host plant onto an intermediate, non-crop, secondary host and then onto lettuce crops. In this case the non-crop host would be acting as a 'green bridge' for *N. ribisnigri*. It is possible that *N. ribisnigri* could remain on a primary winter host until mid-summer and then migrate to outdoor lettuce. Experiment 4 was a small, unreplicated experiment and so this possibility cannot be ruled out. However, this seems unlikely as migrations by most host-alternating aphids occur in spring (Blackman and Eastop, 1984; Dixon and Kundu, 1994). Indeed, in Experiment 4 the numbers of *N. ribisnigri* on the *Ribes* plants which had developed colonies declined during May and neither nymphs nor adults could be found on these plants in June.

Applications to field-caught specimens

Since it has been difficult to capture *N. ribisnigri* in significant numbers in a variety of entomological traps (see Chapter 2), coupled with their low abundance in field margins and on non-crop host plants, the UniPlant method could provide vital insights into the nature of the primary and secondary host plants of *N. ribisnigri* and potentially reveal the existence of different biotypes. For example, by running traps in lettuce crops and/or collecting alates from lettuce plants and using the UniPlant method, it might be possible to determine any non-crop secondary hosts that the species uses prior to migration onto lettuce. This information could inform future IPM strategies aimed at reducing the availability of potential green bridges and reservoirs for *N. ribisnigri*.

Other potential applications of this method could clarify potential transmission pathways for insects which vector plant viruses. For example, combining this technique with current plant virus identification methods, such as Taq-Man assays, would help to elucidate the main transmission pathways involving an aphid, a virus, and the associated host plant.

3.9 Conclusion

In conclusion, *N. ribisnigri* failed to establish on outdoor lettuce despite the proximity from the overwintering host. This supports the hypothesis that *N. ribisnigri* is using an intermediate secondary host prior to transitioning onto lettuce. Further research into whether chicory has the potential to be used as a trap crop in a push-pull IPM approach would need to be conducted. The UniPlant method has been shown to successfully amplify the host plant ITS2 barcode within individual aphids and remains at detectable levels 48-hours post feeding under laboratory conditions. The aphid midgut is thought to be the source of the trace host plant DNA elements as no host plant DNA was amplified in either the stylet or salivary glands. This technique was shown to successfully amplify and identify the previous host plants of field-caught *N. ribisnigri* and has highlighted other potential host plants used by *N. ribisnigri*. With the increasing threat of crop viruses, the applications of this molecular diagnostic

tool could provide a vital link in understanding pest-plant interactions and elucidate currently unknown transmission pathways of insect vectored plant viruses. It is another tool that could be used to improve our knowledge of understanding host plant green-bridges and reservoirs, especially in cryptic species such as *N. ribisnigri*. Providing definitive answers of the host plant species that an insect is feeding on could improve our understanding of the autecology of vulnerable species and promote conservation management strategies.

3.10 References

- AHDB, 2017. FV 441 Project Reports: Celery: Investigation of strategies to control mirid bugs in outdoor crops. Available at: https://projectblue.blob.core.windows.net/media/Default/Research%20Papers/Horticulture/FV%20441_Report_Final_2017.pdf (Accessed: 12 November 2021).
- Anon, 2021. *Orthops campestris* – British Bugs website. Available at: http://www.britishbugs.org.uk/heteroptera/Miridae/orthops_campestris.html (Accessed: 20 November 2021).
- Batovska, J., Piper, A.M., Valenzuela, I., Cunningham, J.P. and Blacket, M.J., 2021. Developing a non-destructive metabarcoding protocol for detection of pest insects in bulk trap catches. *Scientific reports*, pp.1-14.
- Ben-Issa, R., Gomez, L. and Gautier, H., 2017. Companion plants for aphid pest management. *Insects*, p.112.
- Blackman, R.L. and Eastop, V.F., 1984. Aphids on the World's Crops, An Identification Guide. First edn. Chichester: John Wiley & Sons.
- Brittain, W.H., 1919. Notes on *Lygus campestris* Linn. *Nova Scotia. Proc. Ent. Soc. Nova Scotia*, pp.76-81.
- BSBI., 2021. Botanical Society of Britain & Ireland. Available at: <https://bsbi.org/maps?taxonid=2cd4p9h.90y> (Accessed: 28/11/2021).
- Castañera, P., Loxdale, H.D. and Nowak, K., 1983. Electrophoretic study of enzymes from cereal aphid populations. II. Use of electrophoresis for identifying aphidiid parasitoids (Hymenoptera) of *Sitobion avenae* (F.)(Hemiptera: Aphididae). *Bulletin of Entomological Research*, pp.659-665.
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X. and Luo, K., 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS one*, p.e8613.
- Coissac, E., Riaz, T. and Puillandre, N., 2012. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular ecology*, pp.1834-1847.
- Cook, S.M., Khan, Z.R. and Pickett, J.A., 2007. The use of push-pull strategies in integrated pest management. *Annu. Rev. Entomol.*, pp.375-400.
- Cuff, J.P., Wilder, S.M., Tercel, M.P., Hunt, R., Oluwaseun, S., Morley, P.S., Badell-Grau, R.A., Vaughan, I.P., Bell, J.R., Orozco-terWengel, P. and Symondson, W.O., 2021. MEDI: Macronutrient

Extraction and Determination from invertebrates, a rapid, cheap and streamlined protocol. *Methods in Ecology and Evolution*, pp.593-601.

Diaz, B.M. and Fereres, A., 2005. Life table and population parameters of *Nasonovia ribisnigri* (Homoptera: Aphididae) at different constant temperatures. *Environmental entomology*, pp.527-534.

Diaz, B.M., Barrios, L. and Fereres, A., 2012. Interplant movement and spatial distribution of alate and apterous morphs of *Nasonovia ribisnigri* (Homoptera: Aphididae) on lettuce. *Bulletin of entomological research*, pp.406-414.

Dixon, A.F. and Kundu, R.A.N.A.J.I.T., 1994. Ecology of host alternation in aphids. *European Journal of Entomology*, pp.63-70.

Dixon, A.F.G. 1977. Aphid Ecology: Life cycles, polymorphism, and population regulation. *Annual Review of Ecology & Systematics*, pp.329-353.

Dong, W., Liu, J., Yu, J., Wang, L. and Zhou, S., 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS one*, p.e35071.

Favetti, B.M., Braga-Santos, T.L., Massarolli, A., Specht, A. and Butnariu, A.R., 2017. Pearl millet: a green bridge for lepidopteran pests. *Embrapa Cerrados-Artigo em periódico indexado (ALICE)*. Frohlich, G., 1960. Life histories and control of *Lygus campestris* and *L. kalmi* on fennel. *Pharmazie*, pp.334-41.

Frohlich, G., 1960. Life histories and control of *Lygus campestris* and *L. kalmi* on fennel. *Pharmazie*, pp.334-41.

Gielly, L. and Taberlet, P., 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. *Molecular biology and evolution*, pp.769-777.

Hand, S.C. and Hand, L., 1986. Monitoring of the winter populations of cereal aphids near Wageningen, the Netherlands, in 1982/1983. *Netherlands Journal of Plant Pathology*, pp.137-146.

Harrington, R., Hullé, M. and Plantegenest, M., 2007. *Monitoring and forecasting* (Wallingford: CABI), pp. 515-536.

Hough, G. and Collier, R., 2014. The biology and control of the currant lettuce aphid *Nasonovia ribisnigri*. *The biology and control of the currant lettuce aphid Nasonovia ribisnigri.*, pp.51-55.

Hough, G. L., 2013. *Biology and Control of Currant Lettuce Aphid Nasonovia ribisnigri*, University of Warwick.

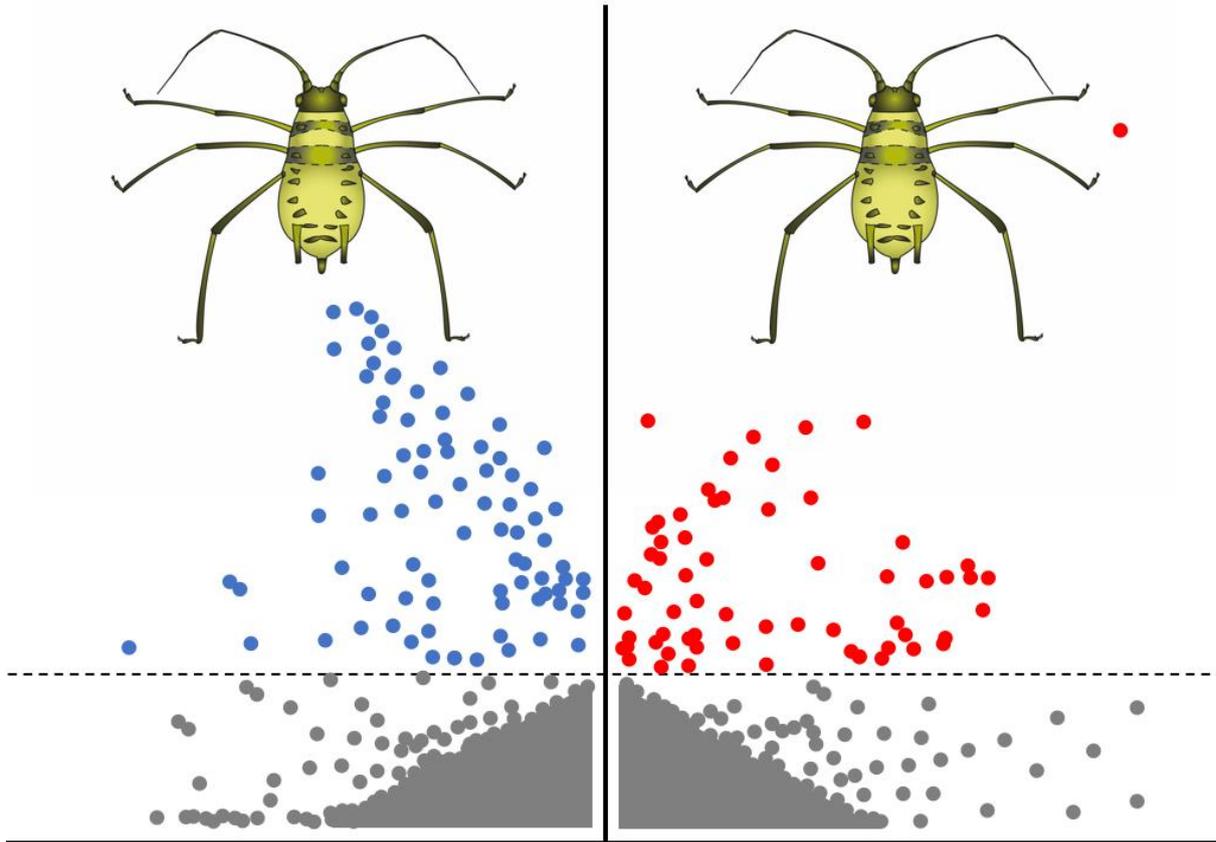
Karley, A.J., Parker, W.E., Pitchford, J.W. and Douglas, A.E., 2004. The mid-season crash in aphid populations: why and how does it occur?. *Ecological Entomology*, pp.383-388.

Khelifa, M., 2019. Detection and quantification of Potato virus Y genomes in single aphid stylets. *Plant disease*, pp.2315-2321.

- Lafontaine, D.L. and Tollervey, D., 2001. The function and synthesis of ribosomes. *Nature Reviews Molecular Cell Biology*, pp.514-520.
- Liu, Z., Zeng, X., Yang, D., Chu, G., Yuan, Z. and Chen, S., 2012. Applying DNA barcodes for identification of plant species in the family Araliaceae. *Gene*, pp.76-80.
- Ma, L., Li, M.Y., Chang, C.Y., Chen, F.F., Hu, Y. and Liu, X.D., 2019. The host range of *Aphis gossypii* is dependent on aphid genetic background and feeding experience. *PeerJ*, p.e7774.
- Macaulay, E.D.M., Tatchell, G.M. and Taylor, L.R., 1988. The Rothamsted Insect Survey '12-metre' suction trap. *Bulletin of Entomological Research*, pp.121-128.
- Maling, T., Diggle, A.J., Thackray, D.J., Siddique, K.H.M. and Jones, R.A.C., 2008. An epidemiological model for externally sourced vector-borne viruses applied to Bean yellow mosaic virus in lupin crops in a Mediterranean-type environment. *Phytopathology*, pp.1280-1290.
- Moorhouse-Gann, R.J., Dunn, J.C., De Vere, N., Goder, M., Cole, N., Hipperson, H. and Symondson, W.O., 2018. New universal ITS2 primers for high-resolution herbivory analyses using DNA metabarcoding in both tropical and temperate zones. *Scientific Reports*, pp.1-15.
- Moran, N.A., 1988. The evolution of host-plant alternation in aphids: evidence for specialization as a dead end. *The American Naturalist*, pp.681-706.
- Powell, G., Tosh, C.R. and Hardie, J., 2006. Host plant selection by aphids: behavioral, evolutionary, and applied perspectives. *Annu. Rev. Entomol.*, pp.309-330.
- Random, 2019. Random.org. Available at: www.random.org/list (Accessed: 10/01/2020).
- Reinink, K., Dieleman, F. L., 1993. Survey of aphid species on lettuce. IOBC/WPRS Bull., 56-68.
- RIS., 2021. The Insect Survey, Rothamsted Research. Available at: <https://www.rothamsted.ac.uk/insect-survey> (Accessed: 21/11/2021).
- Sauer-Kesper CL, Lucia N, Buser H and Vogler U., 2011. Bedeutung und Verbreitung des neuen Biotyps-Nr:1 der Grünen Salatlaus in der Deutschschweiz. *Agrarforschung Schweiz* 10: pp.462–469.
- Sharma, A., Shrestha, G. and Reddy, G.V., 2019. Trap crops: how far we are from using them in cereal crops?. *Annals of the Entomological Society of America*, pp.330-339.
- Siddiqui, W.H., Barlow, C.A. and Randolph, P.A., 1973. Effects of some constant and alternating temperatures on population growth of the pea aphid, *Acyrtosiphon pisum* (Homoptera: Aphididae). *The Canadian Entomologist*, pp.145-156.
- Simon, J.C., Carre, S., Boutin, M., Prunier-Leterme, N., Sabater-Muñoz, B., Latorre, A. and Bournoville, R., 2003. Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, pp.1703-1712.

- Smallshire, D., Robertson, P. and Thompson, P., 2004. Policy into practice: the development and delivery of agri-environment schemes and supporting advice in England. *Ibis*, pp.250-258.
- Stufkens, M.A.W., and D.A.J. Teulon., 2003. Distribution and host range of the lettuce aphid in New Zealand and its effect on the lettuce industry. *N. Z. Plant Prot.*, pp.27-32.
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, pp.2045-2050.
- Thackray, D.J., Diggle, A.J. and Jones, R.A.C., 2009. BYDV PREDICTOR: a simulation model to predict aphid arrival, epidemics of Barley yellow dwarf virus and yield losses in wheat crops in a Mediterranean-type environment. *Plant Pathology*, pp.186-202.
- Toju, H. and Baba, Y.G., 2018. DNA metabarcoding of spiders, insects, and springtails for exploring potential linkage between above-and below-ground food webs. *Zoological letters*, pp.1-12.
- Traugott, M., Bell, J.R., Broad, G.R., Powell, W., Van Veen, F.J.F., Vollhardt, I.M.G. and Symondson, W.O.C., 2008. Endoparasitism in cereal aphids: molecular analysis of a whole parasitoid community. *Molecular ecology*, pp.3928-3938.
- Utzeri, V.J., Schiavo, G., Ribani, A., Tinarelli, S., Bertolini, F., Bovo, S. and Fontanesi, L., 2018. Entomological signatures in honey: an environmental DNA metabarcoding approach can disclose information on plant-sucking insects in agricultural and forest landscapes. *Scientific reports*, pp.1-13.
- van der Arend, A.J., 2003. The possibility of *Nasonovia ribisnigri* resistance breaking biotype development due to plant host resistance: a literature study. *Eucarpia leafy vegetables*, pp.75-81.
- Vialatte, A., Dedryver, C.A., Simon, J.C., Galman, M. and Plantegenest, M., 2005. Limited genetic exchanges between populations of an insect pest living on uncultivated and related cultivated host plants. *Proceedings of the Royal Society B: Biological Sciences*, pp.1075-1082.
- Walton, M.P., Powell, W., Loxdale, H.D. and Allen-Williams, L., 1990. Electrophoresis as a tool for estimating levels of hymenopterous parasitism in field populations of the cereal aphid, *Sitobion avenae*. *Entomologia Experimentalis et Applicata*, pp.271-279.
- Wang, D., Liu, D., Shi, X., Yang, Y., Zhang, N. and Shang, Z., 2020. Transcriptome profiling revealed potentially important roles of defensive gene expression in the divergence of insect biotypes: a case study with the cereal aphid *Sitobion avenae*. *BMC genomics*, pp.1-15.

Chapter 4



CHAPTER 4: Genome assembly and transcriptomic analyses to study the resistance-breaking mechanisms in host plant resistant biotypes to *Nasonovia ribisnigri*

4.1 Introduction

Aphids are an economically important insect group in horticulture worldwide (Van Emden and Harrington, 2017). They are the most important and widespread vectors of plant viruses causing serious damage to many crops throughout the world (Nault, 1997). The sequencing of genomes of horticulturally important aphids is fundamental to facilitate the study of the genetic basis of the aphid-plant-virus interactions (Biello *et al.*, 2020; Jiang *et al.*, 2019; Mathers *et al.*, 2020; Tagu *et al.*, 2008). The assembly of genomes, in particular aphids, enables researchers to understand the complex life history traits of organisms which have a high degree of plasticity and provides information on symbiosis, vector biology and the genetics of host adaption and speciation (Brisson and Stern, 2006).

In recent years, with the advances in sequencing technology, computational power, and the reduction of associated costs, generating reference genomes is becoming routine. In 2010, the only aphid genomic model available to study aphid genetics and aphid-plant interactions was the pea aphid (*Acyrtosiphon pisum*) (International Aphid Genomics Consortium, 2010). This was a milestone for aphid genomics and the genome is still used 11 years later in a variety of genomic comparative studies (Biello *et al.*, 2021; Li *et al.*, 2019; Mathers, 2020; Shahid *et al.*, 2021). Since the *A. pisum* genome, there are now 28 published genomes for 12 different aphid species, ranging from draft assemblies to chromosome-level assembly and annotation (BIPAA, 2021). Along with the increasing number of genomes being assembled and annotated, researchers now have excellent genomic resources to address a multitude of fundamental questions, such as aphid evolution and adaption, the impact of virus acquisition on individuals and the modes of resistance to insecticides (Li *et al.*, 2020; Singh *et al.*, 2021; Troczka *et al.*, 2021). Genome assemblies are also continuously being improved with additional

sequencing data, technology and new information. A recent paper highlighted the importance of revisiting draft assemblies and incorporated a k-mer based duplication pipeline to improve the accuracy of a previous assembled and annotated genome (estimated genome size of 375 Mb and 27,678 protein coding genes) of the black cherry aphid, *Myzus cerasi* (Mathers *et al.*, 2020).

The currant-lettuce aphid, *Nasonovia ribisnigri*, is a major pest of outdoor lettuce (*Lactuca sativa* L.) and is responsible for large economic losses (Diaz *et al.* 2012; Liu and McCreight, 2006; Morales *et al.*, 2013). The nymphs and adults preferentially feed on young developing leaves of lettuce, so they predominantly reside in the centre of the lettuce head, making foliar insecticide application ineffective (Aarts *et al.*, 1999). This is particularly a problem in well-developed lettuce plants, in which the older leaves surround feeding aphids, and provides resources for larger colonies of *N. ribisnigri* leading to contamination of the harvested crop and thus economic loss. Biological control by both predatory insects and parasitoids is also hindered by their location in the centre of the lettuce. Regardless of this, with whole lettuce plants being harvested for consumption, even the presence of beneficial insects is considered a contamination issue by retailers, so reducing the potential for biological control and integrated pest management in field grown lettuce. The use of insecticides on a young lettuce crop, in which the leaves are still open, does enable foliar sprays some effectiveness against *N. ribisnigri* (Natwick and Lopez, 2016). Unfortunately, it is often during the later stages of growth, once the lettuce heads are developed, and soon to be harvested, when the level of infestation passes the quality control threshold. By exploiting this host plant niche, *N. ribisnigri* causes the most damage for lettuce growers compared to other lettuce feeding aphids (Reinink and Dieleman, 1993). With the decreasing number of insecticides being developed and the increasing restrictions on products and applications, the reliance on a small number of active ingredients is leading to the rapid development of insecticide-resistant biotypes, as has been shown for *Myzus persicae* (Bass *et al.*, 2014; Puinean *et al.*, 2010; Singh *et al.*, 2021).

The inherent ability of aphids to reproduce both parthenogenetically and sexually, combined with their telescopic reproduction, provides aphids with enough lifecycle plasticity to overcome selection pressures rapidly (Görür, 2000; Lombart *et al.*, 2009; Smith and Chuang, 2014). Within populations of

aphids, the selective pressure of insecticides causes the frequency of resistance gene(s) to increase (Barber, 2002). It has been shown that *N. ribisnigri* has developed resistance to commonly used insecticides such as pyrethroids and pirimicarb (Kift *et al.* 2004). Strains of five biotypes collected from Almería (south of Spain) and Perpignan (south of France) had a demonstrable high-level of resistance to endosulfan, deltamethrin, pirimicarb, methomyl and acephate (Rufingier *et al.* 1997). Furthermore, the banning of neonicotinoid seed-coating treatment in the UK for non-flowering crops in 2020 will further reduce the methods of control available for *N. ribisnigri* and will put additional pressure on conventional growers of lettuce and other crops (Dewar, 2017).

To tackle the complex issues concerning insecticides and the changing legislation pressures, a lettuce cultivar resistant to *N. ribisnigri* (Nr) was introduced into the growers' repertoire during the early 1980s. The single gene resistance (Nr-gene) originated from a wild species of lettuce (*Lactuca virosa*) that showed natural resistance to *N. ribisnigri* and it was introgressed into lettuce via an interfertile bridging species (*Lactuca serriola*) (Eenink *et al.* 1982). This single dominant gene conferred near complete resistance to *N. ribisnigri* and was incorporated into many lettuce cultivars to reduce the loss of lettuce crops to this aphid. The resistance in *N. ribisnigri* associated with the Nr-gene is unlikely to be mechanical or due to a change in host plant quality (antibiosis), as other aphid species would be likely also affected, so the resistance mechanisms involved are most likely species-specific (Tjallingii and Esch, 1993). Research has shown that the Nr-gene has no effect on other species of lettuce feeding aphids, with complete susceptibility to *M. euphorbiae* and partial resistance to *M. persicae* (Reinink and Dieleman, 1989). The variation of partial resistance observed in three lines of *M. persicae* is most likely a result of the interaction of the Nr-gene with additional genes, moreover the Nr-gene itself (Reinink and Dieleman, 1989). It has been demonstrated, using the electrical penetration graph (EPG) technique, that the susceptible (Nr:0) aphids are able to penetrate the lettuce tissue but are unable to feed and subsequently leave the unsuitable plant (van Helden, 1995). Thus, the source of resistance is thought to reside in the phloem of *L. sativa* and to act during sap ingestion from the sieve element, although the exact resistance mechanism or pathway involved is still unknown (ten Broeke *et al.*, 2013).

The use of resistant cultivars is an effective method for managing many pest insects and is incorporated into a range of integrated pest management (IPM) programmes to ensure that they are more resilient and less reliant on synthetic pesticides (Barriere *et al.*, 2015; Lin, 2011). The ecological benefits of incorporating resistant cultivars are that this strategy will reduce the reliance on insecticides, improve water quality by reducing pollution, and reduce the mortality of non-target and beneficial arthropods. The inherent practises in modern crop production for each plant to be genetically identical to one another, enables a grower to obtain uniform growth characteristics and in turn, cultivation and management practises remain the same. Unfortunately, these monocultures also benefit pest insects and pathogens which can move easily between plants and devastate entire fields of crop in the right conditions (Segoli and Rosenheim, 2012; Tooker and Frank, 2012). As the Nr resistance in lettuce cultivars is conferred by a single gene, and since there are few other effective control strategies, the overuse and reliance on the Nr-gene in lettuce cultivars resulted in the first reports of resistance breakdown in early 2000 (Van der Arend, 2003). This led to the identification of so-called resistance-breaking (Rb) biotypes of *N. ribisnigri* which could overcome the Nr-resistance of lettuce cultivars. In both France and Germany during 2007, these new Rb biotypes were identified on resistant cultivars and subsequently, a Rb biotype was identified in Kent during 2009 (Hough, 2013). It has been demonstrated that insecticide-resistant biotypes of *N. ribisnigri* and others that break host-plant resistance (Rb) have been present in UK aphid populations and in the EU (Barber *et al.* 1999; Broeke *et al.* 2013; Rufingier *et al.* 1997). In the absence of lettuce cultivars that exploit new resistance mechanisms, the lack of effective pesticides, and with the increasing demand for fresh salads, *N. ribisnigri* infestations are predicted to become more frequent and have greater impact on the availability of leafy salads to the consumer (Van der Arend, 2003).

The genome for *N. ribisnigri* has not been sequenced to date, and only the mitochondrial cytochrome c oxidase subunit I (COI) barcode and a selection of odorant binding proteins (OBP) have been sequenced. This project aimed to generate two *de novo* assembled reference genomes for the currant-lettuce aphid, *N. ribisnigri*, corresponding to both a biotype susceptible to host plant resistance (Nr:0) and a resistance-breaking biotype (Nr:1). Secondly, changes in gene expression of two *N. ribisnigri* biotypes (Nr:0 and Nr:1) were investigated using transcriptomic analyses of RNA-seq data to study

potential mechanisms of resistance to the *Nr*-gene in lettuce. By providing two draft genome assemblies of *N. ribisnigri* and an transcriptomics analysis of a susceptible and resistant biotype, this chapter will provide additional insight and knowledge to understand the genetic basis of *N. ribisnigri* host-plant resistance breaking, bringing us closer to elucidating the resistance mechanisms involved.

4.2 Materials and methods

4.2.1 Genome assembly and annotation

4.2.1.1 Aphid cultures, biotypes and near isogenic lines

All aphids used for the genome sequencing were obtained from Warwick Crop Centre (Warwickshire, UK) or field-collected and subsequently cultured at Rothamsted Research (Harpenden, UK) and reared parthenogenetically in a laboratory on whole plants of *Lactuca sativa* cv. 'Auvona RZ', a variety that does not contain the *Nr*-gene resistance. The *N. ribisnigri* strain Nr8 originates from a clone collected from a field in Yorkshire in 1999 and cultured at Warwick Crop Centre in a laboratory at a constant temperature of 21°C with a 16:8h (light:dark) photoperiod. Strains Kent wt and Kent CL were collected from two different fields in Kent during 2010-2011 and were also subsequently cultured at Warwick Crop Centre. The Ely culture was collected in a field containing resistant lettuce varieties (containing the *Nr*-gene) at G's Fresh (Cambridgeshire, UK) in 2018. All clones were determined and validated to be resistant or susceptible by subsequent culturing (same conditions as above) on resistant (*Nr*) varieties (cv. Eluarde) at Rothamsted Research prior to DNA extraction and downstream pipelines.

To create isogenic cultures for each biotype used in the experiments, one female from each strain was placed on a single plant and left to establish a new culture (Table 4.1). Each *N. ribisnigri* line derives from different starting females from the original isogenic culture reared at Rothamsted Research. Once ~10 nymphs had been produced, the founding mother was removed and stored in 100% ethanol. Each of the lines listed below are the cultures and isogenic lines that were used for the sequencing experiments. A total of ~20 individuals from each isogenic line were used to obtain enough high molecular weight DNA for downstream sequencing.

Table 4.1: Near isogenic culture lines of *Nasonovia ribisnigri* which were used for DNA extraction and *de novo* genome assembly. Susceptible = not resistant to the Nr-gene in lettuce varieties (Nr:0); resistant = resistant to the Nr-gene in lettuce varieties (Nr:1). Isogenic lines are indicated numerically i.e. (1).

Culture	Origin	Biotype
Nr8 (1)	Warwick Crop Centre	Nr:0 susceptible
Nr8 (2)	Warwick Crop Centre	Nr:0 susceptible
Nr8 (3)	Warwick Crop Centre	Nr:0 susceptible
WT Kent (1)	Warwick Crop Centre	Nr:0 susceptible
WT Kent (2)	Warwick Crop Centre	Nr:0 susceptible
WT Kent (3)	Warwick Crop Centre	Nr:0 susceptible
UK631 (1)	Warwick Crop Centre	Nr:1 resistant
UK631 (2)	Warwick Crop Centre	Nr:1 resistant
Kent CL (1)	Warwick Crop Centre	Nr:1 resistant
Kent CL (2)	Warwick Crop Centre	Nr:1 resistant
Ely (1)	Rothamsted Research	Nr:1 resistant
Ely (2)	Rothamsted Research	Nr:1 resistant

4.2.1.2 DNA extraction

Nasonovia ribisnigri DNA was extracted using a QIAGEN® Genomic-tip (20/G) extraction kit, specifically designed for isolating high molecular weight DNA for whole genome sequencing (WGS). An adapted User-Developed Protocol for mosquitoes (<https://www.qiagen.com/us/resources/download.aspx?id=b45c3cc3-7f2b-4f4a-aa37-21d814ed3730&lang=en>) and other insects was utilised to help increase overall DNA yield as follows: 10-20 isogenic aphids were placed into a 1.5 ml microcentrifuge tube, submersed in liquid nitrogen for 10 seconds and homogenised using a sterile pestle. Subsequently, 1ml of genomic-tip lysis buffer with DNase-free RNase A (to remove any RNA contaminants) was added to suspend the samples and incubated for 30 minutes at 37°C, followed by Proteinase K (0.8mg/ml) (removing any protein contaminants) and then incubated for two hours at 50°C. This solution was centrifuged for 20 minutes at 15,000 rpm (21168 g) to pellet the insoluble debris. The clarified lysate was transferred to a Genomic-tip. Genomic-tips are gravity fed and therefore can take time to flow-through, but this method is less aggressive than spin columns and enables higher molecular weight DNA to be obtained (Mayjonade *et al.*, 2016). The Genomic-tips were washed four times with a wash buffer and eluted into 2ml Eppendorfs with 1 ml elution buffer. The addition of an equal volume of 100% isopropanol to

the samples was used to precipitate DNA, which was centrifuged for 20 minutes at 15,000 rpm (21168 g), washed with 70% ethanol and centrifuged again at 10 minutes at 5,000 rpm (2352 g). Finally, the DNA pellet was left to dry for ~ 1 minute, resuspended in an appropriate volume of molecular grade water and placed into an orbital shaker for 2 hours, with gentle agitation to dissolve the DNA pellet. The DNA was quantified using a Qubit fluorometer (dsDNA HS Assay Kit, Thermo Fisher Scientific), visualised for quality on an electrophoresis gel and stored in a -20°C freezer for later down-stream processes.

4.2.1.3 Sequencing

DNA that passed inhouse quality checks (Qubit BR Assay Kit (ThermoFisher Scientific) and electrophoresed on a 1% agarose gel)) was sent to Novogene (Cambridge Science Park, Cambridge, UK) to generate Illumina sequence data using x20 HiSeq-PE 150. Some previous Illumina sequence data was provided by Warwick Crop Centre from a resistant biotype (Kent CL (Nr:1 resistant)) and this was incorporated into the assembly of the Kent CL genome. To complement the Illumina sequence data, long read sequence data was generated using a MinION (Oxford NANOPORE Technologies) at Rothamsted Research. Two SpotON flow cells (R9.4.1) were used to produce both resistant (Nr:1) and susceptible (Nr:0) long read sequence data to improve genome assembly consensus for all constructed genomes. High molecular weight *N. ribisnigri* DNA was extracted following the method above (section 4.2.1.2), following the Genomic DNA by Ligation (SQK-LSK109) protocol for library preparation (Oxford NANOPORE Technologies) which includes DNA repair and end-prep, adapter ligation and clean-up followed by priming and loading the SpotON flow cell. For each run, approximately 1µg of *N. ribisnigri* genomic DNA was loaded into each SpotON flow cell. The genomic DNA samples were sequenced using MinKNOW software on a Windows based computer. Kent CL was used to generate the resistant sequence data and Nr8 was used to produce the susceptible (wild type) sequence data.

4.2.1.4 Genome assembly

Prior to assembly, Fast QC (Andrews, 2019) was used to check the quality of all Illumina sequence data using the in-house Galaxy server (Afgan *et al.*, 2018). To utilise the different sequence technologies generated in this study, MaSuRCA (Maryland Super Read Cabog Assembler) was used to create a hybrid assembly for each near isogenic line and biotype (see Appendix 1 for full commands) The advantage of MaSuRCA over other assemblers is its ability to combine the benefits of deBruijn graph and Overlap-Layout-Consensus assembly approaches using short Illumina reads and long high-error prone data, such as MinION and PacBio, to create a consensus genome assembly (Zimin *et al.* 2013). After assembly, a set of summary statistics was computed using CEGMA (Parra *et al.*, 2007) to analyse the quality of the assembly. To improve the draft assemblies, the genomes for all biotypes were polished using the relevant Illumina HiSeq RNA-seq libraries for each biotype with Pilon v.1.23 (Walker *et al.*, 2014). GenomeScope v2 was used to create k-mer spectra plots of both the resistant and susceptible *N. ribisnigri* genomes (<http://qb.cshl.edu/genomescope/genomescope2.0/>) (Appendix 2).

4.2.1.5 Contamination removal

To identify and remove contamination from non-target organisms, Rothamsted Research's dedicated server for DNA and protein similarity searches, DeCypher, was used. Of the contaminants in the DeCypher results, the majority of them were identified to be the host plant lettuce (*Lactuca sativa*) and the symbiotic bacteria found in aphids, *Buchnera aphidicola*. Since all *N. ribisnigri* used for genome assembly were collected from long term laboratory cultures, the risk of contamination from parasitoid wasps was incredibly low. By running the newly assembled aphid genome against the NCBI database of all known nucleotide sequences, DeCypher generates a text file containing all positive nucleotide matches to sequences in the database (Wheeler *et al.*, 2007). The taxonomic analysis program, MEGAN (12.0.1) (Huson *et al.*, 2007), was incorporated, along with the Decypher, to identify any sequences which matched other non-target organisms against the NCBI reference database. MEGAN conveniently organises the DeCypher text file containing the NCBI positive blast hits of the newly assembled genomes into categories (cellular organisms: Bacteria, Archaea, Eukaryota (animal, plant);

and other sequences: viruses, unclassified sequences, No hits and Not assigned). As a result, MEGAN enables the identification of sequences in the newly assembled genome belonging to other organisms. All identified contaminant sequences in the genome in MEGAN (e.g. bacterial and plant) were submitted to the NCBI Web BLAST (nucleotide) database to clarify sequence identity. Any confirmed contaminants were manually removed from the genome using the bioinformatics software Geneious (10.1). The aphid symbiotic bacteria sequences of *B. aphidicola* were removed but retained for future analysis.

4.2.1.6 Assessment of *N. ribisnigri* completeness using Benchmarking Universal Single-Copy Orthologs (BUSCO)

Once all contaminants had been removed, Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão *et al.*, 2015) was used to assess the completeness of the *N. ribisnigri* genomes using an Arthropoda gene set (insecta_odb9) (n=1562) with default settings (Waterhouse *et al.*, 2018).

4.2.1.7 Genome annotation

Prior to annotation, RepeatMasker (v4.1.0) (Smith *et al.*, 2013) was conducted to identify any low complexity DNA sequences within the genomes and soft-masked with an 'N'. RepeatMasker removes low complexity DNA sequences and interspersed repeats and masks them within a given sequence. RepeatMasker was run in conjunction with NCBI BLAST and Crossmatched with the 'Sternorrhyncha', a suborder of the Hemiptera, which includes all aphids and other similar taxonomic groups on the NCBI sequence database. Genome annotation files were uploaded onto Galaxy and annotation was performed using a configured MAKER annotation pipeline (Campbell *et al.*, 2014). Initial gene prediction was run using Augustus using the pea aphid (*Acyrtosiphon pisum*) and the peach-potato aphid (*M. persicae*) as a training set. *Macrosiphum persicae* protein sequences were initially used to indirectly infer gene predictions but not directly from all protein alignments. After the initial round of training, the outputs from the preliminary runs were used to automatically retrain the gene prediction

algorithm to produce higher quality gene models. Genome annotations were considered complete when no further improvements could be made to the gene models evaluated by BUSCO (insecta_odb9). The genome annotations were visually explored using the Integrative Genomics Viewer (IGV) against the newly assembled *N. ribisnigri* genomes (resistant biotype, Kent_CL and wild type, Nr8) (Robinson *et al.*, 2011). Functional annotations of the predicted gene models were generated in OmicsBox using InterProScan (Blum *et al.*, 2020).

4.2.2 RNA-seq methods and differential gene expression analysis

4.2.2.1 RNA extraction and RNA-seq library preparation

RNA-seq data for *N. ribisnigri* was provided from Warwick Crop Centre (Wellesbourne Campus, Warwick). Aphids had been reared on either the susceptible lettuce cultivar Pinokkio or the resistant (Nr-gene) cultivar, Eluarde (Table 4.2). Approximately 50 aphids (mixed apterous and alates) were pooled for RNA extraction from each culture (detailed in Table 4.2). Since the aphids were pooled prior to RNA extraction and downstream analysis, these 50 aphids acted as pseudo replicates for each culture. Total RNA was isolated using a trizol-based method followed by purification on RNeasy clean up columns (Qiagen) and removal of any DNA contamination by treatment with DNase I. RNA quality was checked on a Bioanalyzer (Agilent). 100bp paired-end sequencing was performed on an Illumina HiSeq platform by Welcome Trust Centre for Human Genetics, Oxford, with the resulting read counts ranging from 12634191 to 37227165 per sample.

Quality control of the RNA-seq reads was conducted on all datasets with FastQC prior to further downstream analysis. (Wingett and Andrews, 2018). The newly assembled reference genome of *Nasonovia ribisnigri* susceptible strain (Nr8_123) was used for mapping the reads using HISAT2 software with default options.

Table 4.2: *Nasonovia ribisnigri* cultures that were used for the RNA-seq data extraction and sequencing. These data were used for the transcriptome assembly and RNA-seq analysis for differentially expressed (DE) gene analysis. Nr: 0 = *Nr-gene* susceptible; Nr: 1 = *Nr-gene* susceptible; Pyrethroid R = pyrethroid resistant; Insecticide R = insecticide resistant; (S) = susceptible (does not contain *Nr-gene*); (R) = resistant (contains *Nr-gene*). For the differentially expressed (DE) gene analysis using DeSeq2, RNA-seq data was grouped into three groups: blue (susceptible), orange (resistant_Kent) and green (resistant_UK).

<i>Nasonovia ribisnigri</i> culture	Biotype	Host plant variety	Culture origin
Nr4	Nr:0, Pyrethroid R	Pinokkio (S)	Chichester, 1999
Nr8	Nr:0, Insecticide R	Pinokkio (S)	York, 1999
Nr29	Nr:0, Insecticide R	Pinokkio (S)	Suffolk, 1999
4850a	Nr: 0	Pinokkio (S)	Lincolnshire, 2003 (Hough, 2013)
WT Kent	Nr: 0	Pinokkio (S)	Kent, 2010 (Hough, 2013)
Kent CL	Nr: 1	Eluarde (R)	Kent, 2009 (Hough, 2013)
Kent CL	Nr: 1	Pinokkio (S)	Kent, 2009 (Hough, 2013)
UK631	Nr: 1	Eluarde (R)	UK - Unknown, 2009 (Hough, 2013)
Uk631	Nr: 1	Pinokkio (S)	UK - Unknown, 2009 (Hough, 2013)

4.2.2.2 Transcriptome assembly

After all reads were quality checked using FastQC (Wingett and Andrews, 2018), reads were aligned to the draft *N. ribisnigri* reference genome (Nr_123) using Hisat2 to create short read alignments. Bowtie2 was used to reconstruct the transcripts from short read alignments into gene structures. The final transcriptome was a guided assembly using the draft *N. ribisnigri* reference genome and assembled using Trinity v2.5.1, using default settings (Appendix 1).

4.2.2.3 Differential gene expression analysis

The same RNA-seq data was used for the differential gene expression analysis (section 5.2.2.1). RNA-seq data was aligned to the newly annotated *N. ribisnigri* genome annotation (Nr8_123) using *featureCounts* (Liao *et al.*, 2014; Yang *et al.*, 2014). *featureCounts* aligns all RNA-seq data from all conditions (Table 4.2) to each gene in the *N. ribisnigri* genome annotation to create a gene count table which can be used in further downstream analysis. The DESeq2 package (Anders and Huber, 2010) was initially used to test for differential gene expression between the susceptible group consisting of

all RNA-seq count data *N. ribisnigri* cultures (five) unable to feed on lettuce containing the NR-gene and the resistant group which can break host plant resistance (four) and feed on lettuce containing the NR-gene. Since this study was focused on elucidating the potential mechanisms of resistance to the Nr-gene in lettuce only, *N. ribisnigri* cultures Nr4, Nr8 and Nr29 (which have a level of resistance to either pyrethroids or insecticides; pirimicarb) were grouped together with 4850a and WT Kent as the 'susceptible' group, as they are all unable to feed on lettuce which contains the Nr-gene. To identify whether any differential gene (DE) expression between the two resistant cultures (Kent CL and UK631), a second DESeq2 analysis was conducted between the susceptible group, Kent CL and UK631 (three groups).

For the DESeq2 analyses, data rows with no or single counts were removed, and count data was log transformed (\log_2) prior to analysis. Gene expression log fold changes were considered significant if $P < 0.05$ (alpha). The apegglm package was utilised to generate more accurate \log_2 fold change (LFC) estimates (Zhu *et al.*, 2018). The more recent versions of the DESeq2 package does not perform shrinkage of the LFC estimates by default. The apegglm package was incorporated into the analysis as, when RNA-seq count data of a gene is low, apegglm allows for the shrinkage of the LFC estimates towards zero (i.e. low counts or high dispersion) (Zhu *et al.*, 2018). For the RNA-seq count data for each gene, a negative binomial distribution was used to estimate the variability in the data. A Wald test was used post-hoc to test for differential expression between resistant-breaking (Nr:1) and susceptible (Nr:0) *N. ribisnigri* biotypes. The null hypothesis for each gene was that there is no differential expression between Nr:1 or Nr:0 (i.e. LFC = 0) (see Appendix 3, section 9.3 for further details).

All DeSeq2 analyses was conducted in RStudio (4.0.5) (R Core Team, 2020). The ggplot package was used to create a principal component analysis (PCA) plot of the experiment to visualise the overall similarities of the samples in both the susceptible and resistant groups (two) and susceptible and two resistant groups (three). To visualise the significant DE genes between the susceptible and resistant groups, a volcano plot package (EnhancedVolcano) was used. For the volcano plot only, both DE genes of the resistant (Nr:1) groups were pooled and plotted against the DE genes of the susceptible (Nr:0) for visualisation. A heatmap of the top DE genes was created using the pheatmap package. An initial

P-value of 0.05 (alpha) was used but a more stringent *P*-value of 0.00001 was used to reduce number of identified DE genes. Finally, to identify the number of DE genes between the two resistant cultures (Kent CL and UK631), data of all significant (alpha = 0.05) DE genes were extracted into separate files to create a Venn diagram (<https://bioinformatics.psb.ugent.be/webtools/Venn/>).

4.2.3 Validation of RNA-seq data by qRT-PCR

4.2.3.1 Aphid and plant cultures

To confirm the results of the RNA-seq data, A single apterous *N. ribisnigri* from four biotype cultures (Nr:0 and Nr:1) were placed onto either a susceptible (cv. Pinokkio) or resistant (cv. Eluarde) lettuce host plant (Table 4.3). Two susceptible cultures (Nr:0) and two resistant cultures (Nr:1) were used for the qRT-PCR gene expression validation experiment, which were also used in the initial RNA-seq analysis experiment. Once the founding mother had produced several nymphs, the founder was removed, and the nymphs left to start a new colony. In total there were four treatments, with each culture triplicated to create three biological replicates per treatment (12 conditions) (Table 4.3). Each culture was maintained in laboratory conditions described previously (Section 4.2.1.1) until ~50 individuals were produced. Each isogenic line was collected in 1.5ml eppendorfs containing RNAprotect (Qiagen) and frozen at -80°C prior to RNA extraction.

Table 4.3: Conditions used for qRT-PCR experiment for *N. ribisnigri*. Nr:0 = susceptible biotype (unable to feed on resistant lettuce containing the Nr-gene. Nr:1 = resistant biotype (able to feed and reproduce on resistant lettuce containing the Nr-gene). UK631 were cultured on both susceptible (Pinokkio) and resistant (Eluarde) lettuce lines. Susceptible *N. ribisnigri* cultures were only cultured on susceptible (Pinokkio) lettuce lines.

Treatment	<i>Nasonovia ribisnigri</i> culture	Biotype	Host plant variety	Biological replicate
1	WT Kent	Nr: 0	Pinokkio (S)	1
1	WT Kent	Nr: 0	Pinokkio (S)	2
1	WT Kent	Nr: 0	Pinokkio (S)	3
2	Kent CL	Nr: 1	Eluarde (R)	1
2	Kent CL	Nr: 1	Eluarde (R)	2
2	Kent CL	Nr: 1	Eluarde (R)	3
3	UK631	Nr: 1	Eluarde (R)	1
3	UK631	Nr: 1	Eluarde (R)	2
3	UK631	Nr: 1	Eluarde (R)	3
4	Uk631	Nr: 1	Pinokkio (S)	1
4	Uk631	Nr: 1	Pinokkio (S)	2
4	Uk631	Nr: 1	Pinokkio (S)	3

4.2.3.2 House-keeping gene discovery

The house-keeping genes (HKGs); ribosomal protein S9 (*RPS9*), ribosomal protein S18 (*RPS18*) and ribosomal protein L13a (*RPL13*), were used as internal controls. The *RPS9* HKG was chosen based on previous work which demonstrated that it had a stable expression among soybean aphids (*Aphis glycines* (Matsumura) on susceptible and resistant host plants (Bansal *et al.*, 2012). The ribosomal proteins *RPS18* and *RPL13* have been proven to be good HKGs in gene expression analyses in the mustard aphid (*Lipaphis erysimi* (Kaltenbach) (Koramutla *et al.*, 2016). Protein homology searches for these genes' transcript sequences were performed on the newly assembled *N. ribisnigri* transcriptome (76,782 transcripts) which was uploaded onto an in-house dedicated server (DeCypher v9.1) (TimeLogic Division, 2013). This server uses an onsite high-performance computer with parallel processing chips and enables DNA/protein similarity searches to be performed, like the NCBI BLAST function with a computer hardware acceleration (Luethy *et al.*, 2004). The identity of putative cDNAs was validated using the Blastx search function in NCBI-GenBank Once cDNA identities had been confirmed, specific PCR and qRT-PCR primers for each gene were designed using PrimerQuest (<https://eu.idtdna.com/pages/tools/primerquest>) (Appendix 3, section 9.4).

4.2.3.3 Identification and validation of the differentially expressed gene in *Nasonovia ribisnigri*

One significantly differentially expressed (DE) gene of interest was selected from the RNA-seq experiment for further analysis using qPCR in conjunction with the HKGs. The identified gene was located and visualised in the *N. ribisnigri* transcriptome using OmicsBox v.1.4.11 (OmicsBox, 2019). This transcript was subsequently used in a Blastx homology search in NCBI-GenBank. A protein homology search was conducted for the DE gene of interest, following the method described above for the HKGs using DeCypher. Once cDNA identities had been confirmed, specific PCR and qRT-PCR primers for the potential resistance gene were designed using PrimerQuest available online: (<https://eu.idtdna.com/pages/tools/primerquest>).

4.2.3.4 qRT-PCR analysis on *N. ribisnigri* to validate RNA-seq data

For RNA-seq data validation, aphid samples were collected as described previously. These samples were processed for total RNA extraction using a RNeasy Micro Kit (50) (Qiagen), following the manufacturer's guidelines. In brief, 5-10 individual aphids were used from each condition (Table 4.3) and homogenised using liquid nitrogen in 1.5 ml Eppendorfs. Once homogenised, 350µl of RW1 buffer was added and the lysate was transferred to a RNeasy spin column and centrifuged at 10 000 rpm (9408 g) for 15 s and the flow-through was discarded. A stock solution of 10µl DNase I was added to 70µl of RDD buffer and added to the RNase spin column and mixed by inverting. The DNase I solution was incubated on the benchtop for 15 min at 20°C. The spin column was then centrifuged at 10 000 rpm for 15 s. The flow-through was discarded and 500µl of RPE buffer was added to the spin column and centrifuged at 10 000 rpm for 15 s. This was repeated again but the centrifugation was at 10 000 rpm for 2 min. To elute the RNA, the RNeasy spin column was placed in a new 1.5 ml collection tube and 30µl of RNase-free water were centrifuged at 10 000 rpm for 1 min. All RNA samples were checked for quality and quantity using a nanodrop. First-strand cDNA was prepared with 200 ng RNA (DNA free) for all samples.

The qRT-PCR was conducted using PowerUp SYBR Green Master Mix (ThermoFisher Scientific) on a thermocycler (GeneAmp PCR System 9700). Each reaction was performed in triplicate (three technical replicates for each biological replicate) in 96-well optical-grade PCR plates (ThermoFisher Scientific), which were sealed with optical sealing tape. qRT-PCR primers for the resistance gene and HKGs were designed using PrimerQuest (Table 4.4). Each reaction was carried out with 2 µl of cDNA, 0.4 µl of each primer pair (100ng/µl), 2.2µl RNase-free water and 5 µl of PowerUp SYBR Green Master Mix in a total volume of 10 µl. qRT-PCR reactions were conducted with the following conditions: Uracil-DNA glycosylases (UDG) activation at 50°C for 2 min, followed by Taq DNA polymerase step of 95°C for 2 min, 40 cycles of denature at 95°C for 15 s, annealing of 55°C for 15 s and extension at 72°C for 1 min. Melt curve analysis was performed by heating the PCR samples from 55 to 95°C (1°C per cycle of 10 secs) with SYBR Green I signal intensity measurements. There were three biological replicates (Table 4.3) and three technical replicates used to validate the gene/s. Statistical analysis was performed using t-test through the MeV package, version 4.9 available at www.tm4.org.

Table 4.4: Primer sequences and amplicon characteristics of resistance genes and house-keeping genes for qRT-PCR validation experiment for *N. ribisnigri*.

Gene	Primer sequence	Amplicon length (bp)	Product temp (°C)
2698_1	F: GTTAGTCGGATCGAGTCCAATG R: CGTGAGCGATACTTTGACATTTAC	115	62
2698_2	F: ACGTTGAGTGGCAGTATAGTTATC R: AAGATATGAGAACCCACGTGTTAG	89	62
RPS9	F: CTGTTGACCCTCGAAGAGAAG R: GCCCTCGTCTAATACTCCAATAC	93	62
RPS18	F: CGTATCCTCAGCACCAACAT R: CGGTACATTCTCCAGCTCTTT	142	62
RPL13	F: TCAAATACGATGCAAACCTTCAC R: ACGCACTCCTCATTCCTTAAC	102	62

4.2.3.5 House-keeping gene stability analysis

The software algorithm GeNorm was used to determine the stability and efficiency of the three primer pairs of the house-keeping genes (*RPS9*, *RPS18* and *RPL13a*) (Vandesompele *et al.*, 2002; Andersen *et al.*, 2004). The inputs for GeNorm were the raw expression values for each gene (using the equation, $2^{-\Delta Ct}$). An M-score is calculated by GeNorm, with a lower M value of < 1.5 suggests a more stable gene expression or low variation, and an M-score > 1.5 indicating high variation and therefore not a suitable HKG. Fold change and relative expression level were determined using the comparative Ct method ($2^{-\Delta\Delta Ct}$) (Schmittgen and Livak, 2008). Statistical analysis was performed using a t-test (MeV package).

4.3 Results

4.3.1 Genome *de novo* assembly and annotation

Altogether, 45.2Gb short read Illumina data was generated for the Nr:0 biotype *N. ribisnigri* genome and 20.6Gb produced for the Nr:1 biotype *N. ribisnigri* genome with. From the two SpotON flow cells, a total of 2,966,975 (average sequence length; 2,155) and 5,167,875 (average sequence length; 2,401) long read nanopore sequences were generated for the Kent CL (resistant) and Nr8 (susceptible) genomes, respectively. This produced an additional 30.3Gb of long-read Nanopore data for the hybrid genome assembly. The hybrid *de novo* genome assembly of the susceptible Nr:0 *N. ribisnigri* genome

was assembled onto 4,778 scaffolds with an N50 of 295Kb (Table 4.5). The GC contents of the *N. ribisnigri* genomes were between 29.75-30.1% and suggests that all contaminants were removed.

Table 4.5: Summary statistics of the currant-lettuce aphid, *N. ribisnigri* genomes. Nr8 and WT_Kent are susceptible (Nr:0) biotypes and UK631, Kent CL and Ely are resistant (Nr:1) biotypes. Only the best genome assemblies for the susceptible (Nr:0) and resistant (Nr:1) were selected and used for downstream annotation and analysis, highlighted in red.

	Genome	Genome size (Mb)	N50	Scaffolds	Max scaffold size	GC Content (combined %)	BUSCO (%)
Susceptible (WT, Nr:0)	Nr8_(123)	423	295,999	4,778	3,391,359	29.81	94.2
	WT_Kent (123)	423.2	259,518	5,225	2,467,012	29.8	92.5
	UK631 (1)	367.9	105,098	8,130	967,723	29.79	91.6
	UK631 (2)	359.5	71,978	10,488	721,085	29.87	82.9
	UK631 (12)	367.9	107,002	8,047	1,394,296	29.8	91.3
Resistant (R, Nr:1)	Kent CL (1)	384.2	163,106	6,264	1,500,737	29.76	92.9
	Kent CL (2)	384	162,012	6,305	1,147,043	29.74	92.4
	Kent CL (12)	382.3	99,972	8,234	1,161,853	29.8	91.4
	Ely (1)	366.9	107,464	8,035	956,705	29.75	91.1
	Ely (2)	365.2	101,694	8,289	903,197	29.77	91.5
	Ely (12)	366.9	109,458	7,890	1,394,296	29.7	91.5

The genome size of *N. ribisnigri* was estimated to be ~360Mbp which is comparable to previously sequenced aphid genomes. The number of scaffolds in genome assemblies ranged between 4,778 – 12,004 and the maximum scaffold size between 721Kbp – 3,391Kbp (Table 4.5).

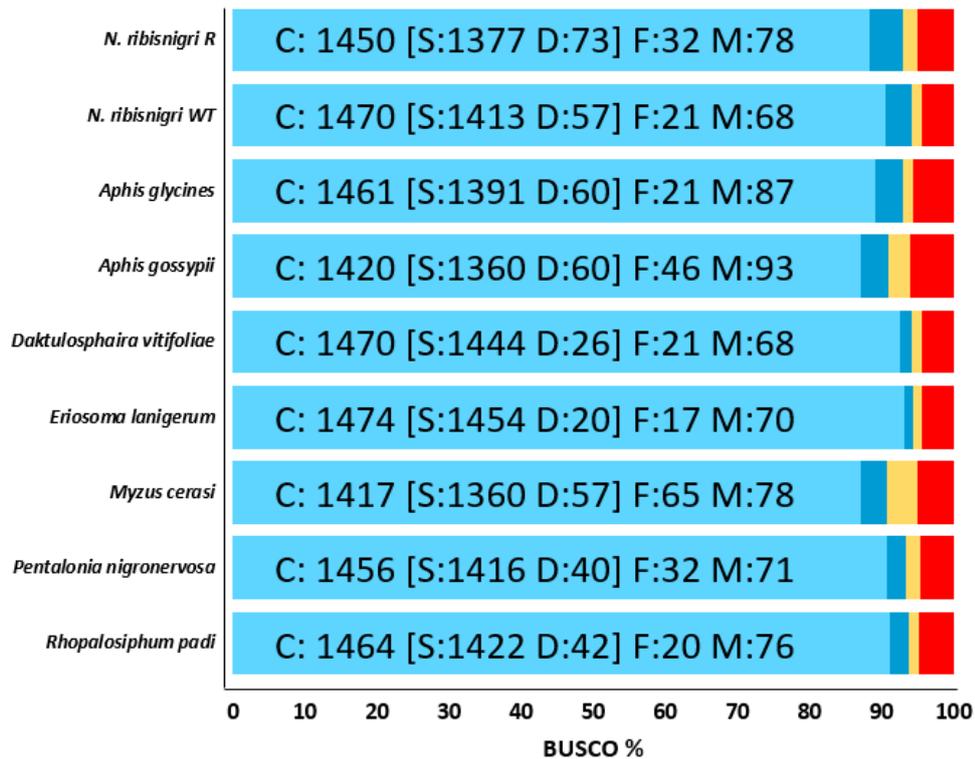


Figure 4.1: BUSCO analysis score of both *N. ribisnigri* genomes (WT = Nr:0, susceptible; R = Nr:1, Resistant) and previously published aphid genomes (obtained from BIPAA genomic resources) using an Arthropod gene set (Insecta_odb9) (n= 1562). C: Complete, S: Single-copy, D: Duplicate, F: Fragmented, M: Missing.

The completeness of the genome assemblies as determined using BUSCO (Insecta_odb9) indicated that the *N. ribisnigri* genomes of the susceptible (WT) and resistance-breaking (R) aphids were 94.2% and 91.4% complete, respectively. A *de novo* hybrid PAN genome assembly was attempted using the combination of reads from all biotypes to determine whether this would improve overall assembly statistics but failed to do so. After removal of contaminants and further quality control, completeness of the PAN genome was 93% (BUSCO, Insecta odb9), Augustus predicted a total of 38,389 gene models, including partial gene models, genes incorrectly shown or duplicated and unsupported *ab initio* models. After running an Interpro Scan (IPS) search of all 38,389 predicted genes, ~8,000 genes had no IPS matches, ~23,000 genes had no GO terms but had IPS matches and ~7,000 had both GO terms and IPS matches. Out of these, a further 3,044 genes had either a start or stop codon missing and were subsequently removed.

4.3.1.1 Transcriptome *de novo* assembly and annotation

A total of 90Gb of strand-specific paired-end RNA-seq data from *N. ribisnigri* isogenic lines was generated and provided by Warwick Crop Centre which was used in the *de novo* assembly of the *N. ribisnigri* transcriptome. This generated a near complete transcriptome (94.8%, Arthropoda gene set (n=1562)) with 76,782 transcripts totalling 83,926,066 bases. The length of the transcripts varied from 201 to 18,524 nucleotides with an average of 1,093 (Fig. 4.2 A). The transcriptome assembly N50 was 2,341, in which 50% of all bases in the assembly are covered by sequences equal or larger and is comparatively high for a non-model organism (Francis *et al.*, 2013). The transcriptome completeness (BUSCO) for the susceptible *N. ribisnigri* biotype was 94.8% and is similar to other available aphid transcriptomes, such as the pea aphid (*Acyrtosiphon pisum*) which was 96.2% (Fig. 4.2 B).

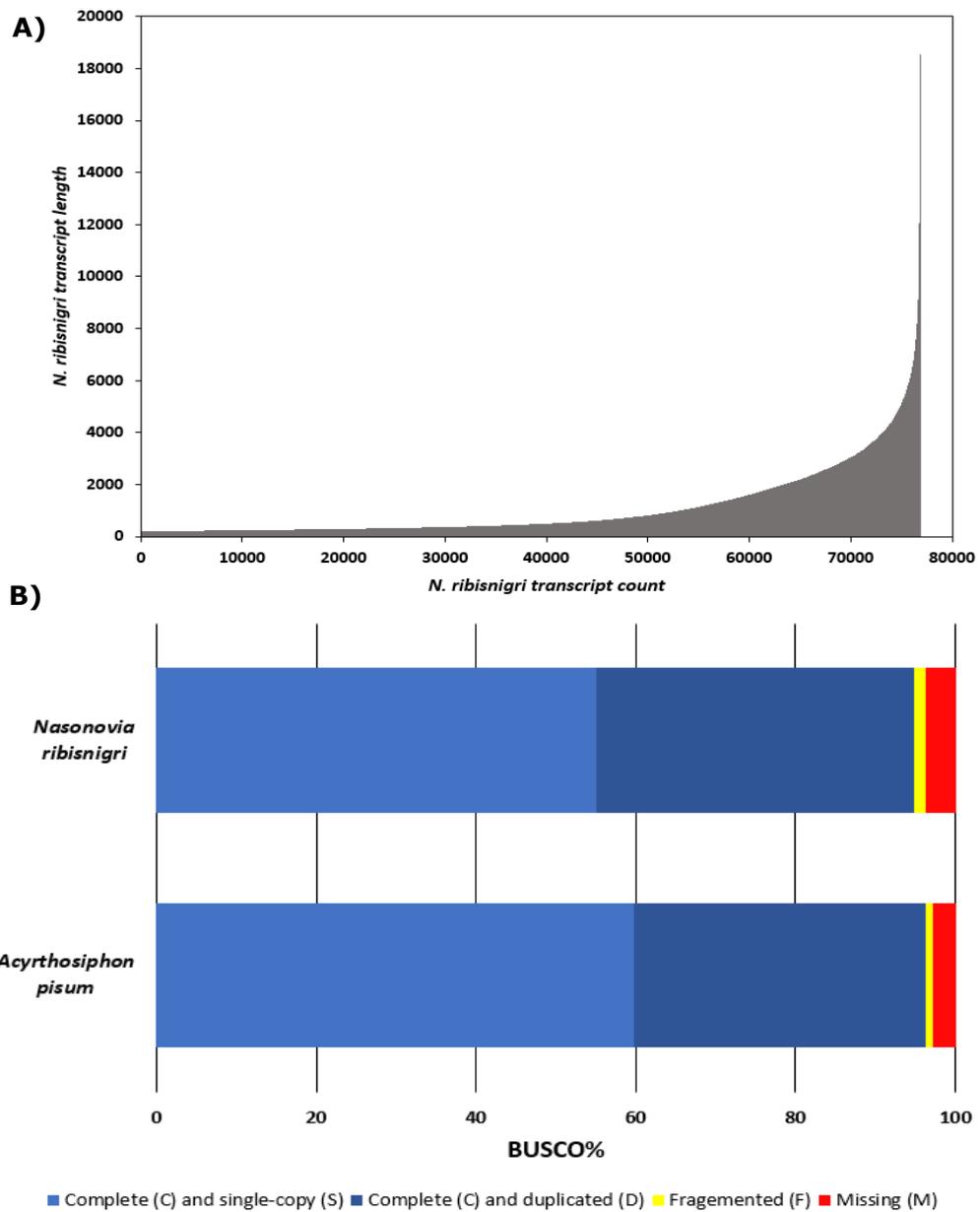


Figure 4.2: A) *Nasonovia ribisnigri* transcript length distribution of 76,782 contigs of de novo assembled transcriptome. Individual contigs are ordered on X-axis based on increasing size. B) BUSCO analysis score of the *N. ribisnigri* transcriptome and the pea aphid (*Acyrthosiphon pisum*) transcriptome (obtained from BIPAA genomic resources) using an Arthropod gene set (n= 1562). C: Complete, S: Single-copy, D: Duplicate, F: Fragmented, M: Missing.

4.3.1.2 RNA-seq analysis

In total, there were 183 million RNA-seq reads for *N. ribisnigri* for all conditions (Table 4.6). Over 115 million RNA-seq reads were obtained for the *N. ribisnigri* which were susceptible to the *Nr*-gene in the lettuce host plant, with the remaining 68 million RNA-seq reads for *N. ribisnigri* which were resistant to the *Nr*-gene in the lettuce host plant.

Table 4.6: HISAT2 results of number of reads which aligned to the *Nr_8* WT genome annotation using feature counts and mapping statistics. Nr:0 = *N. ribisnigri* unable to feed to lettuce containing the *Nr*-gene (susceptible); Nr:1 *N. ribisnigri* able to break the host plant resistance and feed on lettuce containing the *Nr*-gene (resistant).

<i>Nasonovia ribisnigri</i> culture	Biotype	Host plant	No. of HiSeq reads	Assigned	Unmapped	Mapping quality	Chimera
Nr4	Nr:0, Pyrethroid R	Pinokkio (S)	37,726,652	30225819	788040	33682	62599
Nr8	Nr:0, Insecticide R	Pinokkio (S)	11,418,030	9119987	257288	9339	21991
Nr29	Nr:0, Insecticide R	Pinokkio (S)	11,779,986	8174203	1231914	50176	68931
4850a	Nr: 0	Pinokkio (S)	17,018,669	1253797	317543	21170	36128
WT Kent	Nr: 0	Pinokkio (S)	37,227,165	28894399	2041638	49777	95444
Kent CL	Nr: 1	Eluarde (R)	18,270,893	13918495	538177	28144	48192
Kent CL	Nr: 1	Pinokkio (S)	19,054,666	14008966	516558	24536	45233
UK631	Nr: 1	Eluarde (R)	18,356,475	9618800	219916	13510	27106
Uk631	Nr: 1	Pinokkio (S)	12,634,191	14452372	479047	22803	40625

Between the susceptible and resistant group, DeSeq2 identified 18 872 genes, with 689 DE genes (alpha of $P < 0.05$) (Fig. 4.3). Of these 689 DE genes, 351 were up- and 338 down-regulated genes were identified between the susceptible and resistant *N. ribisnigri* groups (Fig. 4.4). Further stringent refinement (alpha = $P < 0.00001$), identified 32 DE genes out of the 18 872 genes.

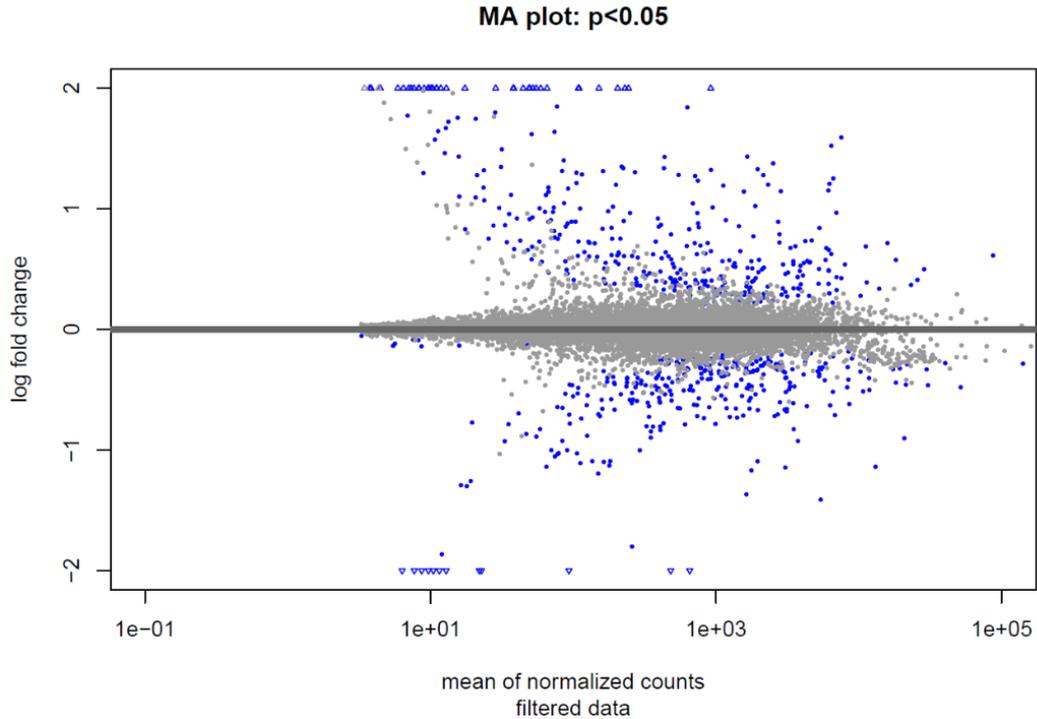


Figure 4.3: Gene expression change in *N. ribisnigri* biotypes feeding on resistant (Nr-gene) and susceptible (Nr-gene absent) lettuce host plants. The shrunken log₂ fold change (removing the noise associated with log₂ fold changes from low count genes) of each gene between *N. ribisnigri* able to feed on resistant lettuce (Nr:1) and *N. ribisnigri* unable to feed on resistant lettuce (Nr:0). Fold changes were considered significant if P values were <0.05. Differentially expressed (DE) genes are coloured as blue circles and DE genes which fall out of the window are indicated by open triangles. The remaining grey circles are genes which are not significantly expressed between the resistant and susceptible biotype.

A principal component analysis (PCA) highlighted some differences between the *N. ribisnigri* biotypes (Fig. 4.4). With the DeSeq2 analysis separated into three groups, the PCA suggested that there was little intra-variation between the resistance-breaking biotypes feeding on either lettuce containing the Nr-gene (Eluarde) or without the gene (Pinokkio) but with a notable inter-variation between different resistance-breaking biotypes. For the susceptible group, a close grouping of three susceptible biotypes (Nr4, Nr8 and Nr29) is apparent, which are *N. ribisnigri* cultures shown to be unable to feed on lettuce containing the Nr-gene but have some resistance to pyrethroids or insecticides. The remaining two susceptible biotypes are loosely clustered together (Fig. 4.4) and are unable to feed on lettuce containing the Nr-gene nor do they have any insecticide resistance.

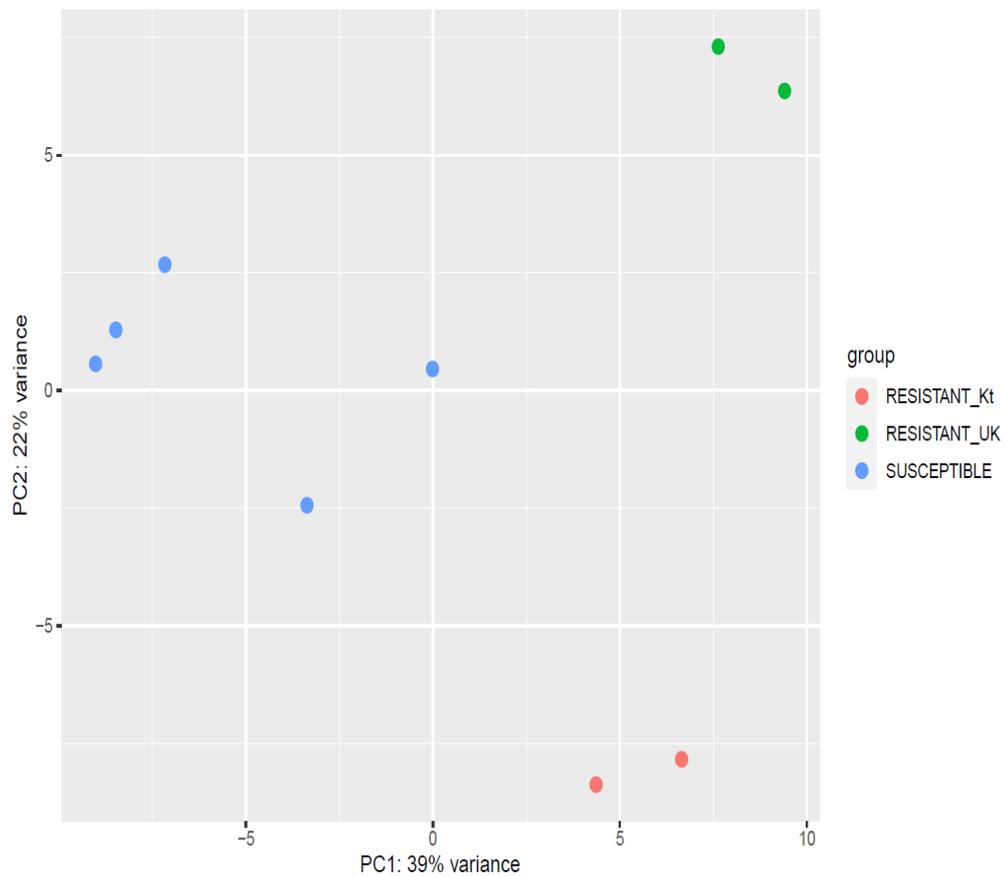


Figure 4.4: Principal Component Analysis of the RNA-read count data (*featurecounts*) for each *N. ribisnigri* biotype grouped into three groups (Susceptible, Resistant_Kt (Kent_CL) and Resistant_UK (UK631)). Susceptible denotes all five *N. ribisnigri* cultures that are unable to feed on lettuce containing the *Nr-gene* (Nr:0).

Between the two groups of susceptible and resistant cultures, the volcano plot (Fig 4.5) highlights all 689 DE genes ($\alpha = P < 0.05$) out of the 18,872 genes identified by the DeSeq2 analysis. Out of the DE genes, maker-Scaffold_2698-augustus-gene-0.8 (2698) was the most significantly up-regulated in the resistant biotype.

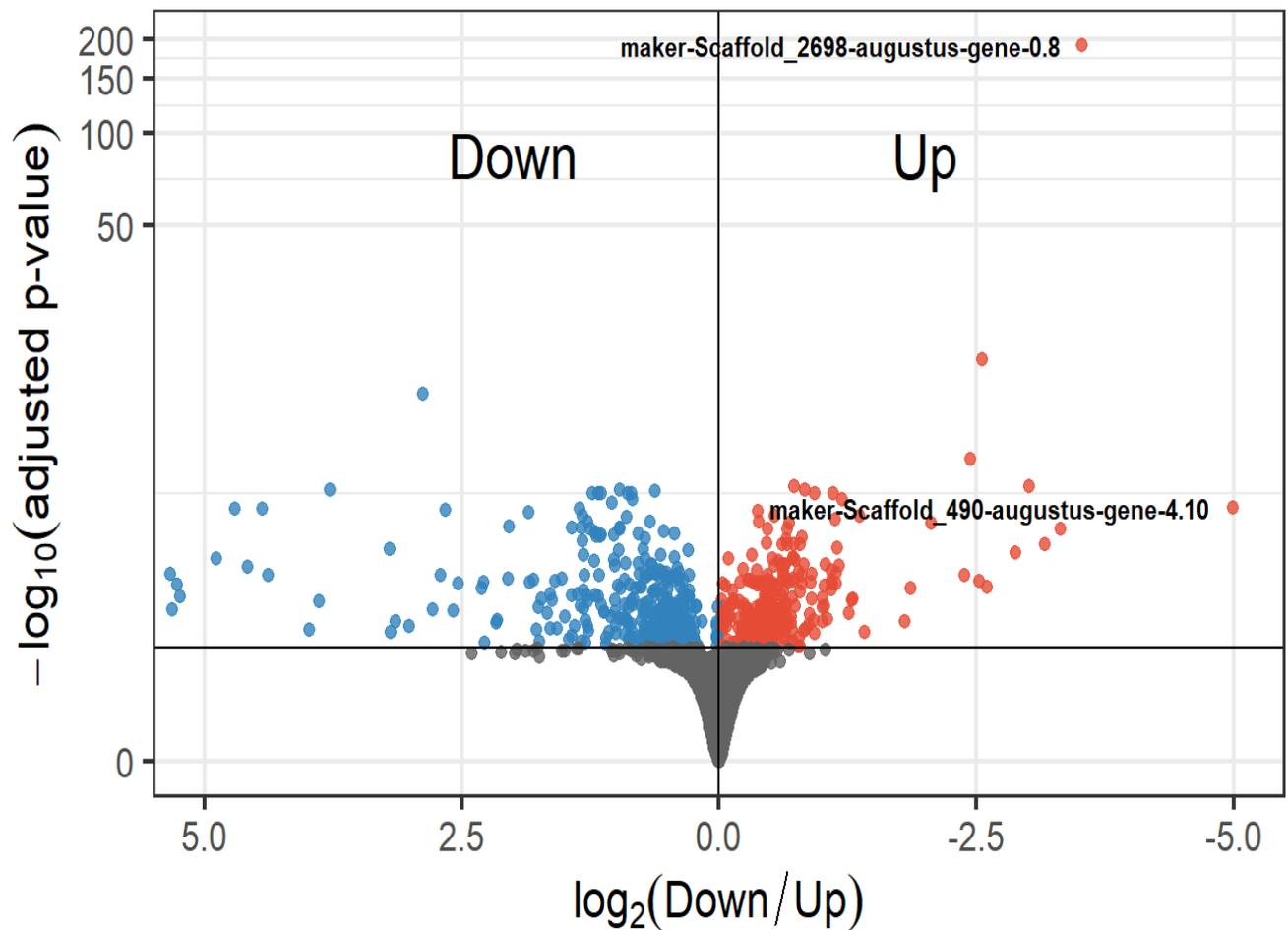


Figure 4.5: Volcano plot of the 18,872 genes in both resistance-breaking (Nr:1) and susceptible (Nr:0) *N. ribisnigri* biotypes. 689 differentially expressed (DE) genes were identified with a P value of <0.05 , highlighted in either blue or red. The volcano plot used the susceptible (Nr:0) *N. ribisnigri* biotype as a foundation for the plot and therefore DE genes are shown to be down regulated in the Nr:0 biotype and up regulated in the resistance-breaking (Nr:1) biotype. All up-regulated DE genes are highlighted in red and all down-regulated DE genes are highlighted in blue.

4.3.1.3 Differential gene expression of susceptible and resistant *N. ribisnigri* biotypes

After further refinement of the initial 689 DE genes ($P < 0.05$), a P value cut-off of $P < 0.00001$ was used to reveal the top 32 DE genes between the resistant-breaking and susceptible *N. ribisnigri* biotypes. The RNA-seq analysis identified clusters of up- and down-regulated genes between the two conditions, highlighted using a heatmap (Fig. 4.6).

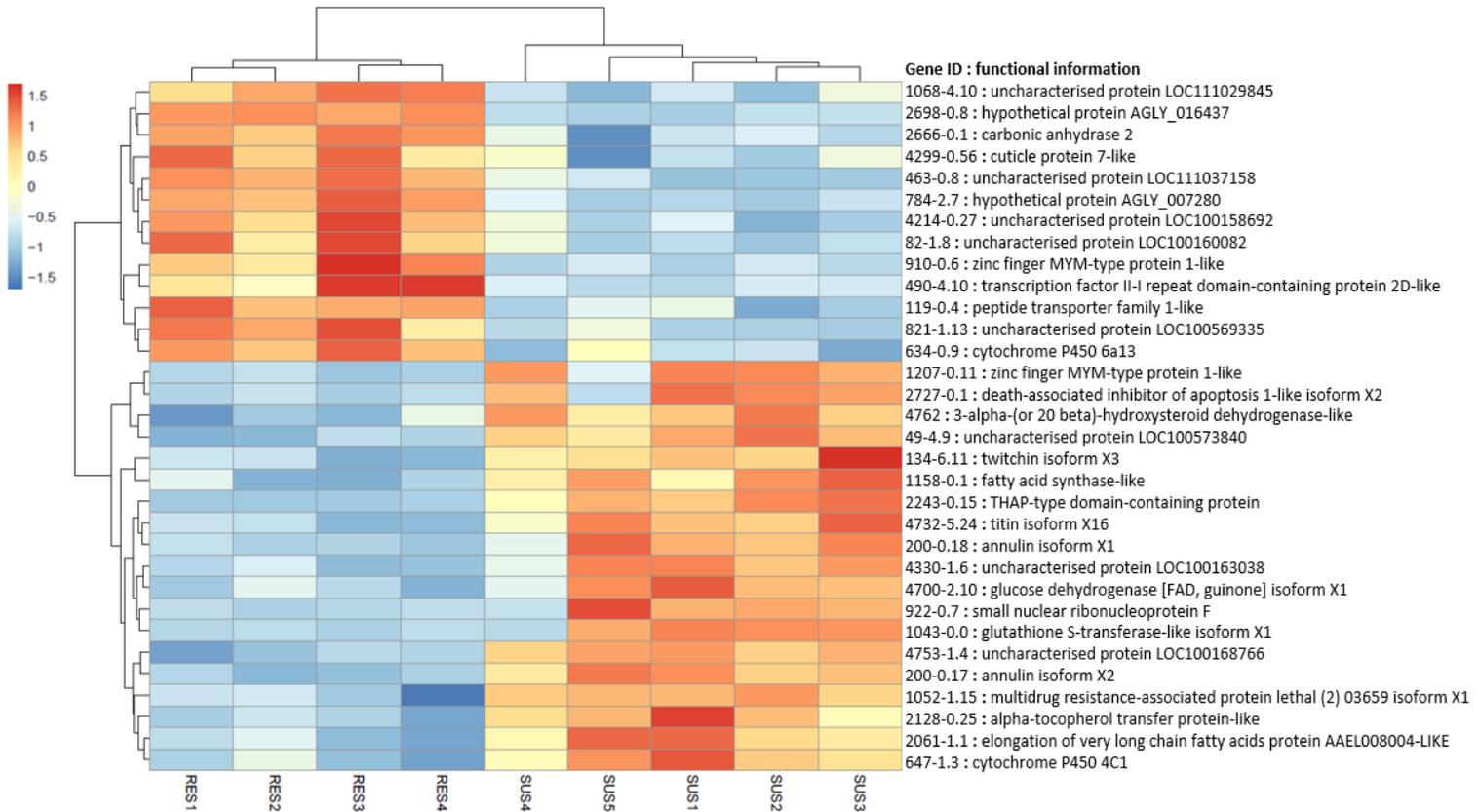


Figure 4.6: Heatmap of the top 32 DE genes ($P < 0.00001$) between *N. ribisnigri* resistant-breaking (Nr:1) and susceptible (Nr:0) biotypes, showing clear division. RES = resistant-breaking (Nr:1); SUS = susceptible (Nr:0). SUS1 = Nr4, SUS2 = Nr8, SUS3=Nr29, SUS4=4850a, SUS5=WT Kent, RES1=Kent_CL (Eluarde), RES2=Kent_CL (Pinokkio), RES3=UK631 (Eluarde), RES4=UK631 (Pinokkio). Gene identification (ID) is indicated on the right-hand side, along with functional information.

Both resistance-breaking biotypes were grouped together according to their origin (RES1 and RES2 – Kent_CL; RES3 and RES4 – UK631) and had similar gene expression levels despite feeding on lettuce with and without the Nr-gene. Out of the top 32 DE genes, one of the susceptible biotypes (SUS4 – 4850a) was down-regulating two of the DE gene clusters that were up-regulated in the other susceptible biotypes. Additionally, there was little differentiation in some of the DE genes down-regulated by the other susceptible biotypes. Out of these 32 DE genes, 13 were up-regulated in the resistance-breaking biotypes, with the remaining 19 being down-regulated. The 13 DE genes in the resistance-breaking biotypes clustered into four main groups. The first cluster consists of uncharacterised proteins and a carbonic anhydrase 2, the second; uncharacterised proteins and a cuticle protein 7, third; transcription factor and a zinc finger MYM-type, and fourth; uncharacterised protein, peptide transporter and a cytochrome P450 (Table 4.7).

Table 4.7: Differentially expressed (DE) genes in resistance-breaking biotypes of *N. ribisnigri* (Nr:1) feeding on both resistant (Nr-gene present) and susceptible (nr-gene absent) lettuce host plants. Both up-regulated and down-regulated DE genes are shown for the resistance-breaking (Nr:1) biotype. Gene names are derived from contig number (e.g. 2698) and transcript ID (0.8). Both up-regulated and down-regulated genes are ordered by *P*-value significance identified from DeSeq2 using LFC estimates. Fold change values were considered significant if *P* < 0.0001.

Gene	Log2 fold Change	<i>P</i> -value	Description
Up-regulated			
2698-0.8	3.52761	2.8E-195	hypothetical protein AGLY_016437
463-0.8	2.56059	1.15E-22	uncharacterised protein LOC111037158
784-2.7	2.4398	1.92E-12	hypothetical protein AGLY_007280
4214-0.27	0.73149	1.22E-10	uncharacterised protein LOC100158692
910-0.6	3.01237	1.37E-10	zinc finger MYM-type protein 1-like (A.pisum)
821-1.13	0.83535	2.85E-10	uncharacterised protein LOC100569335
634-0.9	0.92507	6.39E-10	cytochrome P450 6a13
2666-0.1	1.109	7.44E-10	carbonic anhydrase 2
82-1.8	1.19499	2.09E-09	uncharacterised protein LOC100160082
490-4.10	4.99791	6.75E-09	transcription factor II-I repeat domain-containing protein 2D-like
1068-4.10	0.37322	1.17E-08	uncharacterised protein LOC111029845
119-0.4	0.53792	2.19E-08	peptide transporter family 1-like
4299-0.56	1.36789	2.19E-08	cuticle protein 7-like
Down-regulated			
4753-1.4	-2.876173	3.7E-18	uncharacterised protein LOC100168766
1207-0.11	-3.778138	2.45E-10	zinc finger MYM-type protein 1-like
4762-30.11	-0.965942	3.52E-10	3-alpha-(or 20-beta)-hydroxysteroid dehydrogenase-like
1052-1.15	-0.622874	4.44E-10	multidrug resistance-associated protein lethal (2)03659 isoform X1
4732-5.24	-0.892667	5.96E-10	titin isoform X16
200-0.17	-1.177182	8.78E-10	annulin isoform X2
2061-1.1	-1.234105	9.28E-10	elongation of very long chain fatty acids protein AAEL008004-like
2128-0.25	-0.847765	9.82E-10	alpha-tocopherol transfer protein-like
647-1.3	-1.146218	8.2E-10	cytochrome P450 4C1
134-6.11	-0.841906	2.18E-09	twitchin isoform X3
1158-0.1	-1.045716	3.69E-09	fatty acid synthase-like
2243-0.15	-7.413594	3.65E-09	THAP-type domain-containing protein
200-0.18	-1.350715	7.36E-09	annulin isoform X1
922-0.7	-4.709364	7.82E-09	small nuclear ribonucleoprotein F
1043-0.0	-4.436595	8.36E-09	glutathione S-transferase-like isoform X1
2727-0.1	-2.656668	1E-08	death-associated inhibitor of apoptosis 1-like isoform X2
4330-1.6	-1.848768	1.35E-08	uncharacterised protein LOC100163038
4700-2.10	-1.329245	2.18E-08	glucose dehydrogenase [FAD, quinone] isoform X1
49-4.9	-0.894127	2.51E-08	uncharacterised protein LOC100573840

4.3.1.4 Identification of single gene resistance to the Nr-gene in resistant lettuce

By creating three contrasts between both resistant biotypes (Kent_CL and UK631) and the susceptible biotypes highlighted that out of the 689 DE genes ($\alpha = P < 0.05$) identified, only 1 DE gene was shared between both resistance-breaking biotypes (Fig. 4.7 A). This gene was located on scaffold 2698 and identified as a hypothetical protein in *Acrythosiphon pisum* through a protein homology search using blastx (NCBI) (Table 4.7). Further blastx search results of this gene provided no further information. This protein was shown to be up-regulated in the resistance-breaking biotypes, with a log fold change of 11.47 and a highly significant P value, after LFC shrinkage estimates. The number of RNA-seq counts have a high level of associated transcripts for both resistance-breaking biotypes and a much lower number of transcripts for the susceptible biotypes, for all conditions and replicates (Fig. 4.7 B).

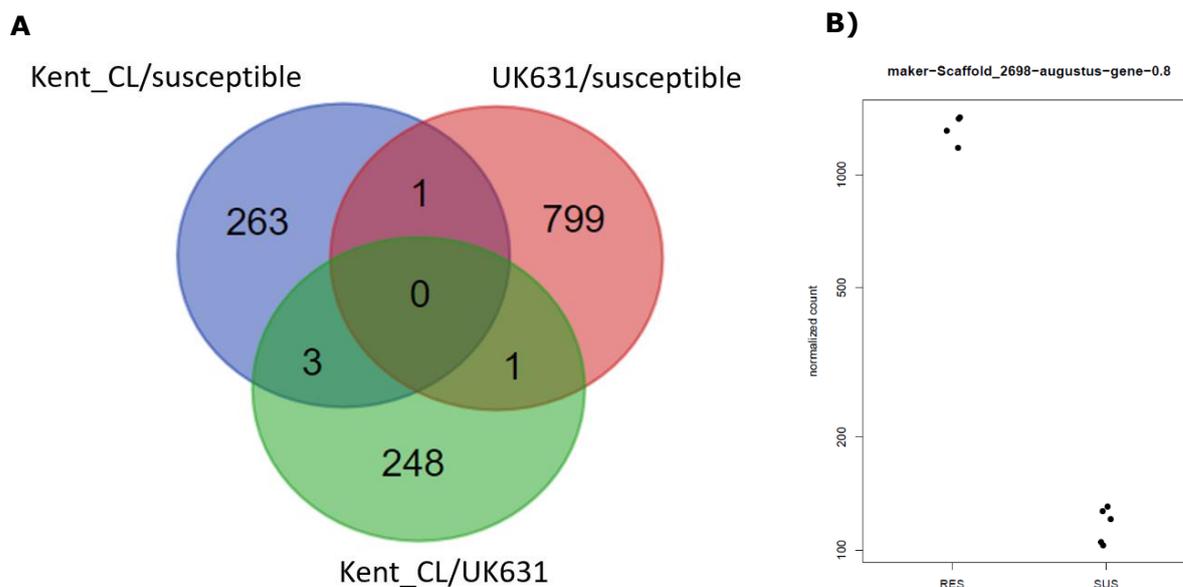


Figure 4.7: (A) 3-way Venn diagram of the differentially expressed genes in three contrasts (Kent_CL/susceptible; UK621/susceptible; Kent_CL/UK631) identified by DeSeq2. Differentially expressed genes were considered significant and used in the diagram if $P < 0.05$. (B) RNA-seq counts for gene 2698 between resistant-breaking (RES) and susceptible (SUS) biotypes.

In gene 2698 (459 amino acids), two SNPs were identified at nucleotide 9 and 50, relative to the first base of the initiating ATG codon. In the first SNP, WT_Kent, susceptible biotype the allele was homozygous for a T residue (97% of 34 sequence reads) in the first codon position of serine residue AGT, whereas the Kent_CL (fed on resistant lettuce) resistant biotype was homozygous for an C residue (100% of 248 sequence reads) which converted the serine to a threonine AGC. The second SNP, WT_Kent, susceptible biotype the allele was homozygous for a A residue (96% of 43 sequence reads) in the last codon position of methionine residue ATG, whereas Kent_CL (fed on resistant lettuce) resistant biotype was homozygous for a T (100% of 290 sequence reads) which converted the methionine to a leucine TTG. These had good RNA-seq coverage, were high quality with no strand bias suggesting that they are not an artifact.

4.3.1.5 qRT-PCR validation of *N. ribisnigri* resistant-breaking gene against the Nr-gene found in lettuce

Three previously used HKG's (*RPS9*, *RPS18*, *RPL13*) were profiled from the newly assembled *N. ribisnigri* transcriptome. The screening results of these HKGs presented two ranges of Ct values between the three HKGs (26.7-27.1 and 31.7-32.4) (Fig. 4.8). There was little variation between Ct values for both conditions (*N. ribisnigri* susceptible biotype feeding on susceptible host plants and *N. ribisnigri* resistant biotype feeding on resistant host plants) for all three HKGs. The qRT-PCR analysis was highly optimised and had amplification efficiencies (E) for primers from 97.28 to 99.31% and R² were > 0.98 (Appendix 4, section 9.4). Genorm highlighted that all three HKGs were stably expressed across both resistant and susceptible *N. ribisnigri* biotypes (M < 1.5). The primer efficiency (E) for *RPL13* was < 95% and was excluded from the validation of the 2698 gene. The primer efficiencies (E) for the 2698 gene were 101.25 and 102.73% with an R² of > 0.98 (Appendix 4, section 9.4). Only one primer pair was used for the gene expression analysis. Both *RPS9* and *RPS18* were used for the qPCR normalisation experiment for the gene expression analysis of gene-2698.

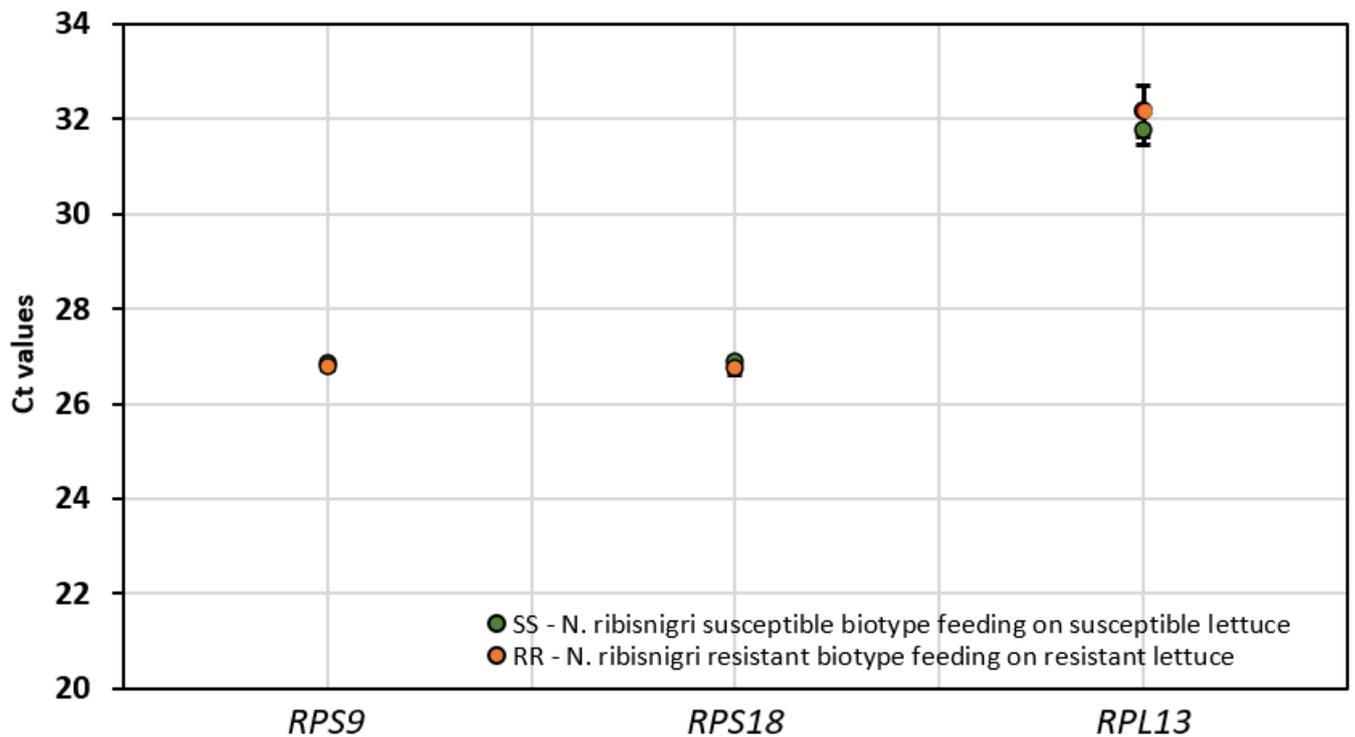


Figure 4.8: Ct values (\pm SE) obtained for the three different HKGs under two experimental conditions. Green circles represent susceptible *N. ribisnigri* feeding on susceptible lettuce, orange circles represent resistant *N. ribisnigri* feeding on resistant lettuce. Three biological and three technical replicates were used to validate the HKGs.

The gene expression of gene-2698 was highly expressed in the resistance-breaking *N. ribisnigri* biotypes compared to the susceptible *N. ribisnigri* biotype (Fig. 4.9). Statistical analysis (t-test – MeV package) confirmed that WT_Kent (SS) was significantly different ($F = 96.54$, $P < 0.001$) from Kent_CL (RR), UK631 (RR) and UK631 (RS). Resistance-breaking *N. ribisnigri* biotypes fed on lettuce containing the Nr-gene (cv. Eluarde) and without (cv. Pinokkio) had similar gene expression levels (non-significant). A significant reduction in gene expression was observed in gene 2698 in *N. ribisnigri* unable to feed on lettuce containing the Nr-gene (Fig. 4.9).

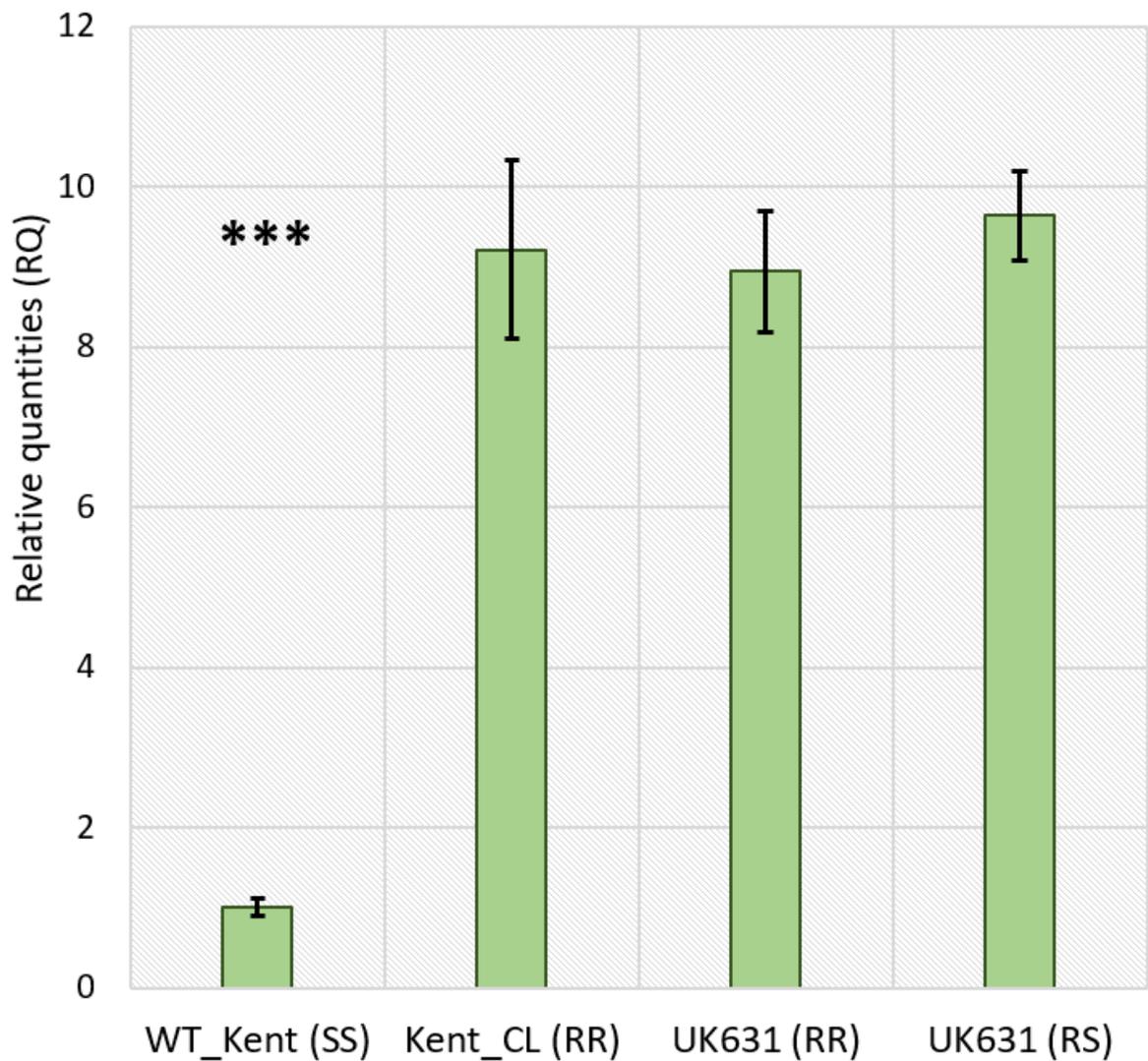


Figure 4.9: Comparison of gene-2698 expression levels using qRT-PCR between four conditions of *N. ribisnigri* cultures feeding on resistant (cv. Eluarde) and susceptible (cv. Pinokkio) lettuce cultivars. Both Kent_CL and UK631 are resistant-breaking biotypes able to feed on resistant lettuce cultivars. WT Kent is a susceptible biotype unable to feed on the resistant lettuce cultivars containing the Nr-gene. RR – resistant *N. ribisnigri* fed on resistant lettuce cultivar (cv. Eluarde), RS – resistant *N. ribisnigri* fed on susceptible lettuce cultivar (cv. Pinokkio), SS – susceptible *N. ribisnigri* fed on susceptible lettuce cultivar (cv. Pinokkio). In total, three biological and three technical replicates were used to compare gene-2698 expression level. *** - $P < 0.001$. Both *RPS9* and *RPS18* were used for the qPCR normalisation for the gene expression analysis of gene-2698.

4.4 Discussion

Construction of the first reference genome for *Nasonovia ribisnigri*

This study is the first to assemble and annotate a draft reference genome of *N. ribisnigri* and has done so for both a resistant-breaking and susceptible biotype. Prior to this project, the only genes of the currant-lettuce aphid (*N. ribisnigri*) that had been sequenced were a selection of odorant binding proteins. As expected from the assembly statistics, the susceptible Nr:0 genome had the highest completeness score (94.2% complete and single copy) and the resistant biotype Nr:1 had the highest assembly statistics and completeness (92.9% complete and single copy) (Table 4.6). These scores represent a high standard of completeness, compared to previously published aphid genomes (Fig. 4.1) (Biello *et al.*, 2021; Mathers, 2020; Li *et al.*, 2019; Shahid *et al.*, 2021). The removal of both the aphid symbiont (*Buchnera aphidicola*) and host plant (*Lactuca sativa*), overall, further improved both the assembly statistics and completeness of the genomes by ~1%. Since all *N. ribisnigri* used for genome assembly were collected from long term laboratory cultures, the risk of contamination from parasitoid wasps was incredibly low and not evident in the NCBI database searches of the genome sequence. However, when collecting from a field environment, it has been highlighted that contamination from parasitoid wasp larvae in aphids is very possible (Mathers, 2020) and this could affect downstream comparative analyses (Koutsovoulos *et al.*, 2016). It is therefore good practice, when collecting field caught aphid specimens for genome assembly or sequence analysis, to first establish a culture in a laboratory environment to clarify whether any parasitoid wasps or other potential contaminants are present prior to DNA extraction and sequencing.

From the assembled genomes, there was a discrepancy between genome size of the Nr:0 and Nr:1 biotypes (Table 4.6). The genome assembles for UK631 (Nr:0) and Ely (Nr:1) biotypes were similar in genome size (~366Mb), N50, number of scaffolds and completeness (BUSCO) but this is likely a result of lower assembly qualities. The higher quality genome assembles of Nr8 (Nr:0), WT Kent (Nr:0) and Kent CL (Nr:1) in fact differ in genome size, 423Mb and 384Mb, respectively. This highlights the advantage of having access to several biotypes for genome assembly projects and

helps prevent drawing spurious conclusions from lower quality assemblies. This difference in genome size between the two biotypes could be attributed to the ability of the Nr:1 biotype to break host plant resistance in *L. sativa* containing the Nr-gene but is more likely an artefact of the sequencing process. Further research would be required to confirm this observation.

The genome size of both the susceptible and resistant-breaking *N. ribisnigri* biotypes were shown to be comparative to other aphid species available. The pea aphid (*Acrythosiphum pisum*) genome is estimated to be a size of ~446.6Mb (The International Aphid Genome Consortium, 2010), a draft assembly of the soybean aphid (*Aphis glycines*) (Matsumura, 1917) has been estimated at 317.1Mb (Wenger *et al.*, 2017), the corn leaf aphid (*Rhopalosiphum maidis*) (Fitch, 1856) near-complete (95.8%) genome is estimated at 321Mb (Chen *et al.*, 2019), Russian wheat aphid (*Diuraphis noxia*) (Kurdjumov, 1913) is 421Mb (Nicholson *et al.*, 2015), the peach-potato aphid (*Myzus persicae*) is estimated to be 409Mb (Wenger *et al.*, 2020) and an improved bird-cherry oat aphid (*Rhopalosiphum padi*) genome is estimated to be of 321Mb (Morales-hojas *et al.*, 2020).

Identification of the potential resistant-breaking gene in *N. ribisnigri*

This is also the first study to provide strong evidence that a single gene is responsible for the ability of *N. ribisnigri* to overcome the Nr-gene resistance in lettuce. The results of the DeSeq2 analysis highlighted that one gene was significantly up-regulated in the resistant-breaking *N. ribisnigri* biotypes (gene-2698). Despite other significantly up- and down-regulated DE genes in the analysis, gene-2698 was very apparent in the volcano plot (Fig. 4.5). In addition, gene-2698 was the only shared DE gene between the two *N. ribisnigri* resistance-breaking biotypes, as shown in the Venn diagram (Fig. 4.7). Further analysis confirmed that gene-2698 was the only DE gene which both resistant-breaking biotypes shared (Fig. 4.6A) and was the most likely candidate gene conferring resistance to the Nr-gene in the lettuce host plant. The qRT-PCR confirmed that this gene is significantly up-regulated in the resistant-breaking biotypes (Fig. 4.9). The monogenic Nr-gene in lettuce cultivars is thought to reside within the phloem of lettuce during sap ingestion from the sieve element but the exact resistance mechanism or pathway involved is still unknown and yet

to be described (ten Broeke *et al.*, 2013) (Chapter 1, section 1.8). The results of the present study therefore suggest that a gene-for-gene resistance exists between the Nr-gene in the lettuce host plant and a corresponding gene (gene-2698) in the resistant-breaking *N. ribisnigri* biotype.

Gene-for-gene resistance has been highlighted in several arthropod and arthropod-associated effectors (Rawat *et al.*, 2012; Wei *et al.*, 2009; Zhao *et al.*, 2015). The gene-for-gene hypothesis states that, 'for each gene conditioning resistance in the host, there is a corresponding gene conditioning pathogenicity in the parasite' (Flor, 1971; Kerr, 1987). Since then, evidence for gene-for-gene resistance has been published a number of times, with the first example being between the Hessian fly (*Matetiola destructor*) (Say, 1817) and wheat (*Triticum* spp.) (Hatchett and Gallun, 1970). More recent examples have identified several *R* genes (Gm1-Gm11) in rice (*Oryza sativa*) against the Asian rice gall midge (*Orseolia oryzae* (Wood-Mason)) which have been mapped in the rice genome (Himabindu *et al.*, (2009) and a gene (*BPH15*) in the brown plant hopper (*Nilaparvata lugens*) (Stal, 1854) that mediates a unique defence mechanism in rice (Lv *et al.*, 2014).

Interestingly, it was shown in the qRT-PCR results (Fig. 4.9) that gene-2698 is upregulated in resistant-breaking biotypes when feeding on susceptible lettuce not containing the Nr-gene as well as on the resistant lettuce. This suggests that gene-2698 is constitutively expressed regardless of the presence or absence of the Nr-gene in the lettuce host plant. In contrast, it has been shown that highly resistant *M. persicae* clones were able to 'switch off' (down-regulate) an esterase gene (*E₄*) which is responsible for the production of high levels of *E₄* and change DNA methylation (Field *et al.*, 1989). Furthermore, this change in DNA methylation was presumed to help reduce the energy cost of making additional esterase in the absence of pesticide (Field *et al.*, 1989).

The RNA-seq data used in the present study contained a low number of true biological replicates, as each sample is derived from ~50 individuals which were pooled for RNA extraction and sequencing. Out of the replicates, gene-2698 was shown to have consistently high RNA counts in the resistance-breaking *N. ribisnigri* biotype compared to the susceptible biotype, with little variation (Fig. 4.7 B). This suggests that, even with the low replication, gene-2698 is present in the resistant-breaking *N. ribisnigri* biotypes. In addition, the results from the qRT-PCR confirm this

gene is present and not an artefact of low RNA-seq sample replication. A minimum number of six replicates has been suggested to be used in RNA-seq experiments, rising to 12, when it is important to identify significantly differentiated genes between conditions for all fold changes (Schurch *et al.*, 2016). Due to the low true replication in this study, DESeq2 was selected for analysis. This tool is regarded as the most robust to identify significantly differentially expressed genes in low replicated RNA-seq experiments (Schurch *et al.*, 2016). In addition, gene-2698 was the only significantly differentially expressed gene to be shared between the two resistant-breaking biotypes at both a high and low alpha parameter (>0.00001 and 0.5 , respectively). There is a chance that other significantly expressed genes were not identified as effectively as may have been possible with a higher replication rate. To confirm no other significantly differentially expressed genes were missed due to low true biological replication, additional replicates would be required.

The overreliance and overuse of a single gene in lettuce (Nr-gene) has resulted in the evolution of resistant-breaking *N. ribisnigri* biotypes (Thabuis *et al.*, 2011). The gene-2698 identified in this study is a strong candidate for the ability of the resistant *N. ribisnigri* biotype to break the lettuce resistance of the single Nr-gene. By combining multiple *R* genes or deleting susceptibility loci greatly reduces the selection pressure (Delmotte *et al.*, 2016). If the pest evolves and can break one control gene, the plant is still protected by the remaining *R* genes. To provide more durable resistance in crop plants, it has been suggested to 'pyramid' *R* genes with mutations in susceptibility genes (Stuart, 2015). By pyramiding *R* genes in this way, generally results in an increase in resistance in 'pyramid crops' compared to monogenetic resistance (Mundt, 2018). This study suggests that pyramid *R* genes could be a more successful approach in controlling *N. ribisnigri* outbreaks in the future. However, pyramiding resistance genes in one cultivar can still be overcome if no other measures to reduce the selection pressure are applied (Barzman *et al.*, 2015).

Thorough homology searches were conducted on gene-2698 but unfortunately no similarities were found. It is currently categorised as a hypothetical protein in a few aphid species and would benefit from further investigation into its exact function. The resistance mechanism has previously

been associated with the phloem sieve elements, as susceptible *N. ribisnigri* biotypes were unable to feed on lettuce containing the Nr-gene (ten Broeke *et al.*, 2010). The use of excised lettuce leaves in experiments resulted in a loss of resistance which suggested that the feeding mechanism is 'mobile' (Liu and McCreight, 2006) but this result could be due to changes in metabolism resulting from the excision of the leaves (Gao *et al.*, 2008). The 'deterrent' effect of the resistance has been shown to not be toxic, as normal feeding behaviour and growth is resumed when susceptible aphids are transferred to susceptible lettuce cultivar (Van Helden *et al.*, 1993). Since the present study has identified a likely candidate gene involved in the ability of *N. ribisnigri* to overcome the Nr-gene resistance in lettuce, more targeted research can be conducted. For example, the use of gene-editing technologies such as CRISPR-Cas9 (Cong *et al.*, 2013) could be incorporated to mediate gene knockout of the SNPs identified in gene-2698 of the *N. ribisnigri* resistant-breaking biotypes. By knocking out gene-2698, it would be possible to determine whether this gene is responsible for the ability of resistance-breaking *N. ribisnigri* to feed on lettuce containing the Nr-gene (loss-of-function strategy) (Housden *et al.*, 2017). In addition, it could be possible to compliment this approach with a gain-of-function, in which susceptible *N. ribisnigri* could be modified to over express gene-2698 and therefore determine if they are able to feed on lettuce containing the Nr-gene (Zimmer *et al.*, 2018).

This study has provided strong foundations for future work on an aphid which has very few effective controls. The availability of a draft reference genome of both the resistant-breaking and susceptible *N. ribisnigri* biotypes provides researchers a tool to study *N. ribisnigri* in much greater resolution. Additionally, elucidating a potential candidate gene of resistance in the resistant-breaking biotype will enable targeted approaches to thoroughly investigate this elusive mechanism of resistance.

4.5 References

Aarts, R., Schut, J.W., Driessen, R. and Reinink, K., 1999. Integrated control of aphids on lettuce varieties resistant to *Nasonovia ribisnigri*. *MEDEDELINGEN-FACULTEIT LANDBOUWKUNDIGE EN TOEGEPASTE BIOLOGISCHE WETENSCHAPPEN*, pp.11-16.

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A. and Guerler, A., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, pp.537-544.

Anders, S. and Huber, W., 2010. Differential expression analysis for sequence count data. *Nature Precedings*, pp.1-1.

Andersen, C.L., Jensen, J.L. and Ørntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research*, pp.5245-5250.

Andrews, S. (n.d.). *FastQC A Quality Control tool for High Throughput Sequence Data*. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

- Bansal, R., Mian, M.R., Mittapalli, O. and Michel, A.P., 2013. Molecular characterization and expression analysis of soluble trehalase gene in *Aphis glycines*, a migratory pest of soybean. *Bulletin of entomological research*, 103(3), pp.286-295.
- Barber, M 2002, 'Insecticide resistance in the currant-lettuce aphid, *Nasonovia ribisnigri*, PhD thesis, University of Bath.
- Barriere, V., Lecompte, F. and Lescourret, F., 2015. Efficacy of pest and pathogen control, yield and quality of winter lettuce crops managed with reduced pesticide applications. *European Journal of Agronomy*, pp.34-43.
- Barzman, M., Bàrberi, P., Birch, A.N.E., Boonekamp, P., Dachbrodt-Saaydeh, S., Graf, B., Hommel, B., Jensen, J.E., Kiss, J., Kudsk, P. and Lamichhane, J.R., 2015. Eight principles of integrated pest management. *Agronomy for sustainable development*, pp.1199-1215.
- Bass, C., Puinean, A.M., Zimmer, C.T., Denholm, I., Field, L.M., Foster, S.P., Gutbrod, O., Nauen, R., Slater, R. and Williamson, M.S., 2014. The evolution of insecticide resistance in the peach potato aphid, *Myzus persicae*. *Insect biochemistry and molecular biology*, pp.41-51.
- Biello, R., Singh, A., Godfrey, C.J., Fernández, F.F., Mugford, S.T., Powell, G., Hogenhout, S.A. and Mathers, T.C., 2020. A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Molecular Ecology Resources*, pp.316-326.
- BIPAA, 2021. Bioinformatics Platform for Agroecosystem Arthropods. Available at: <https://bipaa.genouest.org/is/> (Accessed: 23/06/20).
- Blackman, R.L. and Eastop, V.F., 1984. *Aphids on the World's Crops, An Identification Guide*. First edn. Chichester: John Wiley & Sons.
- Blum M, Chang H, Chuguransky S, Grego T, Kandasamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A and Finn RD The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, Nov 2020, (doi: 10.1093/nar/gkaa977).
- Brisson, J.A. and Stern, D.L., 2006. The pea aphid, *Acyrtosiphon pisum*: an emerging genomic model system for ecological, developmental and evolutionary studies. *Bioessays*, pp.747-755.
- Campbell, M. S., Holt, C., Moore, B., & Yandell, M. 2014. Genome Annotation and Curation Using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, <https://doi.org/10.1002/0471250953.bi0411s48>
- Chen, W., Shakir, S., Bigham, M., Richter, A., Fei, Z. and Jander, G., 2019. Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). *Gigascience*, p.33.

- Collier, R.H., Tatchell, G.M., Ellis, P.R. and Parker, W.E., 1999. Strategies for the control of aphid pests of lettuce. *IOBC WPRS BULLETIN*, pp.25-36.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. and Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, pp.819-823.
- Delmotte, F., Bourguet, D., Franck, P., Guillemaud, T., Reboud, X., Vacher, C. and Walker, A.S., 2016. Combining selective pressures to enhance the durability of disease resistance genes. *Frontiers in plant science*, p.1916.
- Dewar, A.M., 2017. The adverse impact of the neonicotinoid seed treatment ban on crop protection in oilseed rape in the United Kingdom. *Pest management science*, pp.1305-1309.
- Diaz, B.M., Barrios, L. and Fereres, A., 2012. Interplant movement and spatial distribution of alate and apterous morphs of *Nasonovia ribisnigri* (Homoptera: Aphididae) on lettuce. *Bulletin of entomological research*, pp.406-414.
- Dunbar, H.E., Wilson, A.C.C., Ferguson, N.R. and Moran, N.A., 2007. Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS biology*, p.96.
- Eenink, A.H., Groenwold, R. and Dieleman, F.L., 1982. Resistance of lettuce (*Lactuca*) to the leaf aphid *Nasonovia ribis nigri*. 1. Transfer of resistance from *L. virosa* to *L. sativa* by interspecific crosses and selection of resistant breeding lines. *Euphytica*, pp.291-299.
- Field, L.M., Devonshire, A.L. and Forde, B.G., 1989. The combined use of immunoassay and a DNA diagnostic technique to identify insecticide-resistant genotypes in the peach-potato aphid, *Myzus persicae* (Sulz.). *Pesticide biochemistry and physiology*, pp.174-178.
- Flor, H.H., 1971. Current status of the gene-for-gene concept. *Annual review of phytopathology*, pp.275-296.
- Francis, W.R., Christianson, L.M., Kiko, R., Powers, M.L., Shaner, N.C. and Haddock, S.H., 2013. A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. *BMC genomics*, pp.1-12.
- 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*, pp.996-1009.
- Georghiou, G.P., 1983. Management of resistance in arthropods. In *Pest resistance to pesticides*. Springer, Boston, MA. pp. 769-792
- Görür, G., 2000. The role of phenotypic plasticity in host race formation and sympatric speciation in phytophagous insects, particularly in aphids. *Turkish Journal of Zoology*, pp.63-68.

Götz S., Garcia-Gomez JM., Terol J., Williams TD., Nagaraj SH., Nueda MJ., Robles M., Talon M., Dopazo J. and Conesa A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, pp.3420-3435.

Hatchett, J.H. and Gallun, R.L., 1970. Genetics of the ability of the Hessian fly, *Mayetiola destructor*, to survive on wheats having different genes for resistance. *Annals of the Entomological Society of America*, pp.1400-1407.

Himabindu, K., Suneetha, K., Sama, V.S.A.K. and Bentur, J.S., 2010. A new rice gall midge resistance gene in the breeding line CR57-MR1523, mapping with flanking markers and development of NILs. *Euphytica*, pp.179-187.

Hoff, K. J., Stanke, M. 2013. WebAUGUSTUS, a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic acids research*, gkt418. (use this reference for Blast2Go / Augustus)

Hough, Gemma L. (2013) *Biology and control of currant lettuce aphid nasonovia ribisnigri*. PhD thesis, University of Warwick. Chapter 2, pp.15-16.

Housden, B.E., Muhar, M., Gemberling, M., Gersbach, C.A., Stainier, D.Y., Seydoux, G., Mohr, S.E., Zuber, J. and Perrimon, N., 2017. Loss-of-function genetic tools for animal models: cross-species and cross-platform differences. *Nature Reviews Genetics*, pp.24-40.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J. and von Mering, C., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic acids research*, pp.309-314.

Huson, D.H., Auch, A.F., Qi, J. and Schuster, S.C., 2007. MEGAN analysis of metagenomic data. *Genome research*, pp.377-386.

International Aphid Genomics Consortium., 2010. Genome sequence of the pea aphid *Acythosiphon pisum*. *PLoS Biology* (online) 8, e100313.

Jiang, X., Zhang, Q., Qin, Y., Yin, H., Zhang, S., Li, Q., Zhang, Y., Fan, J. and Chen, J., 2019. A chromosome-level draft genome of the grain aphid *Sitobion miscanthi*. *Gigascience*, p.101.

Kerr, A., 1987. The impact of molecular genetics on plant pathology. *Annual Review of Phytopathology*, pp.87-110.

Kift, N.B., Mead, A., Reynolds, K., Sime, S., Barber, M.D., Denholm, I. and Tatchell, G.M., 2004. The impact of insecticide resistance in the currant-lettuce aphid, *Nasonovia ribisnigri*, on pest management in lettuce. *Agricultural and Forest Entomology*, pp.295-309.

Koramutla, M.K., Aminedi, R. and Bhattacharya, R., 2016. Comprehensive evaluation of candidate reference genes for qRT-PCR studies of gene expression in mustard aphid, *Lipaphis erysimi* (Kalt). *Scientific Reports*, pp.1-10.

- Koutsovoulos, G., Kumar, S., Laetsch, D.R., Stevens, L., Daub, J., Conlon, C., Maroon, H., Thomas, F., Aboobaker, A.A. and Blaxter, M., 2016. No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proceedings of the National Academy of Sciences*, pp.5053-5058.
- Li, B., Bickel, R.D., Parker, B.J., Ziabari, O.S., Liu, F., Vellichirammal, N.N., Simon, J.C., Stern, D.L. and Brisson, J.A., 2020. A large genomic insertion containing a duplicated follistatin gene is linked to the pea aphid male wing dimorphism. *ELife*, pp.50-68.
- Li, Y., Park, H., Smith, T.E. and Moran, N.A., 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Molecular biology and evolution*, pp.2143-2156.
- Liao, Y., Smyth, G.K. and Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, pp.923-930.
- Lin, B.B., 2011. Resilience in agriculture through crop diversification: adaptive management for environmental change. *BioScience*, pp.183-193.
- Liu, Y.B. and McCreight, J.D., 2006. Responses of *Nasonovia ribisnigri* (Homoptera: Aphididae) to susceptible and resistant lettuce. *Journal of economic entomology*, pp.972-978.
- Lombaert, E., Carletto, J., Piotte, C., Fauvergue, X., Lecoq, H., Vanlerberghe-Masutti, F. and Lapchin, L., 2009. Response of the melon aphid, *Aphis gossypii*, to host-plant resistance: evidence for high adaptive potential despite low genetic variability. *Entomologia Experimentalis et Applicata*, pp.46-56.
- Luethy, R. and Hoover, C., 2004. Hardware and software systems for accelerating common bioinformatics sequence analysis algorithms. *Drug Discovery Today: BIOSILICO*, pp.12-17.
- Lv, W., Du, B., Shangguan, X., Zhao, Y., Pan, Y., Zhu, L., He, Y. and He, G., 2014. BAC and RNA sequencing reveal the brown planthopper resistance gene BPH15 in a recombination cold spot that mediates a unique defense mechanism. *BMC genomics*, pp.1-16.
- MacIntosh, G.C., 2019. Gene pyramids and the balancing act of keeping pests at bay. *Journal of experimental botany*, pp.4591-4593.
- Mathers, T.C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S.T., Baa-Puyoulet, P., Bretaudeau, A., Clavijo, B., Colella, S., Collin, O. and Dalmay, T., 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome biology*, pp.1-20.
- Mathers, T.C., Mugford, S.T., Hogenhout, S.A. and Tripathi, L., 2020. Genome sequence of the banana aphid, *Pentalonia nigronervosa* Coquerel (Hemiptera: Aphididae) and its symbionts. *G3: Genes, Genomes, Genetics*, pp.4315-4321.
- Mathers, T.C., Wouters, R.H., Mugford, S.T., Swarbreck, D., Van Oosterhout, C. and Hogenhout, S.A., 2021. Chromosome-scale genome assemblies of aphids reveal extensively rearranged

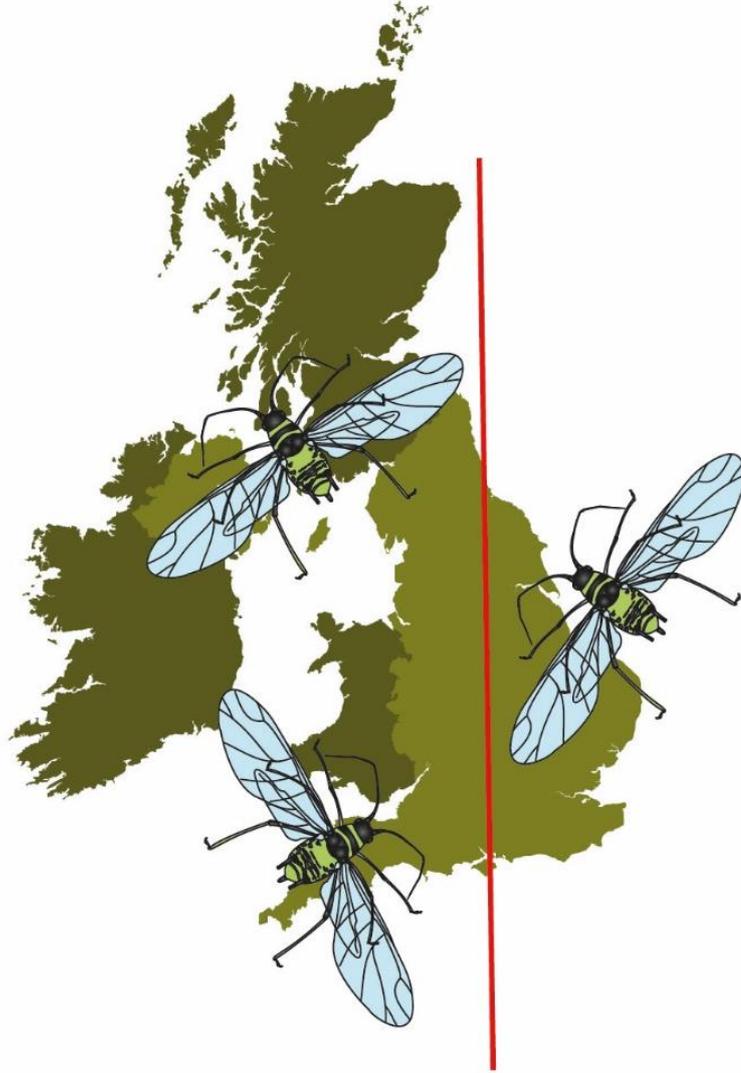
- autosomes and long-term conservation of the X chromosome. *Molecular biology and evolution*, 38(3), pp.856-875.
- Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N. and Muños, S., 2016. Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. *Biotechniques*, pp.203-205.
- Morales, I., Diaz, B.M., De Mendoza, A.H., Nebreda, M. and Fereres, A., 2013. The development of an economic threshold for *Nasonovia ribisnigri* (Hemiptera: Aphididae) on lettuce in central Spain. *Journal of Economic Entomology*, pp.891-898.
- Morales-Hojas, R., Gonzalez-Uriarte, A., Alvira Iraizoz, F., Jenkins, T., Alderson, L., Kruger, T., Hall, M.J., Greenslade, A., Shortall, C.R. and Bell, J.R., 2020. Population genetic structure and predominance of cyclical parthenogenesis in the bird cherry-oat aphid *Rhopalosiphum padi* in England. *Evolutionary applications*, pp.1009-1025.
- Mundt, C.C., 2018. Pyramiding for resistance durability: theory and practice. *Phytopathology*, pp.792-802.
- Natwick, E.T. and Lopez, M.I., 2016. Insecticide Efficacy Against Aphids in Leaf Lettuce, 2015. *Arthropod Management Tests*, p.39.
- Nault, L.R., 1997. Arthropod transmission of plant viruses: a new synthesis. *Annals of the entomological Society of America*, pp.521-541.
- Nicholson, S.J., Nickerson, M.L., Dean, M., Song, Y., Hoyt, P.R., Rhee, H., Kim, C. and Puterka, G.J., 2015. The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC genomics*, pp.1-16.
- OmicsBox - Bioinformatics made easy. BioBam Bioinformatics (Version 1.4.11). March 3, 2019. www.biobam.com/omicsbox.
- Parra, G., Bradnam, K. and Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, pp.1061-1067.
- Puinean, A.M., Foster, S.P., Oliphant, L., Denholm, I., Field, L.M., Millar, N.S., Williamson, M.S. and Bass, C., 2010. Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLOS genetics*, p.e1000999.
- R CORE TEAM 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rawat, N., Neeraja, C.N., Nair, S. and Bentur, J.S., 2012. Differential gene expression in gall midge susceptible rice genotypes revealed by suppressive subtraction hybridization (SSH) cDNA libraries and microarray analysis. *Rice*, pp.1-15.

- Reinink, K. and Dieleman, F.L., 1989. Comparison of sources of resistance to leaf aphids in lettuce (*Lactuca sativa* L.). *Euphytica*, pp.21-29.
- Reinink, K. and Dieleman, F.L., 1993. Survey of aphid species on lettuce. *Survey of aphid species on lettuce.*, pp.56-68.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. and Mesirov, J.P., 2011. Integrative genomics viewer. *Nature biotechnology*, pp.24-26.
- Rufingier, C., Schoen, L., Martin, C. and Pasteur, N., 1997. Resistance of *Nasonovia ribisnigri* (Homoptera: Aphididae) to five insecticides. *Journal of Economic Entomology*, pp.1445-1449.
- Schmittgen, T.D. and Livak, K.J., 2008. Analyzing real-time PCR data by the comparative CT method. *Nature protocols*, pp.1101-1108.
- Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T. and Blaxter, M., 2016. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?. *Rna*, pp.839-851.
- Segoli, M. and Rosenheim, J.A., 2012. Should increasing the field size of monocultural crops be expected to exacerbate pest damage?. *Agriculture, ecosystems & environment*, pp.38-44.
- Shahid, S., Shi, Y., Yang, C., Li, J., Ali, M.Y., Smagghe, G. and Liu, T.X., 2021. CCHamide2-receptor regulates feeding behavior in the pea aphid, *Acyrtosiphon pisum*. *Peptides*, p.170596.
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, pp.3210-3212.
- Singh, K.S., Cordeiro, E.M., Troczka, B.J., Pym, A., Mackisack, J., Mathers, T.C., Duarte, A., Legeai, F., Robin, S., Bielza, P. and Burrack, H.J., 2021. Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest *Myzus persicae*. *Communications Biology*, pp.1-16.
- Smith, A., Hubley, R. and Green, P., 2013. RepeatMasker Open-4.0. *RepeatMasker Open-4.0*.
- Smith, C.M. and Chuang, W.P., 2014. Plant resistance to aphid feeding: behavioral, physiological, genetic and molecular cues regulate aphid host selection and feeding. *Pest management science*, pp.528-540.
- Stuart, J., 2015. Insect effectors and gene-for-gene interactions with host plants. *Current Opinion in Insect Science*, pp.56-61.
- Tagu, D., Klingler, J.P., Moya, A. and Simon, J.C., 2008. Early progress in aphid genomics and consequences for plant-aphid interactions studies. *Molecular plant-microbe interactions*, pp.701-708.

- ten Broeke, C.J., Dicke, M. and van Loon, J.J., 2013. Feeding behaviour and performance of different populations of the black currant-lettuce aphid, *Nasonovia ribisnigri*, on resistant and susceptible lettuce. *Entomologia Experimentalis et Applicata*, pp.130-141.
- Thabuis, A.P.P., Teekens, K.C. and Van Herwijnen, Z.O., Rijk Zwaan Zaadteelt en Zaadhandel BV., 2014. *Lettuce that is resistant to the lettuce aphid Nasonovia ribisnigri biotype 1*. U.S. Patent 8,829,280.
- TimeLogic Division. (2013) *Accelerated BLAST Performance with Tera-BLAST™: a comparison of FPGA versus GPU and CPU BLAST implementations*. Available at: http://www.timelogic.com/documents/TimeLogic_Tera-BLAST_whitepaper_v1.0.pdf (Accessed: 13/09/2021).
- Tjallingii, W.F. and Esch, T.H., 1993. Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals. *Physiological entomology*, pp.317-328.
- Tooker, J.F. and Frank, S.D., 2012. Genotypically diverse cultivar mixtures for insect pest management and increased crop yields. *Journal of Applied Ecology*, pp.974-985.
- Troccka, B.J., Singh, K.S., Zimmer, C.T., Vontas, J., Nauen, R., Hayward, A. and Bass, C., 2021. Molecular innovations underlying resistance to nicotine and neonicotinoids in the aphid *Myzus persicae*. *Pest Management Science*, pp.5311-5320.
- van der Arend, A.J., 2003. The possibility of *Nasonovia ribisnigri* resistance breaking biotype development due to plant host resistance: a literature study. *Eucarpia leafy vegetables*, pp.75-81.
- Van Emden, H.F. and Harrington, R. eds., 2017. *Aphids as crop pests*. Cabi.
- van Helden, M., 1995. *The resistance of lettuce to the aphid Nasonovia ribisnigri*. Wageningen University and Research.
- van Helden, M., Tjallingii, W.F. and Dieleman, F.L., 1993. The resistance of lettuce (*Lactuca sativa* L.) to *Nasonovia ribisnigri*: bionomics of *N. ribisnigri* on near isogenic lettuce lines. *Entomologia Experimentalis et Applicata*, pp.53-58.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, pp.1-12.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K. and Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS one*, p.e112963.
- Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V. and Zdobnov, E.M., 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular biology and evolution*, pp.543-548.

- Wei, Z., Hu, W., Lin, Q., Cheng, X., Tong, M., Zhu, L., Chen, R. and He, G., 2009. Understanding rice plant resistance to the Brown Planthopper (*Nilaparvata lugens*): a proteomic approach. *Proteomics*, pp.2798-2808.
- Wenger, J.A., Cassone, B.J., Legeai, F., Johnston, J.S., Bansal, R., Yates, A.D., Coates, B.S., Pavinato, V.A. and Michel, A., 2020. Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect biochemistry and molecular biology*, p.102917.
- Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. and Feolo, M., 2007. Database resources of the national center for biotechnology information. *Nucleic acids research*, pp.13-21.
- Wingett, S.W. and Andrews, S., 2018. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*, 7.
- Yang Liao, Gordon K. Smyth, Wei Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics*, Volume 30, Issue 7, 1 April 2014, Pages 923–930, <https://doi.org/10.1093/bioinformatics/btt656>.
- Zhao, C., Escalante, L.N., Chen, H., Benatti, T.R., Qu, J., Chellapilla, S., Waterhouse, R.M., Wheeler, D., Andersson, M.N., Bao, R. and Batterton, M., 2015. A massive expansion of effector genes underlies gall-formation in the wheat pest *Mayetiola destructor*. *Current Biology*, pp.613-620.
- Zhu, A., Ibrahim, J.G., Love, M.I. (2018) Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bty895>.
- Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L. and Yorke, J.A., 2013. The MaSuRCA genome assembler. *Bioinformatics*, pp.2669-2677.
- Zimmer, C.T., Garrood, W.T., Singh, K.S., Randall, E., Lueke, B., Gutbrod, O., Matthiesen, S., Kohler, M., Nauen, R., Davies, T.E. and Bass, C., 2018. Neofunctionalization of duplicated P450 genes drives the evolution of insecticide resistance in the brown planthopper. *Current Biology*, pp.268-274.

Chapter 5



CHAPTER 5: Population genetic analysis of *Nasonovia ribisnigri* reveals a temporal divide between the East and West of England

5.1 Introduction

Population genetic analysis is the study of genetic variation between and within populations, specifically the change in allele frequencies over time and space (Catchen *et al.*, 2017). In relation to the management of horticultural pests, population genetics can inform the scale at which populations should be managed. The level of genetic variation can change according patterns of immigration and emigration from both local and geographically distant sources and change in response to mortality rates (Vialatte *et al.*, 2007). Immigrants can introduce genetic variation and enhance the genetic pool of a population and increase its population fitness, whilst emigration will reduce population size and therefore the genetic pool (Keller *et al.*, 2001; Via *et al.*, 1991). The introduction of genetic variation can have direct impacts on the conservation and ecology of vulnerable species (Allendorf, 2017; López-Uribe *et al.*, 2017); and the spread of insecticide resistant populations (Dusfour *et al.*, 2019; Pu *et al.*, 2019; Naveen *et al.*, 2020). Insecticides impose high selection pressure on populations if insects can survive, driving a rapid evolutionary change towards 'breaking' of the resistant mechanisms through either target site or metabolic resistance (Schoville *et al.*, 2018; Swale *et al.*, 2016;). Insecticide resistance alleles can be incorporated to study how allele frequency changes in response to insecticide pressures and therefore, the spread of resistance in space and time (Rinkevich *et al.*, 2007). Combining these resistant alleles, such as cytochrome P450 oxidases and knockdown resistance (*kdr*), with microsatellite loci enables the patterns of movement and the effect of insecticide treatment on the genetic structure of a pest population to be studied (Franck *et al.*, 2007).

The use of microsatellite markers has been used to show host range expansion (Peccoud *et al.*, 2008, Harrison and Mondor, 2011), to quantify the level of gene flow between populations (Orantes *et al.*, 2012), host plant specialisation and biotypes (Vialatte *et al.*, 2005, Via and Hawthorne, 2002) and the ancestry of asexual and sexual morphs (Simon *et al.*, 1999, Halkett *et al.*, 2005). The information

obtained from such studies can provide vital information on aphid migration, dispersal, host plant variation and evolution and spread of resistant biotypes to inform on future mitigative strategies. Whilst population genetics is not dependent on a reference genome (Peterson *et al.*, 2012), having one is advantageous because it enables the identification of resistant genes (Bus *et al.*, 2008; Hill *et al.*, 2012) and other novel control strategies (Heck and Brault, 2018).

Lifecycle plasticity is key to understanding why aphids are pests. Aphids are particularly well adapted to overcome high selection pressures that insecticides and host plant resistance impose on their populations (Loxdale *et al.*, 2020). At their disposal is a combination of asexual and sexual reproduction, parthenogenesis, complex host plant and predator interactions and the ability to disperse over large distances, which all contribute to their success (Sunnucks *et al.*, 1997; Simon and Peccoud, 2018; Roitberg *et al.*, 1979). The incorporation of population genetics will improve our understanding of the population structure, biology and ecology of horticulturally important aphids to inform future management practises and mitigate and minimise crop losses. This is especially important for aphids that are difficult to forecast and have sporadic outbreaks in crops, with limited control.

The currant-lettuce aphid, *Nasonovia ribisnigri* (Mosely), is the most damaging pest aphid on outdoor lettuce crop in the UK, Europe, America and New Zealand (Reinink and Dieleman, 1993; Sauer-Kesper *et al.*, 2011; Stufkens *et al.*, 2002; Nebreda *et al.*, 2004). It is a heteroecious species, host-alternating between currant species (*Ribes spp.*) during the winter/spring and members of the Asteraceae family, that include lettuce (*Lactuca sativa*) on which it feeds during the summer/autumn. Outdoor lettuce is associated with significant crop losses due to the presence of *N. ribisnigri* in the lettuce head, leading to unmarketable, product, due to the inability of foliar insecticides to penetrate and make contact with the enclosed aphids (Natwick and Lopez, 2016). A *Nasonovia*-resistant (*Nr*) lettuce variety was introduced in 1999 and conferred complete resistance to *N. ribisnigri* (ten Broeke *et al.*, 2013). The *Nr*-gene originates from bitter lettuce (*Lactuca virosa*), a wild lettuce species, which was inter-specifically crossed and transferred to lettuce (*L. sativa*) using a bridging species (prickly lettuce, *L. serriola*) (Eenink *et al.*, 1982). Since the *Nr* gene in lettuce cultivars only provided single-gene

resistance, resistant-breaking biotypes were first reported eight years after their first use in 2007 in Europe (Thabuis *et al.*, 2014). *N. ribisnigri* has also shown to develop resistance to insecticides (Barber *et al.*, 1999) and, the removal of systemic neonicotinoids in the UK for non-flowering crops has left the lettuce industry with few management options. Like most aphids, *N. ribisnigri* can either undergo sexual (holocyclic) or asexual (anholocyclic) reproduction (Blackman and Eastop, 2000). A holocyclic population undergoes parthenogenic reproduction except during the autumn/winter months when sexual reproduction occurs and produces overwintering eggs that hatch the following spring. Populations of anholocyclic individuals only undergo parthenogenic reproduction and overwinter as either adults or nymphs and are generally associated with persisting due to higher winter temperatures. Although *N. ribisnigri* are holocyclic, they are considered anholocyclic in temperate regions such as Spain (Nieto Nafria, 1974). Despite its pest status in the UK, Europe, America and New Zealand, the level of gene flow and population structure have not been studied.

In this study, the discovery of novel microsatellite markers in *N. ribisnigri* from a newly *de-novo* assembled genome for the use in population genetic analysis is described. The genetic variation in *N. ribisnigri* has been explored both spatially and temporally in England over a 17-year period using archived samples from the Rothamsted Insect Survey. The study aimed to a) quantify the level of gene flow between populations in England and infer dispersal potential; b) understand whether populations evolve as a response to environmental change, both spatially and temporally; c) determine whether changes in population structure could be attributed to the development of resistant biotypes to the *nr-gene* in lettuce from 2008 onwards and, d) determine the status of anholocyclic reproduction. These aims will provide vital pest ecological and evolutionary information on a species of aphid that is poorly understood and the most damaging to outdoor lettuce. It is expected that this science will deliver insight into future pest control management strategies as well as insight into the past and current demographics of *N. ribisnigri*.

5.2 Materials and methods

5.2.1 Study organism and site locations

All *N. ribisnigri* used in this study were collected from the Rothamsted Insect Survey's (RIS) 12.2m suction trap network between 2003-2020. Aphids caught in the suction traps were identified to species and stored in 100% ethanol:glycerol solution at a ratio of 95:5 in a sample archive. Over 500 *N. ribisnigri* from 10 sites in England (Fig. 5.1) were removed from the archive and then stored in 100% ethanol at -20°C prior to analysis. Due to the inherently low catch incidence of *N. ribisnigri* in the traps, the criteria for site selection merely imposed that there were more than one individual per site-year and that this was caught between April-August, and that the trap had been in operation since 2003, although York (2007) was an exception because it was strategically placed. Due to aphids being small soft-bodied insects which can easily be damaged in the collection and storage process, only *N. ribisnigri* that were undamaged were selected to avoid low DNA yield and reduce the likelihood of contamination in downstream analysis.



Figure 5.1: The Rothamsted Insect Survey (RIS) suction trap sites throughout the UK that were used in the study.

5.2.2 DNA extraction

DNA extractions were conducted on whole individual aphids using the Qiagen QIAmp DNA Micro Kit (Qiagen, Manchester, UK). See Chapter 2 (section 2.3.3) for complete DNA extraction method.

5.2.3 Microsatellite discovery

The reference genome for *N. ribisnigri* (biotype – Nr8) was *de novo* assembled using MaSuRCA (Maryland Super Read Cabog Assembler) with both short (Illumina) and long reads (nanopore, Oxford NANOPORE Technologies) (Chapter 5). A total of 4,778 scaffolds were uploaded into Msatcommander 0.8.2 (Faircloth, 2008) for the detection of microsatellites following the set parameters previously used in the literature (Jun *et al.*, 2011) (Table 5.1). All discovered microsatellites (including flanking regions) were manually selected using Geneious 10.1.3 software (<https://www.geneious.com>). Microsatellites located near to the end of sequence scaffolds were not used due to possible duplications. Forward and reverse primers were designed for each microsatellite in the flanking regions using Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>). A total of 68 microsatellites were selected for PCR (Polymerase Chain Reaction) amplification and sample validation. Microsatellite primers were obtained from Sigma-Aldrich.

Table 5.1: Msatcommander search parameters for microsatellite discovery and the number of microsatellites discovered and passing initial Primer3 parameters.

Motif	Minimum no. of repeats	Microsatellites discovered
Di-nucleotide	8	25
Tri-nucleotide	5	36
Tetra-nucleotide	4	5
Penta-nucleotide	4	1
Hexa-nucleotide	4	1
Total	-	68

5.2.4 Validation, PCR amplification and sequencing of microsatellites

To validate all 68 designed microsatellite primer pairs, seven *N. ribisnigri* cultures were used to determine polymorphism level of each locus using PCR amplification. These samples were representative of the samples being analysed as the cultures originated from various locations throughout the UK. All seven cultures were provided by Warwick Crop Centre from long established colonies of *N. ribisnigri* which were brought to Rothamsted Research to establish isogenic lines of each culture for microsatellite validation (Table 5.2). Each culture was reared on a lettuce (*Lactuca sativa* cv. Pinnokkio) host plant at 20°C, 16:8 (Light:Dark). Isogenic lines of each culture were established by selecting one apterous (founding mother) from each colony onto a new *L. sativa* (cv. Pinnokkio) host plant, in new separate insect rearing chambers. Once the founding mother had produced between 5-10 nymphs, they were removed, and the new isogenic culture left to establish for around 7 days.

Table 5.2: Experimental design of the *N. ribisnigri* cultures that were used for microsatellite validation. Nr: 0 = *Nr*-gene susceptible; Nr: 1 = *Nr*-gene susceptible; Pyrethroid R = pyrethroid resistant; Insecticide R = insecticide resistant; (S) = susceptible (does not contain *Nr*-gene); (R) = resistant (contains *Nr*-gene).

<i>N. ribisnigri</i> culture	Host plant	Culture origin
Nr4	Pinokkio (S)	Chichester, 1999
Nr8	Pinokkio (S)	York, 1999
Nr29	Pinokkio (S)	Suffolk, 1999
4850a	Pinokkio (S)	Lincolnshire, 2003 (Hough, 2013)
WT Kent	Pinokkio (S)	Kent, 2010 (Hough, 2013)
Kent CL	Pinokkio (S)	Kent, 2009 (Hough, 2013)

Reactions were performed in a total volume of 10ul containing 1ul of 10ng/ul of DNA, 5ul of primer master mix (ThermoFisher Scientific), 3.55ul molecular-grade water, 0.2ul of M13 Tail SSR (FAM), 0.05ul of M13 forward-tailed primer (fluorescent labelled dyes; 6-FAM, VIC, NED & PET (ThermoFisher Scientific)) and 0.2ul of reverse primer. All reactions were run on a thermocycler with amplification cycle: 94°C for 1 min; 30 cycles of 94°C for 40 s, annealing at each locus optimal temperature for 1 min, 72°C for 30 s; 8 cycles of 94°C, 56°C for 1 min, 72°C for 30 s; hold 20°C. The annealing temperature was optimised for each microsatellite marker and ranged between 50-58°C, with 2°C increments. For

sequencing, individual reactions were made to a final volume of 10ul (8.8ul Hi Dye Formamide (Applied Biosystems™), 0.2ul Liz 500 Size Standard (ThermoFisher Scientific), 1ul of PCR product. A final extension of 94°C for 6-10 min on the thermocycler was conducted. Immediately after the cycle, samples were transferred to ice for rapid cooling and either sequenced immediately or stored in the fridge for sequencing. The amplicons generated were individually sequenced using the ABI-3100 sequencer (Applied Biosystems™). Microsatellites were scored using the microsatellite plugin for Geneious. Microsatellite analysis was conducted following the recommendation in the online tutorial www.geneious.com/tutorials/microsatellites. In brief, each microsatellite marker for every sample was inspected to confirm the Liz 500 size standard was called correctly and manually adjusted if not. Only the 6-FAM fluorescent dye was used for microsatellite validation (Table 5.3). Each microsatellite was checked for peaks and any peaks present that were within range (e.g. 6-FAM range is 160-450) but not called, were corrected. Once peaks were called, bins were predicted for each sample based on the observed peaks and their size, using the integrated sizing algorithm in Geneious. After this step, an allele table was generated which contained the microsatellite peaks and their size and exported in a csv file for further analysis.

Table 5.3: The fluorescent dye parameters used for microsatellite analysis.

Trace	Expected no. of peaks	Repeat units	Range Start
6-FAM	2	4	160-450
VIC	2	4	200-280
NED	2	4	200-320
PET	2	4	140-480

Initially, only the tri, tetra, penta and hexa-nucleotide microsatellites were sequenced to reduce sequencing costs. From these 43 microsatellites, nine had peaks that were within the 6-FAM range, with only three microsatellites being amplified successfully in one out of the seven *N. ribisnigri* samples tested. A lower annealing temperature (52°C) in the PCRs improved amplification. The remaining 25 dinucleotides were sequenced to identify whether there were more suitable microsatellites. From these 25 microsatellites, six successfully amplified all seven *N. ribisnigri* samples. In total, 15 microsatellite primer pairs were identified (Table 5.4). Eight microsatellites were incorporated and used for the final analysis due to limited DNA quantities obtained from some of the *N. ribisnigri* samples.

Table 5.4: The 15 identified and developed microsatellites for *Nasonovia ribisnigri* and their F/R primers sets. Highlighted in bold are the eight microsatellites used for this study. Ta(°C): Annealing temperature; N_A: Number of alleles.

Locus	Motifs repeat	Primers (5' → 3')	Ta(°C)	Size range (bp)	N _A	HO	HE	Accession no.
Nrib7	(AT)10	F: CACGACGTTGTAACGACTCCGGACATCTTCTTCGCGT R: TCGACTTGACCCGATTCGC	50	160-175	9	0.61	0.75	
Nrib15	(AG)15	F: CACGACGTTGTAACGACTCAACAGTCAACATCCCCAGTGT R: TAGTCCACGCCCAACTGCAT	50	125-140	21	0.47	0.85	
Nrib26	(AGC)5	F: CACGACGTTGTAACGACTGCCTTCTCGTACTGTGTATCG R: ACCCTTGACGAAAAGAAAACACG	58	182-196	12	0.37	0.79	
Nrib30	(AAT)5	F: CACGACGTTGTAACGACTTTGGATATGAGACGAGCGGC R: GGCAAGTTTAAGGGCGACGG	50	175-186	4	0.16	0.66	
Nrib31	(AAT)6	F: CACGACGTTGTAACGACAGCTGTTTTAATGCGTGCGACA R: GATCCCTATCTCCGTCCTCCG	58	192-205	6	0.30	0.57	
Nrib32	(AAT)5	F: CACGACGTTGTAACGACGGTGGTGGTGAGGAAGAGCA R: CACGTGCAGGTCGAGTTCAA	62	180-188	10	0.46	0.75	
Nrib48	(AAG)5	F: CACGACGTTGTAACGACAGGTTTTGTGTTGGCAGCGA R: AGGCAAGGTATATCGGCAGACC	60	144-162	-	-	-	
Nrib49	(AAC)5	F: CACGACGTTGTAACGACCTTTTGCCTTGGCGTCCGAC R: AATACACCACCGCCACCACC	62	140-156	-	-	-	
Nrib51	(AAT)6	F: CACGACGTTGTAACGACTGGCGTATGTTTTCGTGGGT R: CCCACTTGCCACCTAGCTT	58	120-134	9	0.41	0.59	
Nrib53	(ACG)5	F: CACGACGTTGTAACGACGCGCAGAGATTTCCCGTCCG R: ACCTGATCATCGGTCACACAAC	58	130-143	-	-	-	
Nrib58	(AAT)6	F: CACGACGTTGTAACGACTAAATGGGCGTCGGGTTGGA R: GGGTAACTGGCTCGATCGCT	62	104-118	-	-	-	
Nrib59	(AAT)6	F: CACGACGTTGTAACGACACCACAGTTATCCGAGTCCAGA R: ATTCAACTCAAACGCGCTCTAGA	58	108-120	10	0.39	0.47	
Nrib65	(ACAT)4	F: CACGACGTTGTAACGACGGCAGTCGACAGAACCCT R: TCTGCCGGTCGTCCCTTATTT	60	114-130	-	-	-	
Nrib66	(AAAG)7	F: CACGACGTTGTAACGACGCGAGGCCCGTAATTAACAGC R: CCGTCTCTCCCTTGCACTC	62	167-182	-	-	-	
Nrib67	(ATATT)4	F: CACGACGTTGTAACGACTGAGTATTTGCCGCCGGTCT R: CCGCTCGAGTGTGTGTACGA	62	161-177	-	-	-	

5.2.5 Population analysis

To study the genetic differences between populations prior- and post-cultivar resistance breakdown in England (2007), samples were grouped into three categories (2003-2007, 2008-2014 and 2015-2020) (Table 5.5). The second grouping (2008-2014) was incorporated to determine whether a difference in population structure could be observed up to six years after host plant resistance breakdown. The third grouping (2015-2020) was used to understand whether any other longer-term population structure changes had occurred. Additionally, by separating the analysis into three similar number of years, provided a relatively even distribution of representative samples to be analysed and reduce bias from the low sample number. Each year grouping for every site contained a minimum of four individuals. A total of 146 currant-lettuce aphids (*Nasonovia ribisnigri*) were used over 10 locations in the final experiment (Fig. 5.1).

Table 5.5: Number of *Nasonovia ribisnigri* individuals collected from the Rothamsted Insect Survey (RIS) archive and used in the current population genetics study. York suction trap site was not established pre 2008 and therefore no samples were collected.

Sites	2003-2007	2008-2014	2015-2020	Total (2003-2020)
Brooms Barn	6	4	4	14
Hereford	5	5	4	14
Kirton	4	4	6	14
Preston	4	5	4	13
Rothamsted	5	4	6	15
Starcross	6	5	4	15
Wellesbourne	8	4	9	21
Writtle	6	4	4	14
Wye	4	6	4	14
York	-	5	7	12
Total	48	46	52	146

Observed and expected heterozygosities (H_O and H_E), number of alleles per locus by group were obtained using the R packages *ade4* (v.1.7-16) and *adegenet* (v.2.1.3) for analysing genetic and genomic data in RStudio (v.4.0.3) (Jombart, 2008; RStudio Team, 2020). A Principal Coordinate Analysis (PCA) was performed to visualise any structure signal in the microsatellite data for all sites using the R packages *adegenet* and *ade4* (*s.class* function), with a selection of graphical visualisation R packages *wordcloud* (v.2.6), *RColorBrewer* (v.1.1-2), *viridis* (v.0.5.1) and *viridisLite* (v.0.3.0). A Bartlett test of homogeneity of variances was conducted in order to test whether the mean observed heterozygosity was significantly lower than the mean heterozygosity (R package: *Stats* (v.4.0.3)). A subsequent paired t-test was incorporated to determine whether the Bartlett test was significant. If the paired t-test is lower than the alpha ($P < 0.05$), the observed heterozygosity is significantly lower than the mean expected heterozygosity determined from the chi-square distribution.

Structure software (v.2.3.4) was implemented using a Bayesian admixture model with no population assumed *a priori* to identify genetically homogenous groups within the genotyped data set (Pritchard *et al.*, 2000). The K values tested ranged between 1 to 10 with 10 simulations for each. Structure runs were set up with 100,000 burn-in iterations with 500,000 Markov chain Monte Carlo (MCMC) repetitions to ensure convergence was achieved. Within Structure, the Evanno method was used to detect the number of clusters using an ad hoc statistic (ΔK) based on the rate of change in the log probability of data between successive K values (Evanno *et al.*, 2005). To identify and visualise the number of genetic clusters within the different groups, the R package *pophelper* was used (Francis, 2017).

A hierarchical analysis of molecular variance (AMOVA) was conducted using the software Arlequin (v.3.5.2.2) to investigate genetic variation among populations (Excoffier *et al.*, 2005). The parameters were set for 10,000 permutations, with the gametic phase set to unknown and codominant data. The AMOVA was used to test the population structure for: (a) differentiation between all sites (n=10) and all years (2003-2020), (b) all sites and year grouping (2003-2007), (c) all sites and year grouping (2008-2014), (d) all sites and year grouping (2015-2020), (e) differentiation between West (Hereford, Preston, Starcross and Wellesbourne) and East (Brooms Barn, Kirton, Rothamsted, Writtle, Wye and York)

between all years (2003-2020) and year groupings (2003-2007, 2008-2014 and 2015-2020). York was excluded from both the hierarchical AMOVA and pairwise F_{ST} analysis as no samples were taken during 2003-2007 as the suction trap was installed in 2008. Population comparisons were also computed and pairwise F_{ST} values obtained (10,000 permutations and significance set to 0.05). Pairwise F_{ST} information was extracted from Arlequin (10,000 permutations) into RStudio and plotted using the R packages XML (v.3.99-0.5), corrplot (v.0.84), magrittr (v.2.0.1) and dplyr (v.1.0.2). A Mantel test was conducted in Arlequin using 10,000 permutations to identify whether there was a correlation between geographical distance and genetic distances (F_{ST}). Latitude and longitude were obtained from the RIS suction trap network database.

5.3 Results

Over 300 DNA extractions were conducted on *N. ribnisi* samples between 2003-2017 throughout 10 sites across England. Due to very low levels of DNA, over half of the samples failed to yield above 5ng/ μ l and therefore 146 samples were used in the subsequent population genetics analyses (Fig. 5.2). A pairwise Wilcox test highlighted that the main significance resides between the most recent years (2019 and 2020) and early years (2003 and 2004). DNA yield only significantly differed between the most two recent years (2020 and 2019) and the earliest years (2003 and 2004). Despite this data pooling samples from 10 sites removes potential site variation, the DNA yield for 2003 was considerably lower than other years for all samples. DNA yield noticeably varied between years but DNA yields above 20ng/ μ l were much 'rarer' after 2016. This suggests DNA degradation occurs at a higher rate for the initial five years in storage (see Appendix 5). After this 2-year period, DNA quantity levels plateaued and decreases in quantity were more subtle (data not shown). Individual DNA quantity varied widely across sites and within years and little discernible pattern could be observed.

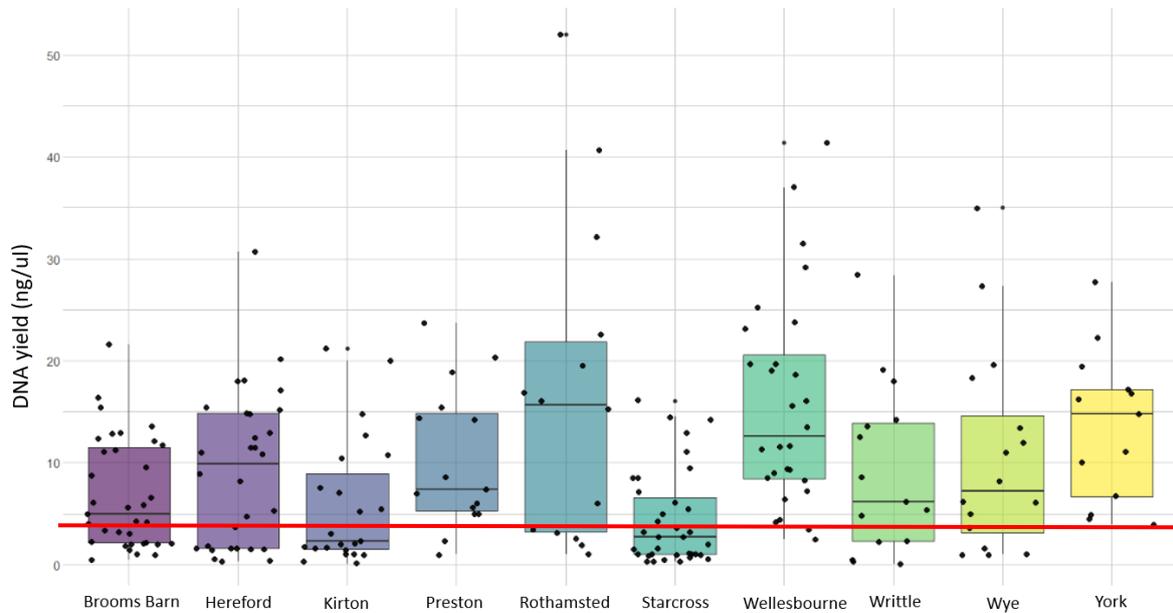


Figure 5.2: Jitter box plot of all DNA extractions using Qiagen QIamp DNA Micro Kit on individual *N. ribisnigri*. Samples below 4ng/ul (indicated by red line on graph) were unable to be used in the study as quantity was below minimum threshold. DNA yields were varied between sites and by year and was particularly low in Starcross, with DNA quantity dropping after the first five years of being stored in the RIS archive (Appendix 5).

5.3.1 The population structure of *Nasonovia ribisnigri*

The scatter plot of the principal coordinate analysis (PCoA) shows a separation of *N. ribisnigri* populations between 2003-2020, with three groupings of Hereford, Preston, Starcross and Wellesbourne; Brooms Barn, Kirton, Wye and York; Writtle and Rothamsted, suggesting that a difference between the populations exists (Fig. 5.3). Analysis of the entire dataset (n=146) in the Structure software using the Evanno method, indicated that the most likely number of genetic clusters is $K = 2$ (Evanno *et al.*, 2005). Unfortunately, this method does not estimate ΔK for $K = 1$ but since Mean $L(K)$ is not maximised for $K = 1$, the number of clusters is 2 (Morales-Hojas *et al.*, 2020a). Supporting the two genetic clusters, the individuals are distinctly distributed into 2 clusters in the bar plot when analysing the data with $K = 2$ (Fig. 5.4). These two population structures corresponded to an East/West divide between sites over the 17-year period.

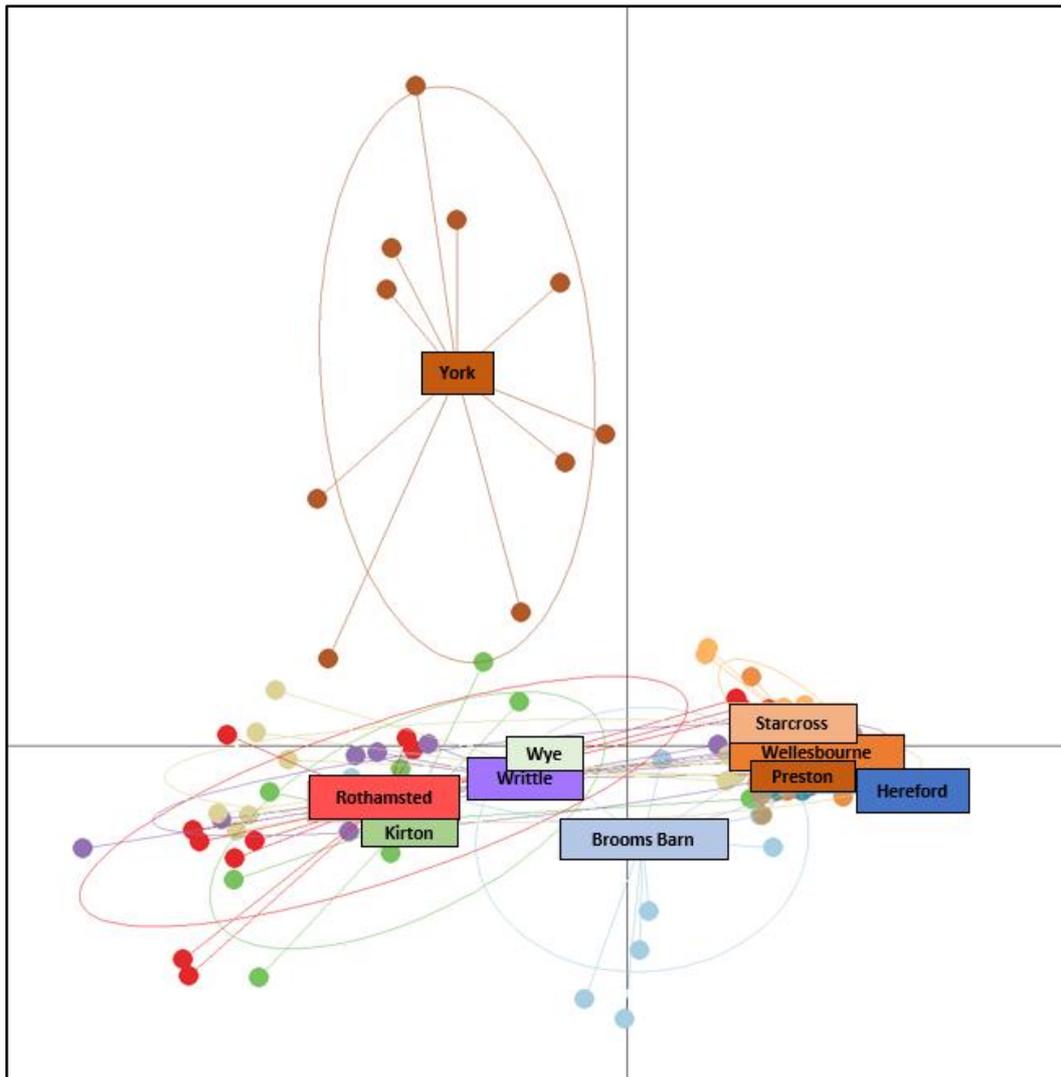


Figure 5.2: Principal coordinate analysis of the microsatellite variation in the currant-lettuce aphid (*Nasonovia ribisnigri*). Loose groupings of the West population (Hereford, Preston, Starcross and Wellesbourne) and the East (Brooms Barn, Kirton, Rothamsted, Writtle, Wye and York) can be seen.

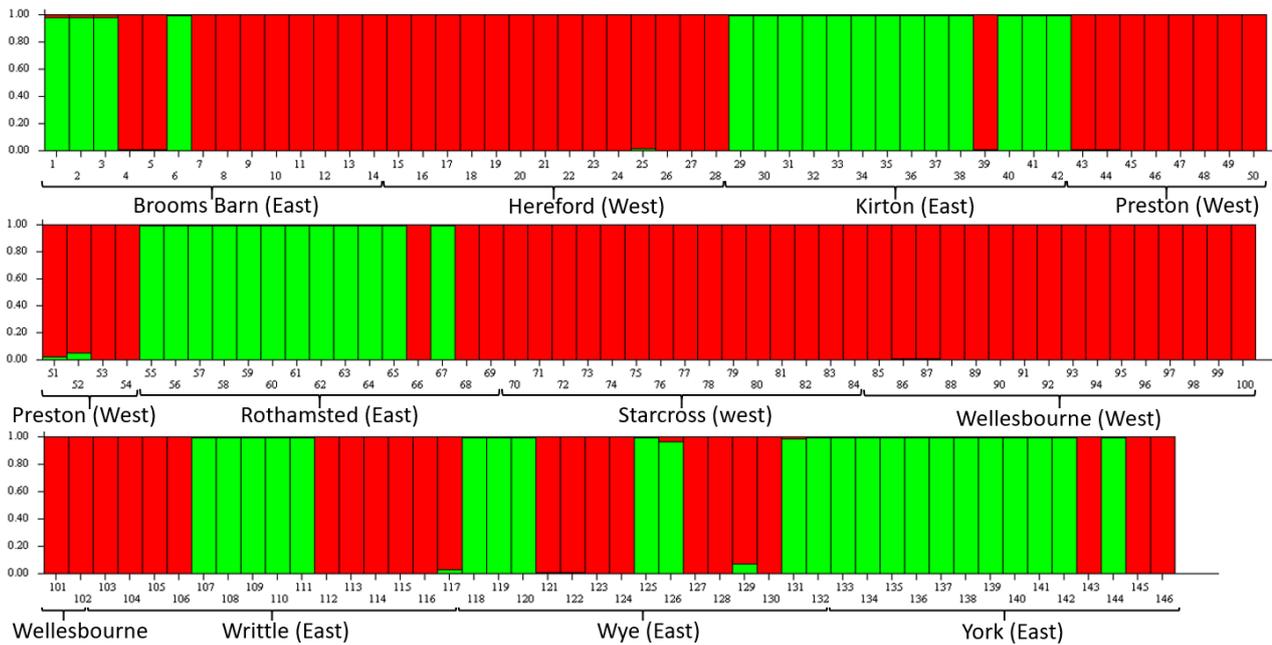


Figure 5.3: Genetic structure bar plot inferred using Bayesian analysis in Structure software with two clusters ($K = 2$). Each bar represents an individual, with the colour of the bar representing the likelihood of membership to either population cluster.

The population genetic differences between sites, estimated by pairwise F_{ST} for *N. ribisnigri* over the 17-year period ranged from moderate to high (Fig. 5.5). The highest genetic differentiation (F_{ST}) resided between populations between the East and West. Despite relatively close geographical location, Hereford and Wellesbourne (West) exhibited a high level of genetic differentiation to Rothamsted (East). Brooms Barn is exhibiting low F_{ST} between all other sites except Writtle over the 17-year period. The Mantel test identified a significant correlation between geographical distance and genetic distances for East and West, which suggests that the differentiation is partly a result of isolation by distance ($F_{ST} 0.307 P = <0.001$).

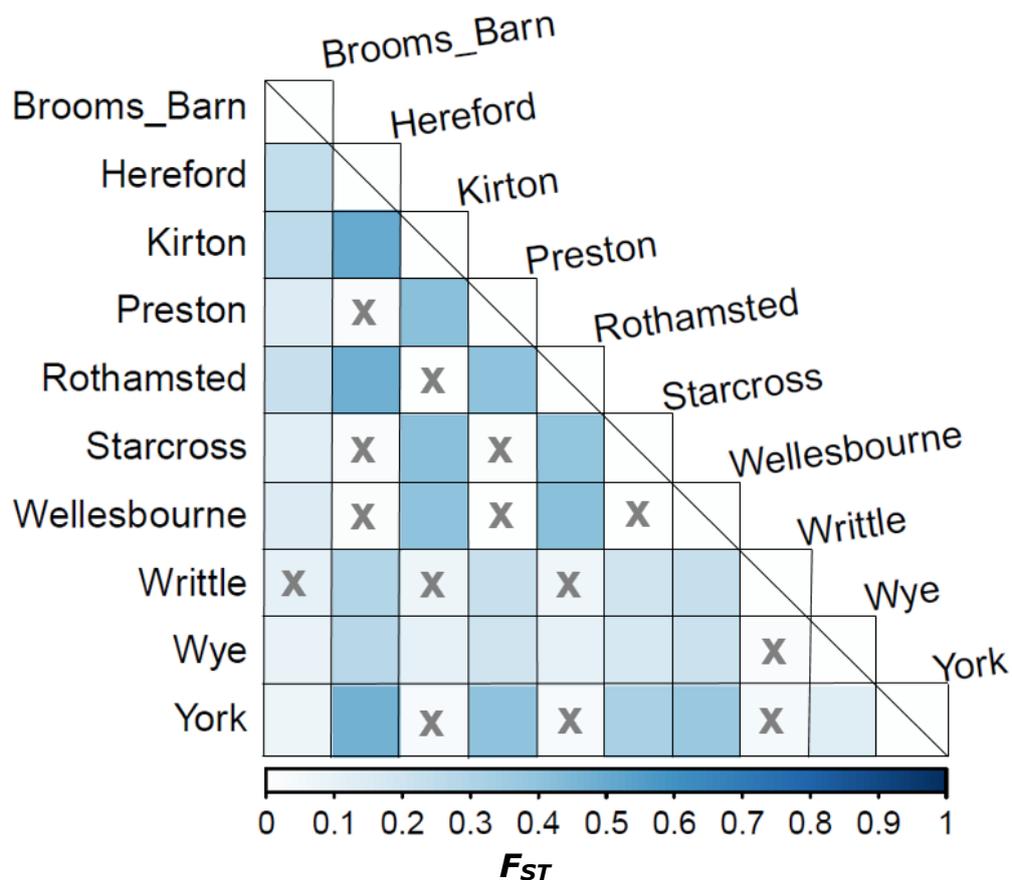


Figure 5.5: Population pairwise F_{ST} of *Nasonovia ribisnigri* between population sites between 2003-2020 showing the level of genetic differentiation. The scale bar on the x axis highlights the pairwise F_{ST} value, the darker the square, the higher the F_{ST} value. X indicates no significant F_{ST} value between two populations (significance set to $F_{ST} > 0.20$, $P < 0.05$).

Including all the samples in the analysis, a total of 292 haplotypes were observed (Table 5.6). The observed heterozygosity for all sites was lower than the expected heterozygosity ($H_O = 0.29$, $H_E = 0.55$), with an inbreeding coefficient (F_{IS}) of 0.64 ($P < 0.001$). 62.5% of the individual loci had significantly different observed heterozygosity (H_O) compared to the expected heterozygosity (H_E) for all sites ($P < 0.05$). The East population (Brooms Barn, Kirton, Rothamsted, Writtle, Wye and York) had 70.8% of individual loci that were significantly different from expected and observed heterozygosity whereas; the west population (Hereford, Preston, Starcross and Wellesbourne) had 50% (Table 5.6). Total H_O , averaged between individual microsatellite loci and sites, was lower in the East genetic cluster (0.27)

compared to the west (0.32). The H_o between both East and West populations overall was similar, however the H_o for York was much lower than all sites (0.09), which reduced the average H_o for the East region. For all sites, mean H_o was lower than H_e and the inbreeding coefficient (F_{IS}) was above 0 and significant. At the haplotype level, there were no deviations from the HWE in most of the populations except for three during 2015-2020 (Kirton, Rothamsted and York). At the individual loci level, a total of 8.5% (25) of the individual loci deviated from the HWE in all populations ($n = 292$), with the majority of the deviations (13.2%) occurring in the East ($n = 166$) and 2.4% (3) deviations in the west ($n = 126$). Eight out of 10 populations had at least one locus which deviated from the HWE, except Hereford and Preston which had none (Table 5.6). York had the lowest observed heterozygosity out of the sites (0.09) and subsequently the highest inbreeding coefficient (F_{IS} 0.94, $P = <0.001$) however, the H_e was similar to the other populations and suggests that an equal range of genetic diversity was sampled from each site.

Overall, the F_{IS} population specific inbreeding coefficients were higher in populations in the East compared to the west. Very high F_{IS} levels can be observed in York, Kirton and Writtle, which could be a result of most of the sample collection occurring June-August for these sites.

Table 5.6: Genetic diversity estimates for the 10 *N. ribisnigri* sites; combined sites; and East and West populations between 2003-2020. *N* – number of gene copies (number of individuals x2), *H* – number of haplotypes, *H_O* – mean observed heterozygosity for all eight microsatellite loci, *H_E* – mean expected heterozygosity for all eight microsatellite loci, *H_O/H_E* – percent of loci which had significantly different *H_O* and *H_E* over all eight microsatellite loci, *F_{IS}* – population specific inbreeding coefficient, and *pHWE* – number of loci for each population which departed from Hardy-Weinberg equilibrium (HWE).

Site	<i>N</i>	<i>H</i>	<i>H_O</i>	<i>H_E</i>	<i>H_O/H_E</i>	<i>F_{IS}</i>	<i>pHWE</i>
Brooms Barn	28	28	0.29	0.69	75	0.69	3
Hereford	28	28	0.3	0.43	50	0.40	0
Kirton	28	28	0.23	0.51	25	0.80	2
Preston	24	24	0.39	0.47	37.5	0.47	0
Rothamsted	30	30	0.3	0.59	75	0.63	3
Starcross	30	30	0.23	0.49	75	0.69	1
Wellesbourne	44	44	0.37	0.5	37.5	0.43	2
Writtle	28	28	0.33	0.63	75	0.78	5
Wye	28	28	0.36	0.62	75	0.63	4
York	24	24	0.09	0.63	100	0.94	5
All	292	292	0.29	0.55	62.5	0.64	25
East	166	166	0.27	0.61	70.8	0.74	22
West	126	126	0.32	0.47	50	0.49	3

5.3.2 Temporal analysis of *Nasonovia ribisnigri*

The hierarchical AMOVA for all sites excluding York (n=9) and among year the three year groupings (2003-2007; 2008-2014; 2015-2020) had a non-significant genetic variation of 4.58% ($F_{ST} -0.045$, $P = 0.98$), with a significant variation of 28.19% ($F_{SC} 0.269$, $P = <0.001$) among individuals within groups (Table 5.7 A). A similar amount of genetic variation (28.13%) resided among populations within the year groups ($F_{IS} 0.368$, $p = <0.001$) and the remaining 48.26% resided within individuals ($F_{IT} 0.517$, $P = <0.001$).

Table 5.7: Hierarchical analysis of molecular variance (AMOVA) for *Nasonovia ribisnigri*. (A) Genetic variation of all individuals between 9 sites among year groupings (2003-2007; 2008-2014; 2015-2020); (B) East geographical cluster determined by Structure comprising of Brooms Barn, Kirton, Rothamsted, Writtle and Wye among year groups ; (C) West geographical cluster determined by Structure comprising of Hereford, Preston, Starcross and Wellesbourne among year groups. York was excluded from temporal analysis as the suction trap was installed in 2008 and therefore no data was collected between 2003-2007.

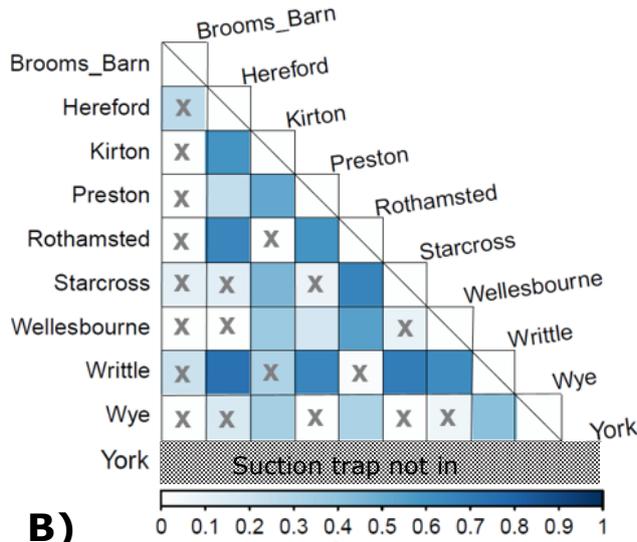
Source of variation	Sum of squares	Variance components	% variation	Fixation indices	P value
(A)					
Among year groups	-1.107	-0.07	-4.58	<i>FCT</i> -0.045	0.98
Among populations within year groups	154.5	0.434	28.19	<i>FSC</i> 0.269	<0.001
Among individuals within populations	188.33	0.433	28.13	<i>FIS</i> 0.368	<0.001
Within individuals	108.5	0.743	48.26	<i>FIT</i> 0.517	<0.001
(B)					
Among year groups	2.003	0.035	-3.06	<i>FCT</i> -0.030	0.732
Among populations within year groups	40.461	0.150	12.97	<i>FSC</i> 0.125	<0.001
Among individuals within populations	94.711	0.392	33.88	<i>FIS</i> 0.376	<0.001
Within individuals	54	0.65	56.21	<i>FIT</i> 0.437	<0.001
(C)					
Among year groups	0.362	0.064	-5.57	<i>FCT</i> 0.055	0.984
Among populations within year groups	23.865	0.142	12.29	<i>FSC</i> 0.116	0.012
Among individuals within populations	62.058	0.136	11.76	<i>FIS</i> 0.126	0.018
Within individuals	59.5	0.944	81.53	<i>FIT</i> 0.184	<0.001

There was no significant difference among groups from the East cluster between the three year groupings, with only 3.06% variation accounted for (F_{CT} -0.030, $P = 0.732$), highlighted by the hierarchical AMOVA (Table 5.7 B). The majority of genetic variation (56.21%) resided between individuals (F_{IT} 0.437, $P = <0.001$), with the remaining 33.88% among individuals within populations (F_{IS} 0.376, $P = <0.001$). These results were similar in the west, in which no significant genetic variation was apparent among different year groups (5.57%, F_{CT} 0.055, $P = 0.984$) and suggests that both the East and west clusters have not significantly changed over time (Table 5.7 C). Additionally, the pairwise F_{ST} values between years was small and non-significant, which further supports an absence of genetic variation during the 17-year period (Table 5.8). This result was mirrored when populations were divided into the two separate clusters (determined by Structure analysis) in the East and west.

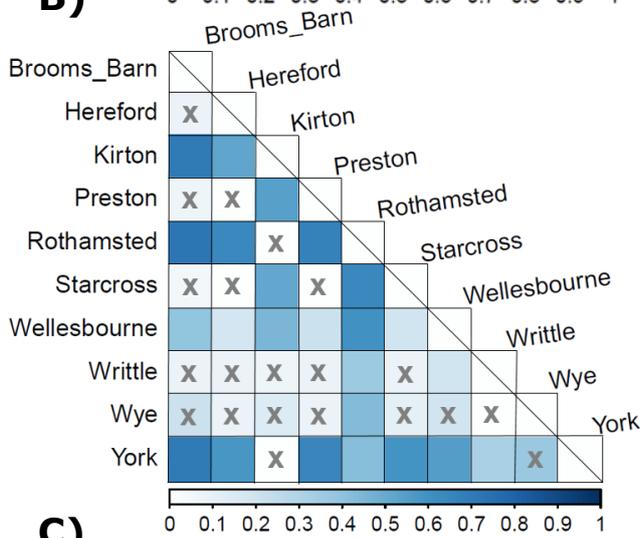
Table 5.8: Pairwise F_{ST} values of all individuals collected in different year groups (A), between individuals in the East (B) and individuals in the West (C). N shows the number of individuals in each group.

	N	2003-2007	2008-2014	2015-2020
(A) All				
2003-2007	48	-		
2008-2014	38	-0.025	-	
2015-2020	48	-0.038	0.016	-
(B) East				
2003-2007	25	-		
2008-2014	19	-0.009	-	
2015-2020	27	-0.05	0.022	-
(C) West				
2003-2007	23	-		
2008-2014	19	-0.04	-	
2015-2020	21	0.023	-0.063	-

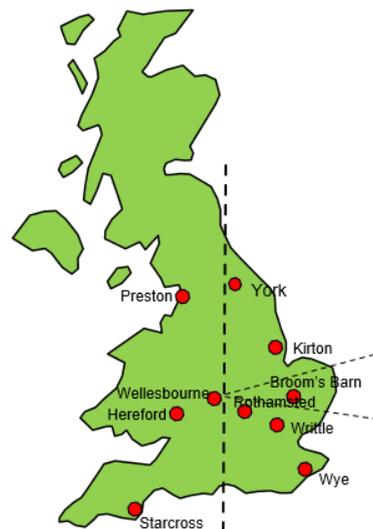
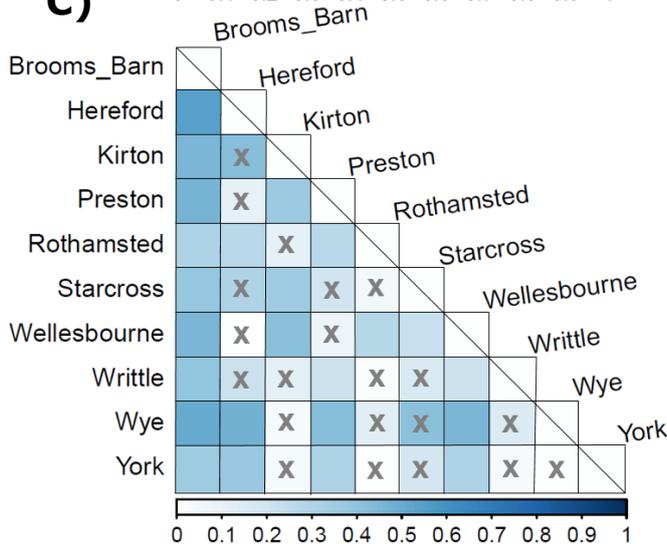
A)



B)



C)



F_{ST}

Figure 5.6: Population pairwise F_{ST} of *Nasonovia ribisnigri* between population sites and separated between year groups showing the level of genetic differentiation; (A) 2003-2007; (B) 2008-2014; (C) 2015-2020. The scale bar on the x axis highlights the pairwise F_{ST} value and X indicates no significant F_{ST} value. The UK map to the right of each pairwise F_{ST} analysis indicates population site locations with the thicker dashed line highlighting the East/West population divide (genetic barrier) and the thinner dashed line highlighting the main disparities between East year groupings and the general divides in the East population.

Overall, pairwise F_{ST} highlights no significant change in the inter population genetic differentiation at each of the three year intervals analysed (2003-2007, 2008-2014, 2015-2020) but some subtle disparities exist when examining individual population pairwise F_{ST} between these sample points (Fig. 5.6). In particular, the Wellesbourne population during 2008-2014 had low significant genetic differences with other populations in the west cluster and moderate-high significant genetic difference with Rothamsted, despite being relatively geographically close (Fig. 5.6 B). Brooms Barn during 2003-2007 had low genetic differentiation between all sites (Fig. 5.6 A), which is likely an artefact of some microsatellite loci being absent during sequencing and analysis and therefore it is difficult to draw conclusions. Between 2015-2020, Brooms Barn had significant pairwise differentiation (F_{ST}) between other populations and indicates that this site had moderate genetic differentiation between these sample points (Fig. 5.6 C). Due to exceptionally low numbers of *N. ribisnigri* caught in the RIS suction traps, sample size was low and therefore the changes in differentiation could be a function of sample size.

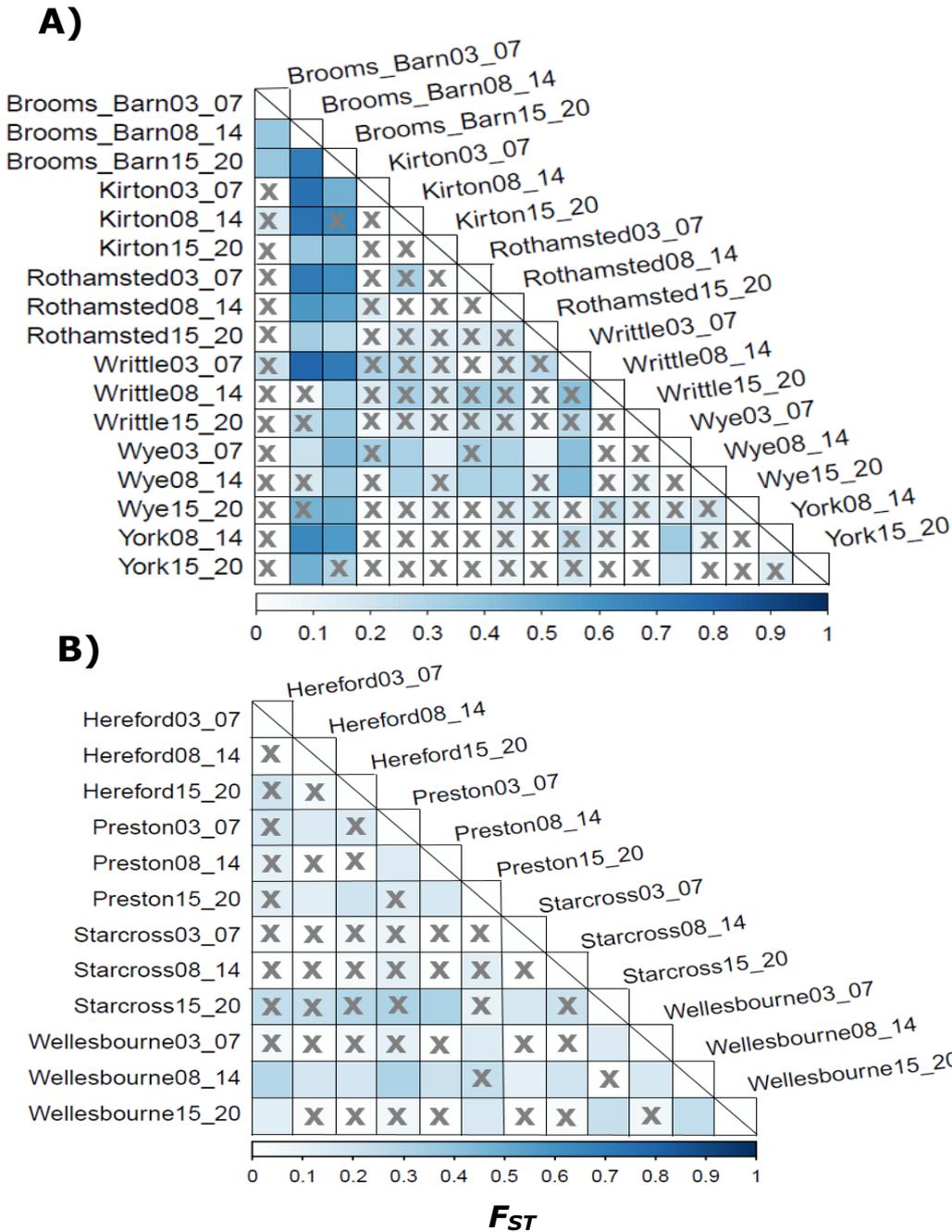


Figure 5.7: Population pairwise F_{ST} of *Nasonovia ribisnigri* between population sites between 2003-2020 showing the level of genetic differentiation between the East (A) and the West (B). 03_07 = 2003-2007; 08_14 = 2008-2014; 15_20 = 2015-2020. The scale bar on the x axis highlights the pairwise F_{ST} value, the darker the square, the higher the F_{ST} value. X indicates no significant F_{ST} value between two populations (significance set to $F_{ST} > 0.20$, $P = < 0.05$).

Overall, in the west, the genetic differentiation between locations has remained relatively stable over the 17 years in each of the three groupings (2003-2007, 2008-2014, 2015-2020), evidenced by the non-significant F_{ST} values throughout (Fig. 5.7 B). The main genetically different group is Wellesbourne during 2008-2014, in which significant F_{ST} values were detected between 0.15-0.26. The overall majority of pairwise differentiation in the East was low and non-significant (Fig. 5.7 A). Brooms Barn was an exception and indicated moderate to high levels of genetic differentiation between sites during 2008-2014 and 2015-2020.

Table 5.9: Hierarchical analysis of molecular variance (AMOVA) for *Nasonovia ribisnigri*. Genetic variation of all individuals between the East and West grouping, as identified in Structure. East geographical cluster comprises of Brooms Barn, Kirton, Rothamsted, Writtle and Wye among all year groups. West geographical cluster comprises of Hereford, Preston, Starcross and Wellesbourne among all year groups. d.f = degrees of freedom.

Year	Source of variation	d.f	Sum of squares	Variance components	% variation	Fixation indices	P value
2003-2020		1				<i>FCT</i>	
	Among groups		81.91	0.552	30.14	0.301	<0.001
	Among populations within groups	27	71.49	0.103	5.67	0.081	<0.001
	Among individuals within populations	117	188.33	0.433	23.64	0.368	<0.001
	Within individuals	145	108.5	0.743	40.55	0.594	<0.001

Separate pairwise F_{ST} analysis of the East and West clusters (as identified in Structure) highlights the genetic differentiation within each region over the 17-year period (Fig. 5.7). The pairwise F_{CT} of both the East and West grouping accounted for 30.14% of the genetic variation of these two clusters over the entire 17-year study period (F_{CT} 0.30, $P = <0.001$) (Table 5.9). The amount of variation among populations within the East and West grouping was low (5.67%) but significant (F_{SC} 0.081, $P = <0.001$). The remaining variation resided among individuals within populations (23.64%, F_{IS} 0.368, $P = <0.001$) and within individuals (40.55%, F_{IT} 0.594, $P = <0.001$).

5.4 Discussion

This study is the first to report the population genetics of *Nasonovia ribisnigri* using microsatellite markers developed from a recently assembled *N. ribisnigri* genome (Chapter 5). Two genetically distant clusters were identified, which correspond to an East and West divide. Throughout the 17-year sample period, genetic variation between the East and West population of *N. ribisnigri* was 30.14%. This result is higher than a recent paper on the population genetic structure of the bird-cherry oat aphid (*Rhopalosiphum padi*) which identified two clusters between populations in the North and the South of England (17%) (Morales-Hojas *et al.*, 2020a). It is unlikely that the two clusters in the present study correspond to sexual (holocyclic) and asexual (anholocyclic) morphs due to the positive and significant inbreeding coefficient (F_{IS}) observed in all populations within both genetic clusters. Given the general low number of individuals sampled and the low observed heterozygosity in all populations over the 17-year period, it is likely that inbreeding is occurring in English populations of *N. ribisnigri*.

The increasing warmer climate and general milder winter conditions in northern Europe, permit some aphid species to maintain an anholocyclic lifecycle and overwinter as an immature nymph or adult morph (Williams and Dixon, 2017). For example, anholocyclic peach-potato aphids, *Myzus persicae*, and English-grain aphids, *Sitobion avenae*, are spreading further north in line with favourable meteorological conditions (Blackman, 1974; Simon *et al.*, 1999; Ramiro-Morales *et al.*, 2020b). Despite anecdotal evidence that *N. ribisnigri* is undergoing anholocyclic reproduction in

the south, the results from this study suggest that the populations of *N. ribisnigri* in England are mostly holocyclic, including the southern-most populations sampled, and overwinter as an egg. This opposes the anholocyclic lifecycle hypothesis that are favoured under milder winter conditions, although it is possible that anholocyclic populations of *N. ribisnigri* do exist in the south, but they were not detected in either southerly location or, if present, are likely to be small isolated populations.

The Structure analysis highlighted that some individuals assigned to the west cluster were present in some of the sites in the East. Interestingly, it was shown that no individuals from the East populations were present in sites in the West. This suggests that some movement from populations in the west to the East does occur but it is not reciprocated, likely due to the dominance of the summer host, grown commercially in the East and is unlikely to be explained by limited migration potential. After all, the hop aphid, *Phorodon humuli* is found several hundreds of kilometres outside of its limited host range in Hereford and Kent, likely a result of conducive convective atmospheric boundary layer dynamics (Taylor *et al.* 1979). This may explain how *N. ribisnigri* migrants were transported a large distance from New Zealand to Tasmania (Stufkens *et al.*, 2004).

Many different barriers exist which can result in population isolation, limit genetic exchange and bottlenecks. A study on Russian wheat aphid (*Diuraphis noxia*) found that clear geographical isolation existed between populations in the north and south of China, with gene flow interrupted by desert regions (Zhang *et al.*, 2012). For *N. ribisnigri*, it is possible that the limited genetic exchange between the East and the West is a result of ecological processes, rather than physical barriers. This ecological differentiation between cultivated and uncultivated host plants could account for the divide in *N. ribisnigri* populations in the East and West of England due to much of the secondary host plant, lettuce (*L. sativa*), being cultivated predominantly in the East side of the UK. This result could also account for the west to East gene flow observed in some of the individuals during the 17-year period. Similar results have been found in other aphids such as the pea aphid (*Acyrtosiphum pisum*) and English-grain aphid (*Sitobion avenae*), which are both highly polyphagous species (Via and Hawthorne, 2002, Vialatte *et al.*, 2005). For example, populations of

English-grain aphid, *S. avenae*, collected on both cultivated and uncultivated host plants within the same area as each other highlighted that limited genetic exchange occurred between populations using microsatellite markers (Vialatte *et al.*, 2005).

It has been shown that *N. ribisnigri* are caught in much lower numbers than many other aphids important to horticulture. For example, in 2020 the total number of *M. persicae*, and *R. padi* caught in the Rothamsted Insect Survey (RIS) 12.2m suction traps across the UK came to: 13 088, and 21 184, respectively (RIS, 2021). The total number of *N. ribisnigri* caught in the same traps in 2020 totalled 71, which is significantly lower than all other pest aphids monitored by RIS. Further supporting information can also be found in Chapter 2, highlighting that different suction trap heights, alternative trap types and monitoring strategies also yielded similar results. Ward *et al.* (1998) suggests only 0.6% of the bird-cherry oat aphid (*Rhopalosiphum padi*) make it to their primary host. Furthermore, the authors propose that the main driver for host alternation is a result of a rendezvous for mating moreover escaping predators. These estimates were correlated using the numbers found in the RIS suction traps, so the percentage of *N. ribisnigri* successfully migrating back to their primary host would likely be much higher. Other literature suggests the major short-term advantage in sex in aphids is production of cold-resistant eggs (Simon *et al.*, 2002). The present study suggests population of *N. ribisnigri* are small and are likely to not migrate as far as many other aphid species. This would also account for the very low numbers caught in the suction traps in other in field traps (Chapter 2 and 3).

Supporting this conclusion, a volunteer network study in 2019 of allotment owners growing lettuce, which had over 150 participants throughout England and Wales, monitored their plots throughout the summer and sent all suspected aphids for identification at Rothamsted Research (Chapter 2). From the aphids identified, none were of *N. ribisnigri* but others such as the peach-potato aphid (*Myzus persicae*), potato aphid (*Macrosiphum euphorbiae*) and the glasshouse-potato aphid (*Aulacorthum solani*), despite *N. ribisnigri* being the most damaging species of aphid on outdoor lettuce crop, in an horticultural environment. *Nasonovia ribisnigri* is elusive and difficult to locate and trap despite their prevalence on outdoor lettuce crop in late summer. Furthermore, it is likely that *N. ribisnigri* has a small population size in England, evidenced by the

F_{is} which indicates that inbreeding is occurring, and would further explain the low numbers caught in the suction traps (Chapter 2).

In one endemic species of aphid in the Arctic Circle, *Acyrtosiphum svalbardicum* (Heikinheimo, 1968), it has been shown that populations exist in a metapopulation structure, in which local conditions are unfavourable for prolonged periods of time (e.g. unavailable suitable host plants), could lead to local extinctions and could take time for recolonisation to occur when conditions improve (Hanski, 1999). A similar event could be occurring in *N. ribisnigri*, in which few suitable host plants exist in certain areas of England and therefore limit movement between populations. The known host range of *N. ribisnigri* is members of the currant genus (*Ribes* spp.) during the winter and members of the Asteraceae family (which includes lettuce) in the spring and summer. Although there are numerous species from this family in the UK, it has been shown that *N. ribisnigri* cannot feed on all of them. A semi-field experiment on two *N. ribisnigri* cultures demonstrated that they accepted and fed on four out of five host plants from the Asteraceae family but with varying levels of success (Hough *et al.*, 2013). In particular, *N. ribisnigri* failed to establish on both field sow thistle *Sonchus arvensis* and smooth sow thistle *S. oleraceus*, with very low numbers for one culture on *S. arvensis*. Additionally, the field experiment in Chapter 3 highlighted that *N. ribisnigri* colonies were only found on chicory (*Chichorium intybus*) despite having access to other members of the Asteraceae family, including lettuce throughout the trial period.

Species of aphids such as *M. persicae*, which have over 600 known host plants are regarded as a cosmopolitan species and can sometimes referred to as generalists (Troncoso *et al.*, 2005; Truong *et al.*, 2015; Vargas *et al.*, 2005). This however has some caveats as not all *M. persicae* are able to feed on all known host plants and have clones which only accept and feed on a smaller selection of plants (Loxdale and Harvey, 2016). Therefore, the term 'generalist' is a misnomer and it is in fact different populations occupying a distinct ecological niche which would either be mono- oligo- or polyphagous but never generalist (Loxdale *et al.*, 2019). *Nasonovia ribisnigri*, with their comparatively small population size is likely to have unique ecological specialisms as a result of their low dispersal rate, host plant availability and are therefore more likely to have their own

niche and diet breadth. The long-term trends of *N. ribisnigri* in the UK highlight a significant decline in abundance since 1965 which are being partly attributed to the change in land use and loss of both winter host plants and alternative summer host plants (Chapter 2, section 3.1.1). It can be surmised that *N. ribisnigri* is in fact a 'rare' aphid and, with the common feature among rare species, is that they complete fewer generations in a year (Dixon and Kindlmann, 2020). Therefore, it is not only the abundance of the host plant that limits the abundance of aphids, but also their population growth rate (r_p), and subsequently their realised population growth rate (R). Furthermore, there is strong theoretical grounds, supported by empirical data, which suggest that host specific aphids which live on uncommon plants incur great losses in finding their host plants (Dixon and Kindlmann, 1990). As a result, they have a lower realised r (m) and are rarer than aphids living on common plants (Dixon and Kindlmann, 1990).

5.5 Conclusion

In conclusion, the use of microsatellite markers has enabled the identification of the population structure of previously unknown pest of outdoor lettuce in England. The analysis has revealed two distinct clusters within the East and West of England which is likely a result of a divide between cultivated outdoor lettuce in the East and uncultivated alternative hosts in the West. Temporal analysis over the 17-year period showed little genetic difference and therefore no evidence for long-term demographic change. This study highlights that *N. ribisnigri* predominantly undergo holocyclic reproduction, even in the south of England, and overwinter as an egg as opposed to an anholocyclic. Despite the subsequent breakdown of cultivar resistance (*NR-gene*) in 2007 and strong selection pressures due to insecticides, a demographic change was not observed, and indicates that selective pressures and environmental change have little effect on the genetic diversity. In addition, there is evidence for a deficiency of heterozygotes, which could be explained by inbreeding. This supports the hypothesis that the population size in England is small as indicated by the low number of individuals collected in the suction traps and on the lettuce crops.

These results have implication on future control and management strategies for this species however, further studies are required to fully understand their population dynamics.

Acknowledgements

I would like to thank the staff of Rothamsted Insect Survey for identifying *N. ribisnigri* caught in the suction traps during 2018-2020 as this helped speed up the analysis and sample condition. Additionally, I would like to thank the genomics group (particularly Steve Hanley and Rachel Rossiter) at Rothamsted Research for helping with the sequencing of the microsatellite loci used in the study.

5.6 References

- Allendorf, F.W., 2017. Genetics and the conservation of natural populations: allozymes to genomes.
- Barber, M.D., Moores, G.D., Tatchell, G.M., Vice, W.E. and Denholm, I., 1999. Insecticide resistance in the currant–lettuce aphid, *Nasonovia ribisnigri* (Hemiptera: Aphididae) in the UK. *Bulletin of Entomological Research*, pp.17-23.
- Blackman, R. L. 1974. Life-Cycle variation of *Myzus persicae* (Sulz) (Hom Aphididae) in different parts of the world, in relation to genotype and environment. *Bulletin of Entomological Research*, pp.595–607.
- Blackman, R.L. and Eastop, V.F., 2000. *Aphids on the world's crops: an identification and information guide* (No. Ed. 2). John Wiley & Sons Ltd.
- Bus, V.G.M., Chagné, D., Bassett, H.C.M., Bowatte, D., Calenge, F., Celton, J.M., Durel, C.E., Malone, M.T., Patocchi, A., Ranatunga, A.C. and Rikkerink, E.H.A., 2008. Genome mapping of three major resistance genes to woolly apple aphid (*Eriosoma lanigerum* Hausm.). *Tree Genetics & Genomes*, pp.223-236.
- Catchen, J.M., Hohenlohe, P.A., Bernatchez, L., Funk, W.C., Andrews, K.R. and Allendorf, F.W., 2017. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, pp.362-365.
- Dixon, A.F. and Kindlmann, P., 2020. Data on aphids indicate that rarity in herbivorous insects may be a consequence of a low rate of potential population growth. *Ecological Entomology*, pp.10-12.
- Dixon, A.F.G. and Kindlmann, P., 1990. Role of plant abundance in determining the abundance of herbivorous insects. *Oecologia*, pp.281-283.
- Dusfour, I., Vontas, J., David, J.P., Weetman, D., Fonseca, D.M., Corbel, V., Raghavendra, K., Coulibaly, M.B., Martins, A.J., Kasai, S. and Chandre, F., 2019. Management of insecticide resistance in the major Aedes vectors of arboviruses: Advances and challenges. *PLoS Neglected Tropical Diseases*, p.e0007615.
- Eenink, A.H., Groenwold, R. and Dieleman, F.L., 1982. Resistance of lettuce (*Lactuca*) to the leaf aphid *Nasonovia ribisnigri*. 1. Transfer of resistance from *L. virosa* to *L. sativa* by interspecific crosses and selection of resistant breeding lines. *Euphytica*, pp.291-299.
- Evanno, G., Regnaut, S. and Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, pp.2611-2620.
- Excoffier, L., Laval, G. and Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, p.117693430500100003.

- Faircloth, B.C., 2008. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular ecology resources*, pp.92-94.
- Francis, R. M. 2017. pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, pp.27–32.
- Franck, P., Reyes, M., Olivares, J. and Sauphanor, B., 2007. Genetic architecture in codling moth populations: comparison between microsatellite and insecticide resistance markers. *Molecular Ecology*, pp.3554-3564.
- Haine, E., 1955. Aphid take-off in controlled wind speeds. *Nature*, 175, pp.474-475.
- Halkett, F., Plantegenest, M., PRUNIER-LETERME, N., Mieuze, L., Delmotte, F. and Simon, J.C., 2005. Admixed sexual and facultatively asexual aphid lineages at mating sites. *Molecular Ecology*, pp.325-336.
- Hanski, I., 1999. *Metapopulation ecology*. Oxford University Press.
- Harrison, J.S. and Mondor, E.B., 2011. Evidence for an invasive aphid “superclone”: extremely low genetic diversity in oleander aphid (*Aphis nerii*) populations in the southern United States. *Plos One*, p.e17524.
- Heck, M. and Brault, V., 2018. Targeted disruption of aphid transmission: a vision for the management of crop diseases caused by Luteoviridae members. *Current Opinion in Virology*, pp.24-32.
- Hill, C.B., Chirumamilla, A. and Hartman, G.L., 2012. Resistance and virulence in the soybean *Aphis glycines* interaction. *Euphytica*, pp.635-646.
- Hough, G. L., 2013. *Biology and Control of Currant Lettuce Aphid Nasonovia ribisnigri*, University of Warwick.
- Jombart T. 2008. “adeget: a R package for the multivariate analysis of genetic markers.” *Bioinformatics*, pp.1403-1405.
- Jun, T.H., Michel, A.P. and Mian, M.R., 2011. Development of soybean aphid genomic SSR markers using next generation sequencing. *Genome*, pp.360-367.
- Keller, L.F., Jeffery, K.J., Arcese, P., Beaumont, M.A., Hochachka, W.M., Smith, J.N. and Bruford, M.W., 2001. Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, pp.1387-1394.
- López-Urbe, M.M., Soro, A. and Jha, S., 2017. Conservation genetics of bees: advances in the application of molecular tools to guide bee pollinator conservation. *Journal of Conservation Genetics*, pp.501-506

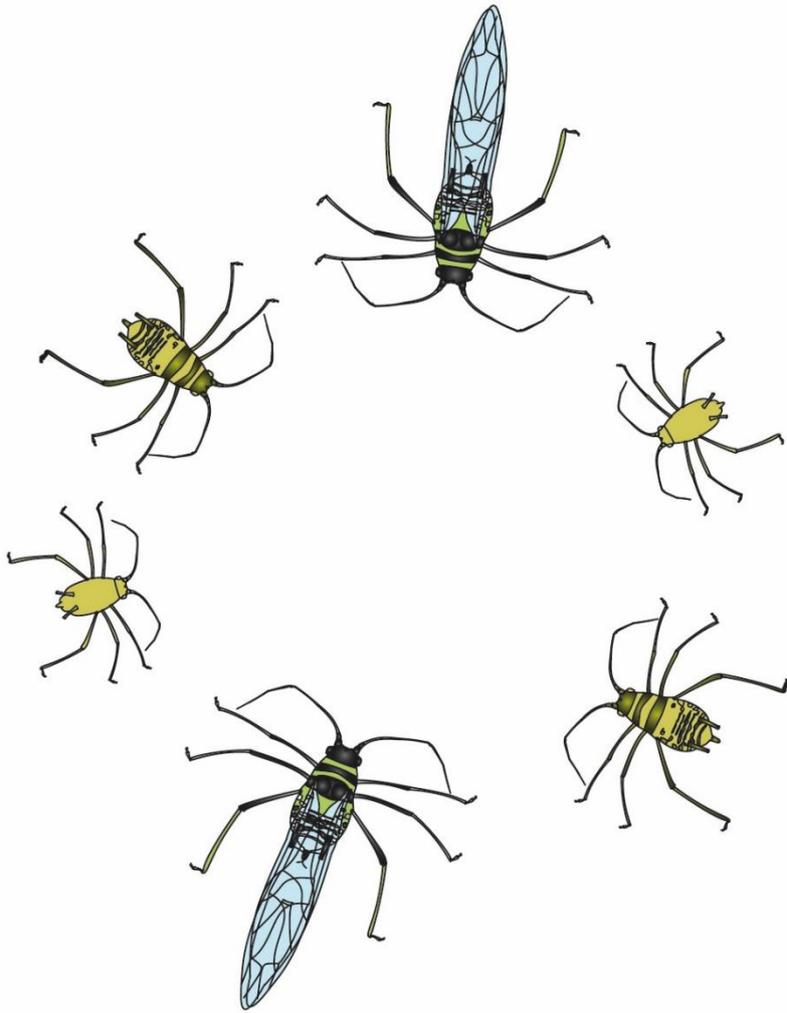
- Loxdale, H.D. and Harvey, J.A., 2016. The 'generalism' debate: misinterpreting the term in the empirical literature focusing on dietary breadth in insects. *Biological Journal of the Linnean Society*, pp.265-282.
- Loxdale, H.D., Balog, A. and Biron, D.G., 2020. Aphids in focus: unravelling their complex ecology and evolution using genetic and molecular approaches. *Biological Journal of the Linnean Society*, 129(3), pp.507-531.
- Loxdale, H.D., Balog, A. and Harvey, J.A., 2019. Generalism in nature... The Great Misnomer: Aphids and wasp parasitoids as examples. *Insects*, p.314.
- Loxdale, H.D., Hardie, J.I.M., Halbert, S.U.S.A.N., Footitt, R.O.B.E.R.T., Kidd, N.A. and Carter, C.I., 1993. The relative importance of short-and long-range movement of flying aphids. *Biological Reviews*, pp.291-311.
- Morales-Hojas, R., Gonzalez-Uriarte, A., Alvira Iraizoz, F., Jenkins, T., Alderson, L., Kruger, T., Hall, M.J., Greenslade, A., Shortall, C.R. and Bell, J.R., 2020a. Population genetic structure and predominance of cyclical parthenogenesis in the bird cherry-oat aphid *Rhopalosiphum padi* in England. *Evolutionary Applications*, pp.1009-1025.
- Morales-Hojas, R., Sun, J., Alvira Iraizoz, F., Tan, X. and Chen, J., 2020b. Contrasting population structure and demographic history of cereal aphids in different environmental and agricultural landscapes. *Ecology and Evolution*, pp.9647-9662.
- Natwick, E.T. and Lopez, M.I., 2016. Insecticide Efficacy Against Aphids in Leaf Lettuce, 2015. *Arthropod Management Tests*, pp.2023-2030
- Naveen, N.C., Chaubey, R., Kumar, D., Rebijith, K.B., Rajagopal, R., Subrahmanyam, B. and Subramanian, S., 2017. Insecticide resistance status in the whitefly, *Bemisia tabaci* genetic groups Asia-I, Asia-II-1 and Asia-II-7 on the Indian subcontinent. *Scientific Reports*, pp.1-15.
- Nebreda, M., Moreno, A., Pérez, N., Palacios, I., Seco-Fernández, V. and Fereres, A., 2004. Activity of aphids associated with lettuce and broccoli in Spain and their efficiency as vectors of Lettuce mosaic virus. *Virus Research*, pp.83-88.
- Nieto Nafria, J.M., 1974. Aphidinea de la Cordillera Central y provincia de Salamanca. INIA Monographs, p.168.
- Orantes, L.C., Zhang, W., Mian, M.A.R. and Michel, A.P., 2012. Maintaining genetic diversity and population panmixia through dispersal and not gene flow in a holocyclic heteroecious aphid species. *Heredity*, pp.127-134.
- Palumbo, J. C., and T. A. Hannan. 2002. Population growth of lettuce aphid, *Nasonovia ribisnigri*, on resistant butter and head lettuce cultivar. Vegetable Report. University of Arizona, College of Agriculture and Life Sciences, Tucson, AZ. ([http:// ag.arizona.edu/pubs/crops/az1292/az1292_2f.pdf](http://ag.arizona.edu/pubs/crops/az1292/az1292_2f.pdf)).

- Peccoud, J., Figueroa, C.C., Silva, A.X., Ramirez, C.C., Mieuze, L., Bonhomme, J., Stoeckel, S., Plantegenest, M. and Simon, J.C., 2008. Host range expansion of an introduced insect pest through multiple colonizations of specialized clones. *Molecular Ecology*, pp.4608-4618.
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. and Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS one*, p.e37135.
- Pritchard, J.K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, pp.945-959.
- Pu, J., Wang, Z. and Chung, H., 2020. Climate change and the genetics of insecticide resistance. *Pest management science*, pp.846-852.
- Reinink, K., Dieleman, F. L. 1993: Survey of aphid species on lettuce. IOBC/WPRS Bull., 56-68.
- Rinkevich, F.D., Hamm, R.L., Geden, C.J. and Scott, J.G., 2007. Dynamics of insecticide resistance alleles in house fly populations from New York and Florida. *Insect Biochemistry and Molecular Biology*, pp.550-558.
- RIS., 2021. The Insect Survey, Rothamsted Research. Available at: <https://www.rothamsted.ac.uk/insect-survey> (Accessed: 21/11/2021).
- Roitberg, B.D., Myers, J.H. and Frazer, B.D., 1979. The influence of predators on the movement of apterous pea aphids between plants. *Journal of Animal Ecology*, pp.111-122.
- RStudio Team (2020). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
- Sauer-Kesper CL, Lucia N, Buser H & Vogler U (2011) Bedeutung und Verbreitung des neuen Biotyps-Nr:1 der Grünen Salatlaus in der Deutschschweiz. *Agrarforschung Schweiz*, 462–469.
- Schoville, S.D., Chen, Y.H., Andersson, M.N., Benoit, J.B., Bhandari, A., Bowsher, J.H., Brevik, K., Cappelle, K., Chen, M.J.M., Childers, A.K. and Childers, C., 2018. A model species for agricultural pest genomics: the genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Scientific Reports*, pp.1-18.
- Simon, J.C. and Peccoud, J., 2018. Rapid evolution of aphid pests in agricultural environments. *Current Opinion in Insect Science*, pp.17-24.
- Simon, J.C., Baumann, S., Sunnucks, P., Hebert, P.D.N., Pierre, J.S., Le Gallic, J.F. and Dedryver, C.A., 1999. Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Molecular Ecology*, pp.531-545.
- Simon, J.C., Rispe, C. and Sunnucks, P., 2002. Ecology and evolution of sex in aphids. *Trends in Ecology & Evolution*, pp.34-39.

- Stufkens, M.A.W. and Bulman, S.R., 2002. *Nasonovia ribisnigri* a new aphid pest found on lettuces (*Lactuca sativa*) and *Ribes* spp in Canterbury. *New Zealand Plant Protection*, pp.437-437.
- Stufkens, M.A.W. and Teulon, D.A.J., 2003. Distribution host range and flight pattern of the lettuce aphid in New Zealand. *New Zealand Plant Protection*, pp.27-32.
- Stufkens, M.A.W., Nichol, S.E., Bulman, S.R. and Drayton, G.M., 2004. The incursion of the lettuce aphid into Tasmania could it have blown over from New Zealand. *New Zealand Plant Protection*, pp.340-340.
- Sunnucks, P., De Barro, P.J., Lushai, G., Maclean, N. and Hales, D., 1997. Genetic structure of an aphid studied using microsatellites: cyclic parthenogenesis, differentiated lineages and host specialization. *Molecular Ecology*, pp.1059-1073.
- Swale, D.R., Engers, D.W., Bollinger, S.R., Gross, A., Inocente, E.A., Days, E., Kanga, F., Johnson, R.M., Yang, L., Bloomquist, J.R. and Hopkins, C.R., 2016. An insecticide resistance-breaking mosquitocide targeting inward rectifier potassium channels in vectors of Zika virus and malaria. *Scientific reports*, pp.1-11.
- Taylor, L., Woiwod, I., & Taylor, R. 1979. The Migratory Ambit of the Hop Aphid and its Significance in Aphid Population Dynamics. *Journal of Animal Ecology*, pp.955-972.
- ten Broeke, C.J., Dicke, M. and van Loon, J.J., 2013. Performance and feeding behaviour of two biotypes of the black currant-lettuce aphid, *Nasonovia ribisnigri*, on resistant and susceptible *Lactuca sativa* near-isogenic lines. *Bulletin of Entomological Research*, pp.511-521.
- Thabuis, A.P.P., Teekens, K.C. and Van Herwijnen, Z.O., Rijk Zwaan Zaadteelt en Zaadhandel BV, 2014. *Lettuce that is resistant to the lettuce aphid Nasonovia ribisnigri biotype 1*. U.S. Patent 8,829,280.
- Troncoso, A.J., Vargas, R.R., Tapia, D.H., Olivares-Donoso, R. and Niemeyer, H.M., 2005. Host selection by the generalist aphid *Myzus persicae* (Hemiptera: Aphididae) and its subspecies specialized on tobacco, after being reared on the same host. *Bulletin of Entomological Research*, pp.23-28.
- Truong, D.H., Bauwens, J., Delaplace, P., Mazzucchelli, G., Lognay, G. and Francis, F., 2015. Proteomic analysis of *Arabidopsis thaliana* (L.) Heynh responses to a generalist sucking pest (*Myzus persicae* Sulzer). *Plant Biology*, pp.1210-1217.
- Vargas, R.R., Troncoso, A.J., Tapia, D.H., Olivares-Donoso, R. and Niemeyer, H.M., 2005. Behavioural differences during host selection between alate virginoparae of generalist and tobacco-specialist *Myzus persicae*. *Entomologia Experimentalis et Applicata*, pp.43-53.
- 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*, pp.76-88.

- Via, S., 1991. The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones. *Evolution*, pp.827-852.
- Vialatte, A., Dedryver, C.A., Simon, J.C., Galman, M. and Plantegenest, M., 2005. Limited genetic exchanges between populations of an insect pest living on uncultivated and related cultivated host plants. *Proceedings of the Royal Society B: Biological Sciences*, pp.1075-1082.
- Ward, S.A., Leather, S.R., Pickup, J. and Harrington, R., 1998. Mortality during dispersal and the cost of host-specificity in parasites: how many aphids find hosts?. *Journal of Animal Ecology*, pp.763-773.
- Williams, I., & Dixon, A. 2007. Life cycles and polymorphism. In H. F. van Emden, & R. Harrington (Eds.), *Aphids as crop pests*, 1st ed. Wallingford, Oxfordshire: CABI Publishing. pp. 58–59.
- Zhang, B., Edwards, O.R., Kang, L. and Fuller, S.J., 2012. Russian wheat aphids (*Diuraphis noxia*) in China: native range expansion or recent introduction?. *Molecular Ecology*, pp.2130-2144.

Chapter 6



Chapter 6: General Discussion

This project has been successful in elucidating the unusual ecology of a cryptic species of aphid using intensive fieldwork, statistical modelling and molecular methods. The overall aim of the project was to bridge knowledge gaps in the ecological processes leading to infestations in lettuce crops. The major findings of this project are **i)** A molecular method has been developed for aphids, to identify environmental DNA (eDNA) that has been acquired from a previous host plant and, from this, two new alternative hosts have been identified **ii)** A draft genome of the aphid has been assembled and annotated and the potential gene responsible for the resistance mechanism has been identified and, **iii)** Significant declines in the abundance of *N. ribisnigri* in the UK since 1965 have been identified, but there is an east/west divide in trends between populations.

Nasonovia ribisnigri is a problem on outdoor lettuce due to its behaviour to preferentially feed in the centre of lettuce plants, rendering many mitigative strategies ineffective and causing yield loss. Recently, aphid biotypes with resistance to the once resistant lettuce cultivars containing the Nr-gene have multiplied, further limiting control strategies. This project took a very wide view of the ecology and host-plant biology of *N. ribisnigri*, employing a cross section of resources and activities. Here, a summary of those activities and the results are drawn together to build the first detailed scientific assessment of *N. ribisnigri*.

6.1 New insights into the field behaviour of *Nasonovia ribisnigri*

6.1.1 Declines in *Nasonovia ribisnigri* populations in the UK

This project provides evidence that *N. ribisnigri* populations have declined significantly in the UK since 1965. During the 55-year period (1965-2020), the abundance of *N. ribisnigri*, as measured by captures in the RIS suction trap network, has declined by an overall -76% across the UK. Surprisingly, there is a mixed response across many other aphid species, which have either declined, remained relatively stable or slightly increased during this period (Bell *et al.*, 2020). The

most likely drivers of this severe decline in *N. ribisnigri* are postulated to be a result of changes in land use, horticultural intensification, climate change, increased insecticide usage and the introduction in the early 1980s of a lettuce cultivar resistant to *N. ribisnigri* (Nr-gene) (Chapter 2, section 5.1). In the eastern region where there are extensive areas of outdoor lettuce cultivation, populations declined the most (-85%). Since the majority (~75%) (Norman, personal communication) of the lettuce cultivars currently contain the Nr-gene, this is likely the largest contributing factor to the sharp decline in abundance from the mid-1980s. The abundance of *N. ribisnigri* in the north has declined slightly less (-72%), with significant declines observed between 1975-1985 and 2008-2012. Overall, the west experienced the lowest decline in abundance (-62%), with significant declines between 1972-1984, with the reasons behind the declines more likely connected to land use and host plant declines as opposed to the introduction of resistant outdoor lettuce cultivars, discussed in Chapter 2, section 5.1. Despite these significant declines in abundance, *N. ribisnigri* still remains the most economically damaging pest to outdoor lettuce in the UK.

Availability of overwintering host plants will limit aphid abundance and therefore the size of the population the following spring (Bell *et al.*, 2012). Since *N. ribisnigri* is predominantly associated with currant (*Ribes* spp.), winter host plant occupancy would be likely to be less than for other heterecious aphids that overwinter on grasses and herbs, for example. Despite evidence that *Ribes* spp. occupy approximately 73% of all the UK's 3,000 10km squares (Bell *et al.*, 2012), the preference of *N. ribisnigri* to feed primarily on new growth, further limits population growth. The increased demand for productive agricultural land and loss of semi-natural areas to urbanisation has likely reduced potential resources for *N. ribisnigri* and therefore further limiting population growth.

Evolution of *N. ribisnigri* biotypes have overcome the host plant resistance mechanism containing the Nr-gene (Thabuis *et al.*, 2013). Consequently, the abundance of the resistance-breaking biotypes would be expected to increase due to the positive bottom-up forces: the secondary host is now suitable for feeding and free of the systemic neonicotinoid treatment that would otherwise limit population growth (Chapter 2, section 3.1.2). The most severe observed decline in *N.*

ribisnigri abundance was in the East, which is suggested to be linked to control strategies in crops (Chapter 2, section 5.1) and likely to be a component of the overall cause for decline but not the main factor.

6.1.2 Migration and Flight Behaviour

As a direct result of this project, five 2 metre pop-up suction traps have now been incorporated into the G's Fresh (industry partner) pest surveillance programme in five counties (Cambridgeshire, Suffolk, Norfolk, Kent and Sussex), which will lead to better decision making based on incoming field data. This has provided valuable information to growers on aphid migration, phenology, and likelihood of infestation of crops. Pop-up suction trap at lower heights did not improve the capture rate of *N. ribisnigri* and the numbers captured in the lower traps were comparable to the 12.2 m metre Rothamsted Insect Survey (RIS) suction trap situated at Broom's Barn, Suffolk (Chapter 2, section 4.3). Additionally, no *N. ribisnigri* were captured in the different coloured water traps, on the yellow sticky traps nor in the Malaise traps.

What has been made evident in this project is that population density of *N. ribisnigri* is much lower than other horticulturally important aphids. It can be considered that, amongst pest aphids as a group, *N. ribisnigri* is a 'rare' aphid and exists in relatively small populations. Surprisingly, even when an outbreak occurs on outdoor lettuce, it would be expected that the exponential population growth would spill over and be recorded in traps situated in the proximity of the affected field. This however was not the case, and very few *N. ribisnigri* were captured during periods of outbreaks on outdoor lettuce (Chapter 2, section 5.2). This would suggest *N. ribisnigri* do not migrate far or that they migrate at a much greater height than the traps used in this study. Various observations and communications with growers and agronomists about alate *N. ribisnigri* on outdoor lettuce suggest they tend to walk when disturbed, despite possessing wings (Norman, personal communication). It has been shown that apterous and nymphal pea aphids, *Acyrtosiphon pisum*, were able to walk 13.5m away from a release point in 7 hours in semi-natural habitat (Ben-Ari *et al.*, 2015). It has also been shown that different clones of *M. persicae*

can walk at different rates and at different temperatures, with walking speeds declining with decreasing temperature (Alford *et al.*, 2012). An investigation of within-field walking behaviour, using the latest image recognition surveillance, would elucidate how *N. ribisnigri* disperses in outdoor lettuce which could improve future management practises.

When an outbreak of *N. ribisnigri* occurs in an outdoor lettuce crop, it is often isolated to one or two fields. Considering the two modes of alate migration (long distance and appetitive), the results of this project suggest that long-distance migration is very infrequent, including when aphids are feeding on outdoor lettuce. Most likely the alates produced during an outbreak are mainly displaying appetitive flight behaviour and moving within the flight boundary layer below 10 m and migrating less than 200 m (Parry *et al.*, 2013). This conclusion would explain the low numbers found in traps and the small, usually isolated, infestations on lettuce. However, further semi-field experiments would need to be conducted to test this hypothesis.

During the Citizen Scientists (CS) project in which over 130 volunteers grew lettuce in their allotments and gardens in the hope of attracting *N. ribisnigri* during the spring and summer months, none were found. The CS throughout England and Wales were reasonably distributed geographically (Chapter 2 section 4.4) and their locations included areas of both low and high agricultural intensification. Despite their diligence, only other known pest aphids and beneficials were identified during the study. The results of this CS study re-enforce the notion that *N. ribisnigri* is a rare aphid. If *N. ribisnigri* were present in the area of the lettuce plants, they were either existing at low population densities (as this project suggests); dispersing short distances (appetitive flight) and therefore not able to locate the lettuce; and/or preferentially selecting other alternative summer hosts (such as chicory, *Cichorium intybus*, or smooth hawksbeard, *Crepis capillaris*).

Comparisons with the phenology of other more abundant pest aphids has provided no strong evidence for their use as a proxy for the development of an improved forecasting model (Chapter 2 section 4.2). It was shown that by regressing accumulated day degree and first flight of *N. ribisnigri* a weak relationship was apparent. However, the low abundance of *N. ribisnigri*

introduced uncertainty about actual first flight dates into the model by producing false negatives. Between 1964-2019, the first flight events of *N. ribisnigri* generally appeared earlier in more recent warmer years, which conforms to previous literature that demonstrates that increases in accumulative day degrees advances the development and appearance of aphids earlier in the year (Harrington *et al.*, 2007).

6.1.3 Alternative host plants

This project is the first to successfully identify use of a particular host plant by individual aphids using a UniPlant method adapted from Moorhouse-Gann *et al.* (2018). With such a rare and cryptic aphid, molecular techniques are critical to help mitigate potential infestations on outdoor lettuce. This method was applied to a number of field-caught *N. ribisnigri* and identified two potential ‘green bridge’ host plants that might act as reservoirs. The main limitation of this study is the low population density and catch rate of *N. ribisnigri*, thus limiting the opportunity to sequence DNA. Although this method was developed for a horticulturally important pest aphid, this method can be applied to other small herbivorous insects to help elucidate their complex feeding behaviour (such as red listed invertebrates to aid conservation efforts).

It was shown that field margins sampled at both conventionally and organically managed outdoor lettuce farms did not harbour detectable numbers of *N. ribisnigri*, despite the presence of alternative host plants (Chapter 3 section 3.1). The work in Chapter 3 (section 3.2) showed that *N. ribisnigri* was only found on chicory (*Cichorium intybus*) in the monitoring plots containing a range of hosts, despite the availability of outdoor lettuce in the same plots throughout the experiment (April – October 2020). During this particular year, outbreaks of *N. ribisnigri* were very low in general, with only one or two small incidents, which resulted in negligible crop losses (Norman, personal communication). It could be surmised that there is the potential for chicory to be used as a trap crop to limit the likelihood of infestation onto an outdoor lettuce crop (Ben-Issa *et al.*, 2017), or potentially as monitoring plot (Chapter 3). These push-pull dynamics have shown to be effective methods to control aphids in crops (Xu *et al.*, 2018) and the use of trap crops has

been shown to be successful in several horticultural settings (Lithourgidis *et al.*, 2011; Lopes *et al.*, 2016; Parker *et al.*, 2013) and could help to displace *N. ribisnigri* away from outdoor lettuce, and attract and encourage natural enemies by providing both food and resources (Ben-Issa *et al.*, 2017). *Nasonovia ribisnigri* typically are not a problem until mid-late summer and therefore companion cropping, or trap cropping could be of benefit during this time. Further research would need to be conducted to establish whether chicory could be used as a trap crop or alternatively excluded from field margin seed mixes altogether and replaced with a suitable alternative that is not a host of *N. ribisnigri*.

6.2 Population processes and resistance mechanisms in *Nasonovia ribisnigri*

6.2.1 Population genetics

The analysis of *N. ribisnigri* populations in England using microsatellite markers showed an east-west genetic clustering (Chapter 5, section 3.1). This clustering demonstrated that a low level of gene flow is apparent in populations of *N. ribisnigri* in both the east and west of England. Furthermore, gene flow appeared unidirectional, in that aphids had the potential to move from west to east but not from east to west. This could be either a result of prevailing south westerly winds, or that strong ties to the lettuce host determine migratory behaviour. During the 17-year period (2003 to 2020), the long-term trends revealed relatively stable populations, despite the breakdown of host plant resistance and other environmental changes, such as increasing temperature. Additionally, *N. ribisnigri* has a low observed heterozygosity and high inbreeding coefficient (F_{IS}) indicating inbreeding occurs (Chapter 5, section 5.3.1). This observed low level of gene flow between populations of *N. ribisnigri* further supports the hypothesis that limited dispersal is an explanation for the unusually low catch rates in the RIS suction traps.

The results of the population genetics study support the differences observed in the population trend analysis, in which an evident difference in the decline in abundance of *N. ribisnigri* is apparent between the eastern and western regions. It has been surmised that this is the result of

the presence of different biotypes in the west which feed on alternative summer hosts but do not feed on lettuce (Chapter 5). However, at present, this suggestion is untested and further field work would need to be conducted (Vialatte *et al.*, 2005).

6.2.2 Genome and insights into resistance-breaking biotypes

This is the first study to have assembled and annotated a draft genome of *N. ribisnigri*. Before this project, only the mitochondrial cytochrome c oxidase subunit I (COI) barcode and a selection of odorant binding proteins existed. Furthermore, the new wild type genome sequence enabled the development of microsatellite markers used in the population genetics study (Chapter 4) which would have not been possible otherwise. Since assembly was only to a scaffold level, despite combining short-read Illumina sequences with long-read Oxford Nanopore data, incorporating additional long-read sequence data, such as from PacBio, would help to improve genome assembly (Biello *et al.*, 2021; Li *et al.*, 2019; Mathers *et al.*, 2021). The newly assembled genomes of both a susceptible and resistance-breaking biotype of *N. ribisnigri* will provide vital resources for future work, but will benefit from further improvement beyond their current 'draft' quality status.

In addition to the genome assembly and annotation, **the use of RNA-seq has elucidated a single gene potentially responsible for the resistance to the *Nr*-gene in lettuce.** This result is a great advancement for understanding the potential resistance mechanism in the resistant-breaking *N. ribisnigri* biotypes. Further understanding would benefit from using gene editing technologies such as zinc-finger nucleases (ZFNs) (Porteus and Baltimore, 2003), transcription activator-like effector nucleases (TALENs) (Hockemeyer *et al.*, 2011) and CRISPR-Cas9 (Cong *et al.*, 2013) to mediate knockout of the gene thought to be responsible for the resistance in the resistance-breaking biotype. By knocking out the gene, it would be possible to understand whether the resistance-breaking biotype is still able to feed on lettuce cultivars containing the *Nr*-gene. The results of this project bring exciting news with regard to the potential prevention of crop losses by an aphid which currently has very few effective methods of control.

6.3 Future trends and management of *Nasonovia ribisnigri*

A long-term trend in the decline of the abundance of *N. ribisnigri* in the agricultural landscape is evident and these declines are more notably severe in the eastern region, where outdoor lettuce is cultivated predominantly. The fragmentation of the agricultural landscape, and the removal of semi-natural areas (Kruess, and Tschardtke, 1994; Wagner et al, 2021) are likely to have impacted *N.ribisnigri* through the removal of scrub that contains the primary host (*Ribes* spp.). However, due to the habitus of *N. ribisnigri* and its ability to reproduce exponentially when able to colonise lettuce crops, with low initial detectability; *N. ribisnigri* still poses a major threat to outdoor lettuce.

Agricultural policy in the UK is moving towards agro-ecological approaches to ameliorate biodiversity decline and GHG emissions linked with intensive agriculture (Dimbleby, 2021; HMGGOV, 2018; 2021). Agro-ecological principles promote the integration of semi-natural areas in farmland; popular management options include field margins, conservation headlands, and regeneration of semi-natural areas within farmland (DEFRA 2020; NFFN, 2021). Indeed, such measures are already in place at the field sites sampled in Chapters 2 and 3. As this project has demonstrated that green bridges may be the primary pathway for crop colonisation, the expansion of areas supporting weedy vegetation in farmland creates considerable potential for *N. ribisnigri* to become an even larger problem. Added to this situation, it is likely that domestic production of outdoor salad crops will increase, particularly in the areas of organic production, to meet public demand, whilst moving towards net zero agriculture (Dimbleby, 2021; HMGGOV 2021). It is therefore important that the work within this project is continued, addressing unresolved issues, such as the identification of an unknown host used between the primary winter host and lettuce. In doing so, horticulturalists will be better able to move towards integrated pest management and control of *N. ribisnigri*.

6.3.1 Concluding remarks

Utilising a multidisciplinary approach, each component of this project provides evidence for the low abundance of *N. ribisnigri* in the UK, and insight into the potential gene responsible for the resistance-breaking biotype in outdoor lettuce. The methodology utilised in this project provides a crucial stepping stone towards further work to elucidate the cryptic behaviour and life cycle of this horticulturally important aphid. The work not only furthers knowledge about this species but is also relevant for broader research relating to dietary analysis in other cryptic herbivorous insects.

6.4 Key findings

- A molecular method to identify use of a particular host plant by individual aphids has been developed and two new alternative secondary hosts have been identified.
- A draft genome has been assembled and annotated and the potential gene responsible for the mechanism for overcoming host plant resistance has been identified.
- Significant declines since 1965 in the abundance of *N. ribisnigri* in the UK have been identified, but there is an East/West divide in trends between populations.
- Population density is low in the field and densities of migrating aphids are comparatively small, which is likely to be due to a weak propensity for flight.
- An intermediate, unidentified alternative secondary host plant is being used as a green bridge prior to *N. ribisnigri* colonising outdoor lettuce.

6.4.1 Recommendations for growers

- Remove chicory and crested dogs-tail from field margin mixes situated in proximity to outdoor lettuce.
- Increase pest surveillance and monitoring (2-metre pop-up suction traps), including monitoring lettuce plants.
- Conduct field trials to test potential trap or companion crops, and examine the utility of push-pull horticulture.
- Monitor the senescence of field margin host plants that may indicate migration away from these hosts into a crop, leading to infestation.

6.4.2 Recommendations for research

- Investigate within-field walking behaviour using the latest image recognition surveillance.
- Identify more potential secondary hosts using dietary analysis using the UniPlant method.
- Improve the draft genome assembly.
- Investigate genetic exchange between *N. ribisnigri* populations on uncultivated and cultivated host plants.
- Further verification of the potential host plant resistance-breaking gene and use gene editing technologies to mediate gene knockout.

6.5 References

- Alford, L., Hughes, G.E., Blackburn, T.M. and Bale, J.S., 2012. Walking speed adaptation ability of *Myzus persicae* to different temperature conditions. *Bulletin of entomological research*, pp.303-313.
- Bell, J.R., Blumgart, D. and Shortall, C.R., 2020. Are insects declining and at what rate? An analysis of standardised, systematic catches of aphid and moth abundances across Great Britain. *Insect conservation and diversity*, pp.115-126.
- Bell, J.R., Taylor, M.S., Shortall, C.R., Welham, S.J. and Harrington, R., 2012. The trait and host plant ecology of aphids and their distribution and abundance in the United Kingdom. *Global Ecology and Biogeography*, pp.405-415.
- Ben-Ari, M., Gish, M. and Inbar, M., 2015. Walking aphids can partake in within-field dispersal to distant plants. *Basic and Applied Ecology*, pp.162-171.
- Ben-Issa, R., Gomez, L. and Gautier, H., 2017. Companion plants for aphid pest management. *Insects*, p.112.
- Biello, R., Singh, A., Godfrey, C.J., Fernández, F.F., Mugford, S.T., Powell, G., Hogenhout, S.A. and Mathers, T.C., 2021. A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Molecular Ecology Resources*, pp.316-326.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. and Zhang, F., 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*, pp.819-823.
- DEFRA, 2020. Environmental Land Management Policy discussion document, February 2020. DEFRA.
- Dimbleby, H., 2021. The National Food Strategy. Independent review.
- Harrington, R., Clark, S.J., Welham, S.J., Verrier, P.J., Denholm, C.H., Hulle, M., Maurice, D., Rounsevell, M.D., Cocu, N. and European Union Examine Consortium, 2007. Environmental change and the phenology of European aphids. *Global change biology*, pp.1550-1564.
- HMG0V, 2018. A Green Future: Our 25 Year Plan to Improve the Environment. DEFRA
- HMG0V, 2021. Net Zero Strategy: Build Back Greener. Presented to Parliament pursuant to Section 14 of the Climate Change Act 2008.
- Hockemeyer, D., Wang, H., Kiani, S., Lai, C.S., Gao, Q., Cassady, J.P., Cost, G.J., Zhang, L., Santiago, Y., Miller, J.C. and Zeitler, B., 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology*, pp.731-734.

- Kruess, A. and Tscharntke, T., 1994. Habitat fragmentation, species loss, and biological control. *Science*, pp.1581-1584.
- Li, Y., Park, H., Smith, T.E. and Moran, N.A., 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Molecular biology and evolution*, pp.2143-2156.
- Lithourgidis, A.S., Dordas, C.A., Damalas, C.A. and Vlachostergios, D., 2011. Annual intercrops: an alternative pathway for sustainable agriculture. *Australian journal of crop science*, pp.396-410.
- Lopes, T., Hatt, S., Xu, Q., Chen, J., Liu, Y. and Francis, F., 2016. Wheat (*Triticum aestivum* L.)-based intercropping systems for biological pest control. *Pest management science*, pp.2193-2202.
- Mathers, T.C., Wouters, R.H., Mugford, S.T., Swarbreck, D., Van Oosterhout, C. and Hogenhout, S.A., 2021. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Molecular biology and evolution*, pp.856-875.
- Moorhouse-Gann, R.J., Dunn, J.C., De Vere, N., Goder, M., Cole, N., Hipperson, H. and Symondson, W.O., 2018. New universal ITS2 primers for high-resolution herbivory analyses using DNA metabarcoding in both tropical and temperate zones. *Scientific Reports*, pp.1-15.
- NFFN, 2021. Rethink farming: a practical guide for farming, nature & climate. Nature Friendly Farming Network.
- Parker, J.E., Snyder, W.E., Hamilton, G.C. and Rodriguez-Saona, C., 2013. Companion planting and insect pest control. *Weed and Pest Control-Conventional and New Challenges*, p.55044.
- Parry, H.R., 2013. Cereal aphid movement: general principles and simulation modelling. *Movement Ecology*, pp.1-15.
- Porteus, M.H. and Baltimore, D., 2003. Chimeric nucleases stimulate gene targeting in human cells. *Science*, p.763.
- Thabuis, A.P.P., Teekens, K.C. and Van Herwijnen, Z.O., Rijk Zwaan Zaadteelt en Zaadhandel BV., 2014. *Lettuce that is resistant to the lettuce aphid Nasonovia ribisnigri biotype 1*. U.S. Patent 8,829,280.
- Vialatte, A., Dedryver, C.A., Simon, J.C., Galman, M. and Plantegenest, M., 2005. Limited genetic exchanges between populations of an insect pest living on uncultivated and related cultivated host plants. *Proceedings of the Royal Society B: Biological Sciences*, pp.1075-1082.
- Wagner, D.L., Grames, E.M., Forister, M.L., Berenbaum, M.R. and Stopak, D., 2021. Insect decline in the Anthropocene: Death by a thousand cuts. *Proceedings of the National Academy of Sciences*.

Xu, Q., Hatt, S., Lopes, T., Zhang, Y., Bodson, B., Chen, J. and Francis, F., 2018. A push–pull strategy to control aphids combines intercropping with semiochemical releases. *Journal of pest science*, pp.93-103.

Appendix 1: Chapter 2 Supplementary material

Table 1: The list of all horticulturally important aphids currently reported in the RIS pest bulletin (2021).

RIS Bulletin aphid species	
Common name	Latin
Pea aphid	<i>Acyrtosiphon pisum</i>
Black bean aphid	<i>Aphid fabae</i>
Potato glasshouse aphid	<i>Aulacorthum solani</i>
Leaf-curling aphid	<i>Brachycaudus helichrysi</i>
Brassica aphid	<i>Brevicorne brassicae</i>
Willow-carrot aphid	<i>Caveriella aegopodii</i>
Sycamore aphid	<i>Drepanosiphum platanoidis</i>
Green spruce aphid	<i>Elatobium abietinum</i>
Mealy plum aphid	<i>Hyalopterus pruni</i>
Blackcurrant-sowthistle aphid	<i>Hyperomyzus lactucae</i>
Potato aphid	<i>Macrosiphum euphorbiae</i>
Rose-grain aphid	<i>Metopolophium dirhodum</i>
Shallot aphid	<i>Myzus ascalonicus</i>
Peach-potato aphid	<i>Myzus persicae</i>
Currant-lettuce aphid	<i>Nasonovia ribisnigri</i>
Hop aphid	<i>Phorodon humuli</i>
Corn leaf aphid	<i>Rhopalosiphum maidis</i>
Apple-grass aphid	<i>Rhopalosiphum oxycanthae</i>
Bird cherry-oat aphid	<i>Rhopalosiphum padi</i>
English grain aphid	<i>Sitobion avenae</i>
Blackberry-grass aphid	<i>Sitobion fragariae</i>

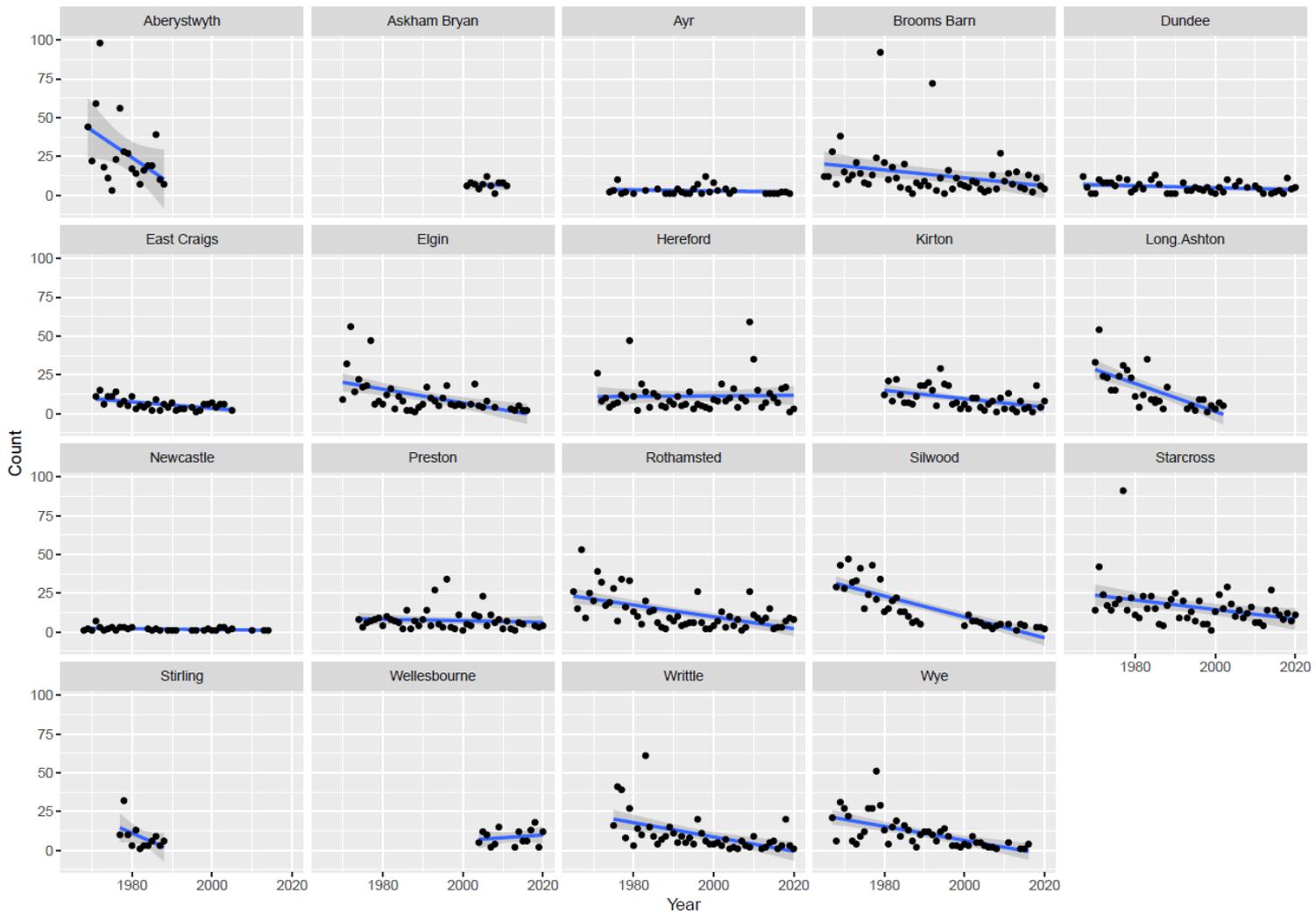


Figure 1: Linear year effect of abundance of *Nasonovia ribisnigri* between 1965-2020 plotted for all sites used in the long-term abundance trend analysis. Blue line indicates trend per site and 95% confidence interval in grey. Each black point represents actual abundance of *N. ribisnigri* per year for each site.

Table 2: Log-linear and non-linear year effects of *Nasonovia ribisnigri* abundance declines in the UK with 95% confidence intervals.

	Linear year effect	95% confidence interval	Non-linear year effect	95% confidence interval
All	-76%	-83%, -66%	-76%	-84%, -65%
North	-72%	-87%, -44%	-65%	-77%, -44%
South	-78%	-86%, -64%	-79%	-86%, -67%
East	-83%	-90%, -73%	-85%	-90%, -77%
West	-62%	-80%, -33%	-54%	-73%, -26%

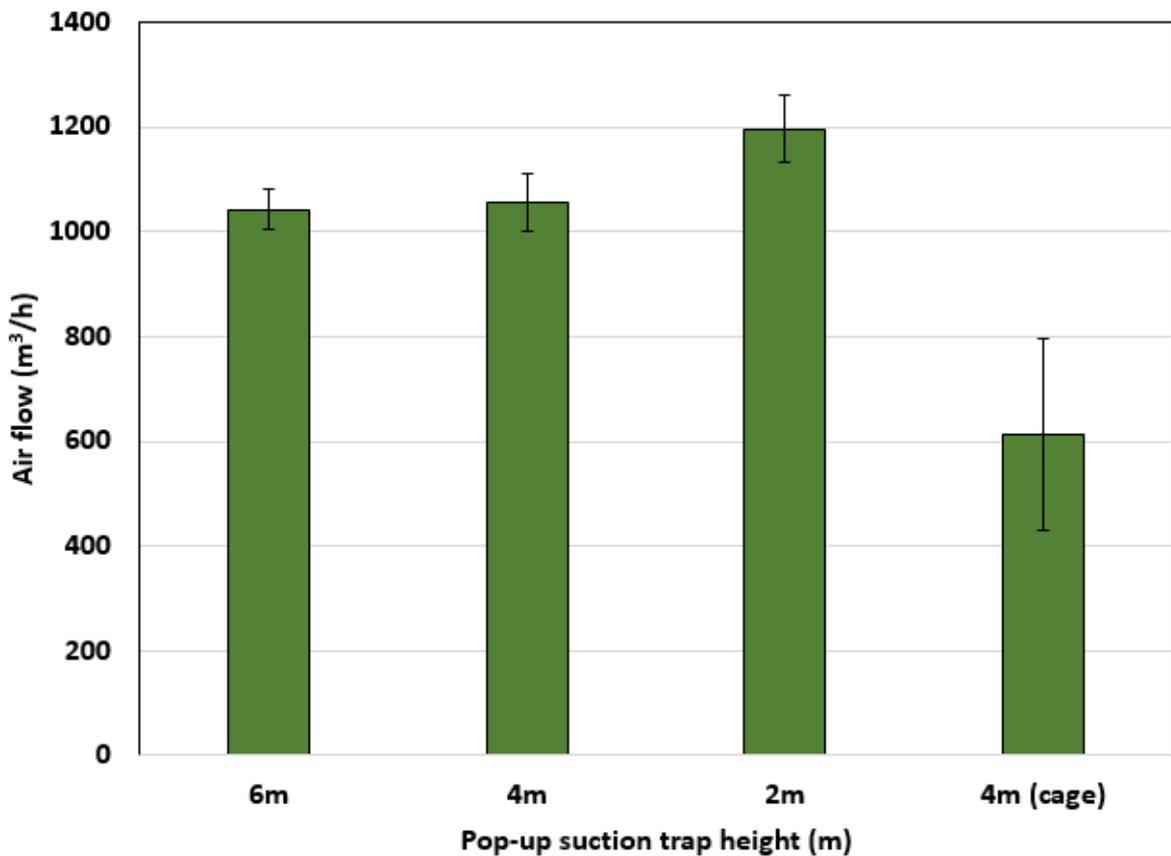


Figure 2: Pop-up suction trap air flow measured using 405i Smart Probe (Testo). To reduce air flow fluctuations, measurements were taken using a specifically cut hole in the pop-up traps 1m above the fan. 20 air flow readings were taken during a 4-hour period during the same day for each pop-up trap to produce an accurate average air flow. The 2m pop-up suction had a higher air flow compared to both the 4m and 6m, which is likely due to the additional PVC tubing. The addition of a 'cage' on top of the 4m meter almost halved the measured air flow and created higher fluctuations in the air flow measurements.

Keen lettuce growers wanted! Summer 2019

- Request from Dion Garrett at Rothamsted Research

I am a PhD student at Rothamsted Research and studying the population structure of the currant-lettuce aphid (*Nasonovia ribisnigri*), the most damaging species of aphid to lettuce crop. Once they get into crop they can stunt growth and, in large numbers, can lead to rejection by supermarkets and consumers.



I would like to request any keen growers if they would be willing to grow a patch of lettuce (approximately 20 plants) between **June-September**. If you see any aphids on the lettuce (especially in the centre), I would ask if you could send them to me at Rothamsted Research.

I will provide the lettuce seed, coir pellets (peat-free), sample collection pots, envelopes and stamps! I plan to send out the lettuce packs in May.

The variety is a sweet open-heart romaine, perfect for salads and ideal for growing outdoors. I hope to provide enough lettuce seed for you to eat too!

If you are interested, please send an email to: dion.garrett@rothamsted.ac.uk and I will get back to you with further details.

Figure 1: Winged currant-lettuce aphid on lettuce. These particular aphids prefer the younger leaves in the centre of the lettuce which makes them very difficult to control.

I hope to understand the evolution and spread of resistance in these aphids by looking at the genetic differences in the population. By collecting live specimens of the lettuce-currant aphid (**Fig. 1**), I will be able to measure the level of genetic variation between populations over the UK. This in turn will provide valuable information to growers on future management strategies for their farms, give insight into how other species of aphids may build-up resistance in arable areas and help reduce the loss of crop (**Fig. 2**).



Figure 2: One of the lettuce fields in Cambridgeshire which had high levels of currant-lettuce aphid present in the heart of the lettuce and was unsuitable for the supermarket!

Your help would be greatly appreciated.



@D_Garrett_Ento



Thank you very much for expressing your interest in helping me out with my project.

What is included in the pack:

- 1 x sample pot
- Lettuce seeds
- 10 x Compost plugs
- 1 x envelope
- 1 x first class large stamp

Plant the seeds in May for them to be in the garden/allotment from June to September. You can plant as many seeds as you like but bear in mind not all will germinate – so more is always better!

Checking lettuce for aphids:

Look in the centre of the lettuce (this is where the species I am after generally reside). If aphids are in the centre of the lettuce, this is quite likely that these are the aphids I'm after! This is the only species in the UK which prefer to feed in the middle of the lettuce. If you are unsure please send me a picture to my email address (dion.garrett@rothamsted.ac.uk) or, if you have a Twitter account, post on Twitter using my tag (@D_Garrett_Ento) and I will get back to you ASAP.

I would rather have a sample than not, even if it turns out to be the wrong species 😊

Some identification features/tips:

Winged aphids have dark long legs and dark patches on their body with long antennae (Figure 1).



Figure 1. The currant-lettuce aphid on iceberg lettuce. **Note:** the dark markings on legs and body.

Wingless aphids vary in colour from pink to green but again are most likely to be in the middle of the lettuce whereas other aphids that feed on lettuce are generally found on the outer leaves. They have some dark markings on their body but they are not as visible as the winged ones (Figure 2).



Figure 2: Wingless curren-lettuce aphids on lettuce. **Note:** Pink-red colour with markings less clear than the winged aphids. Colour can vary from pink/red to green.

Taking a sample:

- If aphids found on lettuce, please cut off (or tear) a small piece with some of the aphids on and put it in the provided sample pot.
Note: Make sure not to 'stuff' a big piece of lettuce in as this will likely squish all of the aphids and I need ones which are still alive!
- Write your name on the sample pot or put a small note in the envelope, so I know who it is from.
- Place sample pot in envelope and post it to:

Dion Garrett
BCP
Rothamsted Research
West Common
Harpenden
AL5 2JQ

R code - Trend analysis (Experiment 1)

```
install.packages("poptrend")
install.packages("mgcv")
library(poptrend)
library(mgcv)
library(ggplot2)

#Read in data

#All sites
# "fSITE" = Site number as factor
# "Year" = Year as continuous variable
# "Annual_count" = Total number of moths caught that site-year
# "fYear" = Year as a factor (for random effect)
# "Site_name_db" = Site site as it appears in database (not necessary for running
model )
# The data is structured so that site, year and abundance are all columns.
# Make the factors into factors

NRIB_ANNUAL_03DEC21$fYear <- paste0("Y_",NRIB_ANNUAL_03DEC21$Year)
NRIB_ANNUAL_03DEC21$Trap <- as.factor(NRIB_ANNUAL_03DEC21$Trap)
NRIB_ANNUAL_03DEC21$fYear <- as.factor(NRIB_ANNUAL_03DEC21$fYear)

# Models can be linear or non-linear. Poptrend models use the mgcv syntax

# Run a non-linear first
# Non linear (wiggly) #####
# Run a GAMM with a non-linear year effect
Wiggly_poptrend <- ptrend(Count ~ trend(Year,

                                tempRE = TRUE, # This adds a random effect for
year (tempRE = "Temporal Random Effect")

                                type = "smooth", # This makes it smooth (wiggly)
rather than linear

                                k = 18) + # This is the wiggleness of the curve. A
rule of thumb is to set as one third of

                                # time series length - ie., 49 years divided by 3 = roughly 16.
Can be reduced for a straighter curve or

                                # increased for a more wiggly one

s(Year, Trap, bs = "re"), # This adds a random slope within each year
```

```

        family = nb, # Negative binomial error distribution is usually good for
count data

        data = NRIB_ANNUAL_03DEC21) # Takes a little while...

change(Wiggly_poptrend, 1965, 2020) # To get the percentage change over the
specified years plus 95% CIs.
# NOTE: The non-linear models are not appropriate for giving a good estimate of
percentage change as they
# are too dependent on the first and last year - use log-linear model for that instead
(see below)

# Plot the model
plot(Wiggly_poptrend)

# Now do a linear version
# # # #
# Linear (log-linear) ####
# Run a GAMM with a linear year effect
Linear_poptrend <- ptrend(Count ~ trend(Year,
                                     tempRE = TRUE, # This adds a random effect for
year (tempRE = "Temporal Random Effect")
                                     type = "loglinear") +
s(Year, Trap, bs = "re"), # This adds a random slope within each
year
                                     family = nb, # Negative binomial error distribution is usually good
for count data
                                     data = NRIB_ANNUAL_03DEC21)

change(Linear_poptrend, 1965, 2020) # To get the percentage change over the
specified years. NOTE: The non-linear models are
# not appropriate for giving a good estimate of percentage change as they are too
dependent on the first and last year

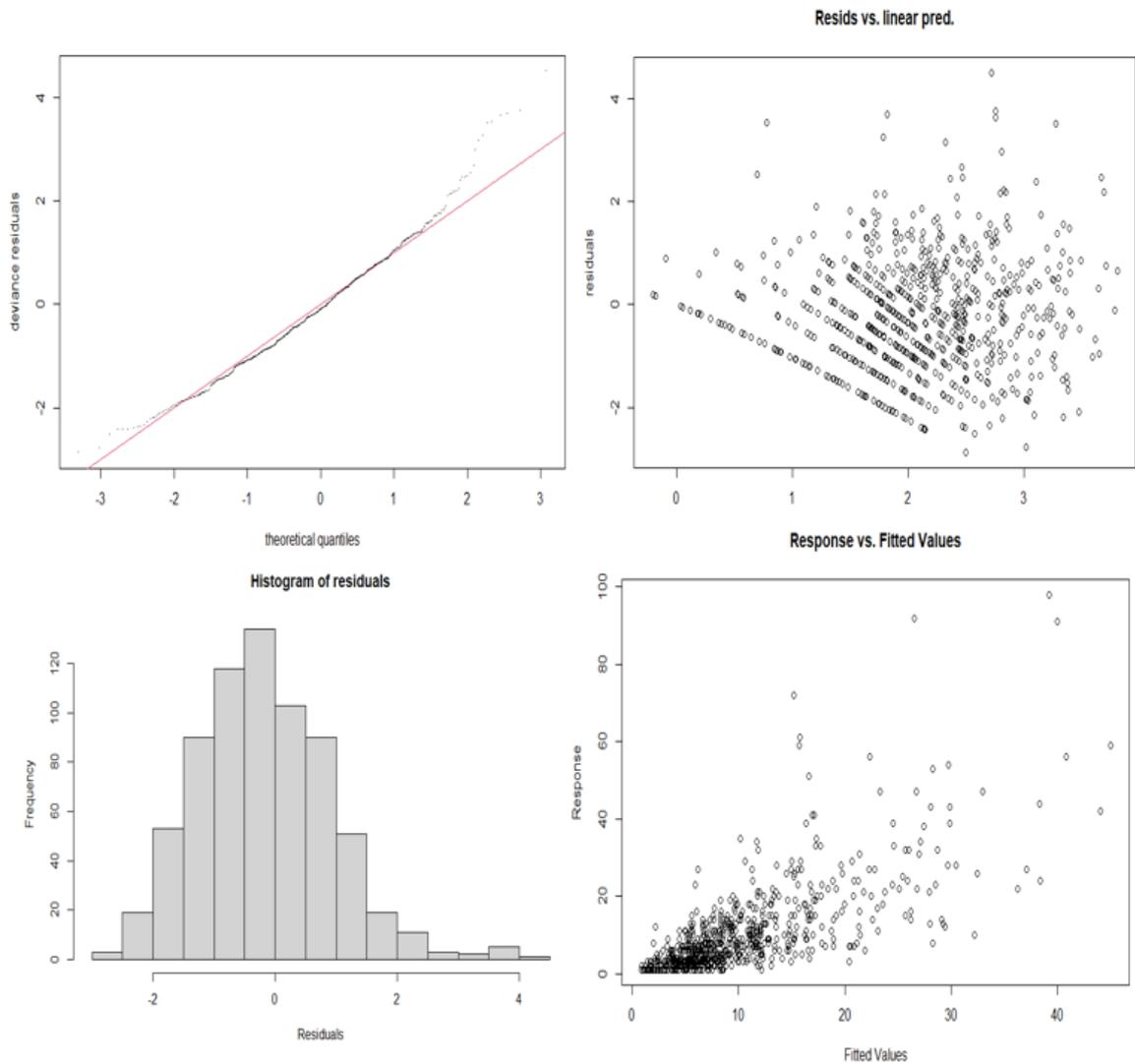
# Plot the model
plot(Linear_poptrend)

#CheckFit
par(mfrow=c(2,2))

AphTrend.nb.loglinear<-ptrend (Count ~ trend(Year, tempRE = TRUE, type =
"loglinear") + s(Trap, bs = "re"), family = nb(theta = 5.789227, link = "log"), data =
NRIB_ANNUAL_03DEC21, gamModel=TRUE)

```

```
checkFit(AphTrend.nb.loglinear)
```



```
par(mfrow=c(1,1))
```

```
#Check percentage change from Year 1965 to 2020  
plot(AphTrend.nb.loglinear, ciBase = 1965)  
change(AphTrend.nb.loglinear,1965,2020)  
summary(AphTrend.nb.loglinear$gam)
```

Family: Negative Binomial(5.789)

Link function: log

Formula:

Count ~ Year + s(Trap, bs = "re") + s(Year__Fac, bs = "re")

Parametric coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 53.945796 6.593923 8.181 2.81e-16 ***

Year -0.026023 0.003308 -7.866 3.67e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

edf Ref.df Chi.sq p-value

s(Trap) 17.39 18 468.2 <2e-16 ***

s(Year__Fac) 44.13 54 314.1 <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.456 Deviance explained = 59.2%

-REML = 2190 Scale est. = 1 n = 702

###

```
ggplot(data = NRIB_ANNUAL_03DEC21, aes(x = Year, y = Count)) +  
  geom_smooth(method = "lm") +  
  geom_point() +  
  facet_wrap(~Trap)
```

#without data points

```
ggplot(data = NRIB_ANNUAL_03DEC21, aes(x = Year, y = Count)) +  
  geom_smooth(method = "lm") +  
  facet_wrap(~Trap)
```

#WEST analysis

```
# Make the factors into factors)
NRIB_ANNUAL_WEST_03DEC21$fYear <-
paste0("Y_",NRIB_ANNUAL_WEST_03DEC21$Year)
NRIB_ANNUAL_WEST_03DEC21$Trap <-
as.factor(NRIB_ANNUAL_WEST_03DEC21$Trap)
NRIB_ANNUAL_WEST_03DEC21$fYear <-
as.factor(NRIB_ANNUAL_WEST_03DEC21$fYear)

# Models can be linear or non-linear. Poptrend models use the mgcv syntax
# Run a non-linear first
# Non linear (wiggly) #####

# Run a GAMM with a non-linear year effect
Wiggly_poptrend <- ptrend(Count ~ trend(Year,
                                     tempRE = TRUE, # This adds a random effect for year
                                     (tempRE = "Temporal Random Effect")
                                     type = "smooth", # This makes it smooth (wiggly) rather
than linear
                                     k = 16) + # This is the wiggliness of the curve. A rule of
thumb is to set as one third of
                                     # time series length - ie., 49 years divided by 3 = roughly 16.
Can be reduced for a straighter curve or
                                     # increased for a more wiggly one
                                     s(Year, Trap, bs = "re"), # This adds a random slope within each
year
                                     family = nb, # Negative binomial error distribution is usually good
for count data
                                     data = NRIB_ANNUAL_WEST_03DEC21) # Takes a litle while...

change(Wiggly_poptrend, 1965, 2020) # To get the percentage change over the
specified years plus 95% CIs.
# NOTE: The non-linear models are not appropriate for giving a good estimate of
percentage change as they
# are too dependent on the first and last year - use log-linear model for that instead
(see below)

# Plot the model
plot(Wiggly_poptrend)
```

```

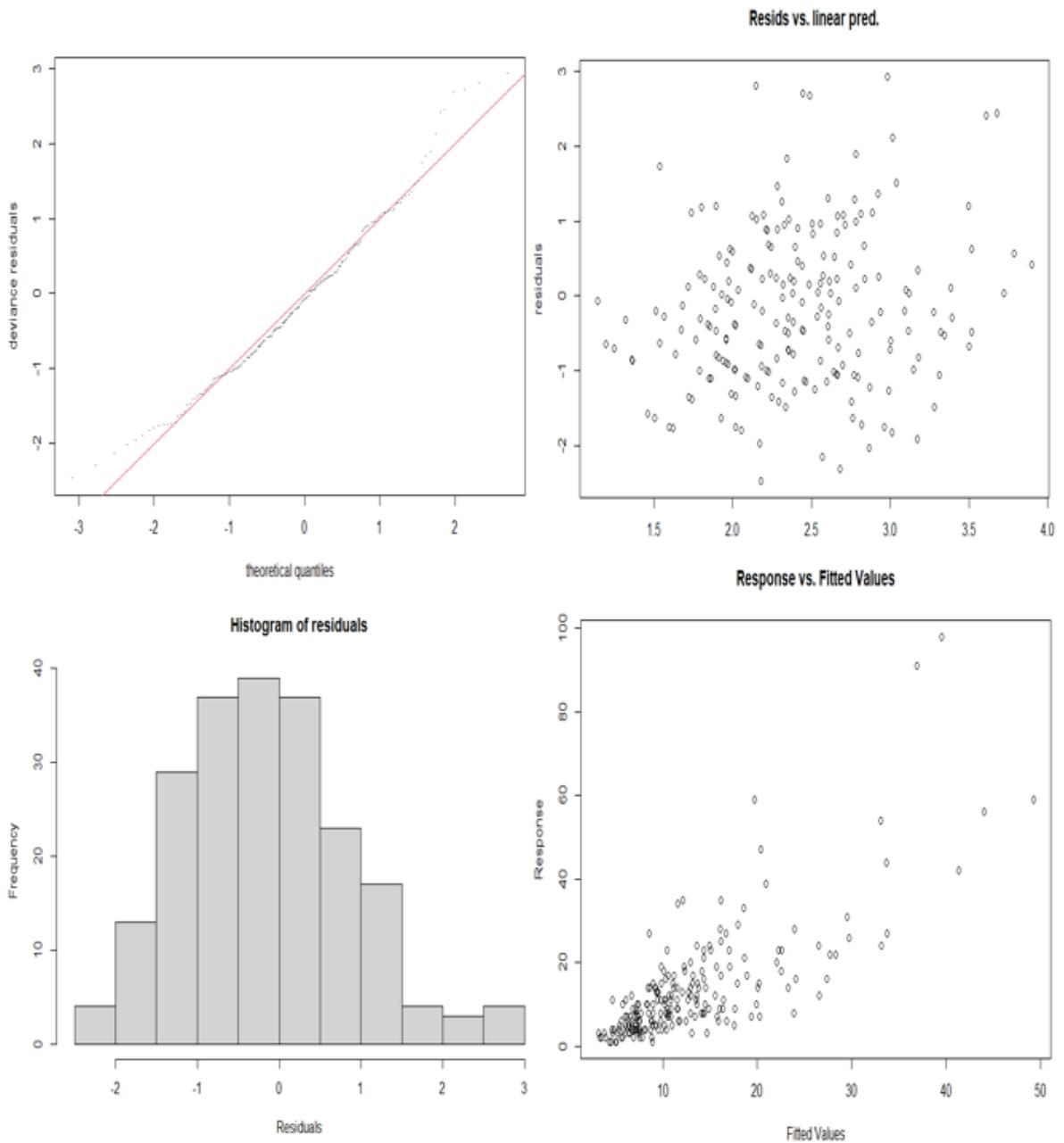
# Now do a linear version
# # # #
# Linear (log-linear) #####

# Run a GAMM with a linear year effect
Linear_poptrend <- ptrend(Count ~ trend(Year,
                                     tempRE = TRUE, # This adds a random effect for year
                                     (tempRE = "Temporal Random Effect")
                                     type = "loglinear") +
                        s(Year, Trap, bs = "re"), # This adds a random slope within each
year
                        family = nb, # Negative binomial error distribution is usually good
for count data
                        data = NRIB_ANNUAL_WEST_03DEC21)

change(Linear_poptrend, 1965, 2020) # To get the percentage change over the
specified years. NOTE: The non-linear models are
# not appropriate for giving a good estimate of percentage change as they are too
dependent on the first and last year

#CheckFit
par(mfrow=c(2,2))
AphTrend.nb.loglinear<-ptrend (Count ~ trend(Year, tempRE = TRUE, type =
"loglinear") + s(Trap, bs = "re"), family = nb(theta = 5.789227, link = "log"), data =
NRIB_ANNUAL_WEST_03DEC21, gamModel=TRUE)
checkFit(AphTrend.nb.loglinear)

```



```
par(mfrow=c(1,1))
```

```
#Check percentage change from Year 1965 to 2020
plot(AphTrend.nb.loglinear, ciBase = 1965)
```

```
change(AphTrend.nb.loglinear,1965,2020)
```

```
summary(AphTrend.nb.loglinear$gam)
```

```
Family: Negative Binomial(5.789)
```

```
Link function: log
```

```
Formula:
```

```
Count ~ Year + s(Trap, bs = "re") + s(Year__Fac, bs = "re")
```

```
Parametric coefficients:
```

```
Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept) 32.077970 9.954226 3.223 0.00127 **
```

```
Year -0.014869 0.004991 -2.979 0.00289 **
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
```

```
edf Ref.df Chi.sq p-value
```

```
s(Trap) 4.508 5 58.31 <2e-16 ***
```

```
s(Year__Fac) 36.712 50 152.98 <2e-16 ***
```

```
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) = 0.425 Deviance explained = 59.8%
```

```
-REML = 721.91 Scale est. = 1 n = 210
```

```
# Plot the model
```

```
plot(Linear_poptrend)
```

```
ggplot(data = NRIB_ANNUAL_WEST_03DEC21, aes(x = Year, y = Count)) +
```

```
geom_smooth(method = "lm") +
```

```
geom_point() +
```

```
facet_wrap(~Trap)
```

```
#without data points
ggplot(data = NRIB_ANNUAL_WEST_03DEC21, aes(x = Year, y = Count)) +
geom_smooth(method = "lm") +
facet_wrap(~Trap)
```

#EAST analysis

```
# Make the factors into factors)
```

```
NRIB_ANNUAL_EAST_03DEC21$fYear <-
paste0("Y_",NRIB_ANNUAL_EAST_03DEC21$Year)
NRIB_ANNUAL_EAST_03DEC21$Trap <-
as.factor(NRIB_ANNUAL_EAST_03DEC21$Trap)
NRIB_ANNUAL_EAST_03DEC21$fYear <-
as.factor(NRIB_ANNUAL_EAST_03DEC21$fYear)
```

```
# Models can be linear or non-linear. Poptrend models use the mgcv syntax
# Run a non-linear first
# Non linear (wiggly) #####
```

```
# Run a GAMM with a non-linear year effect
Wiggly_poptrend <- ptrend(Count ~ trend(Year,
```

```
tempRE = TRUE, # This adds a random effect for year
(tempRE = "Temporal Random Effect")
```

```
type = "smooth", # This makes it smooth (wiggly) rather
than linear
```

```
k = 16) + # This is the wiggleness of the curve. A rule of
thumb is to set as one third of
```

```
# time series length - ie., 49 years divided by 3 = roughly 16.
Can be reduced for a straighter curve or
```

```

# increased for a more wiggly one
s(Year, Trap, bs = "re"), # This adds a random slope within each
year

family = nb, # Negative binomial error distribution is usually good
for count data

data = NRIB_ANNUAL_EAST_03DEC21) # Takes a little while...

change(Wiggly_poptrend, 1965, 2020) # To get the percentage change over the
specified years plus 95% CIs.
# NOTE: The non-linear models are not appropriate for giving a good estimate of
percentage change as they
# are too dependent on the first and last year - use log-linear model for that instead
(see below)

# Plot the model
plot(Wiggly_poptrend)

# Now do a linear version
# # # #
# Linear (log-linear) #####

# Run a GAMM with a linear year effect
Linear_poptrend <- ptrend(Count ~ trend(Year,

                        tempRE = TRUE, # This adds a random effect for year
(tempRE = "Temporal Random Effect")

                        type = "loglinear") +

s(Year, Trap, bs = "re"), # This adds a random slope within each
year

family = nb, # Negative binomial error distribution is usually good
for count data

data = NRIB_ANNUAL_EAST_03DEC21)

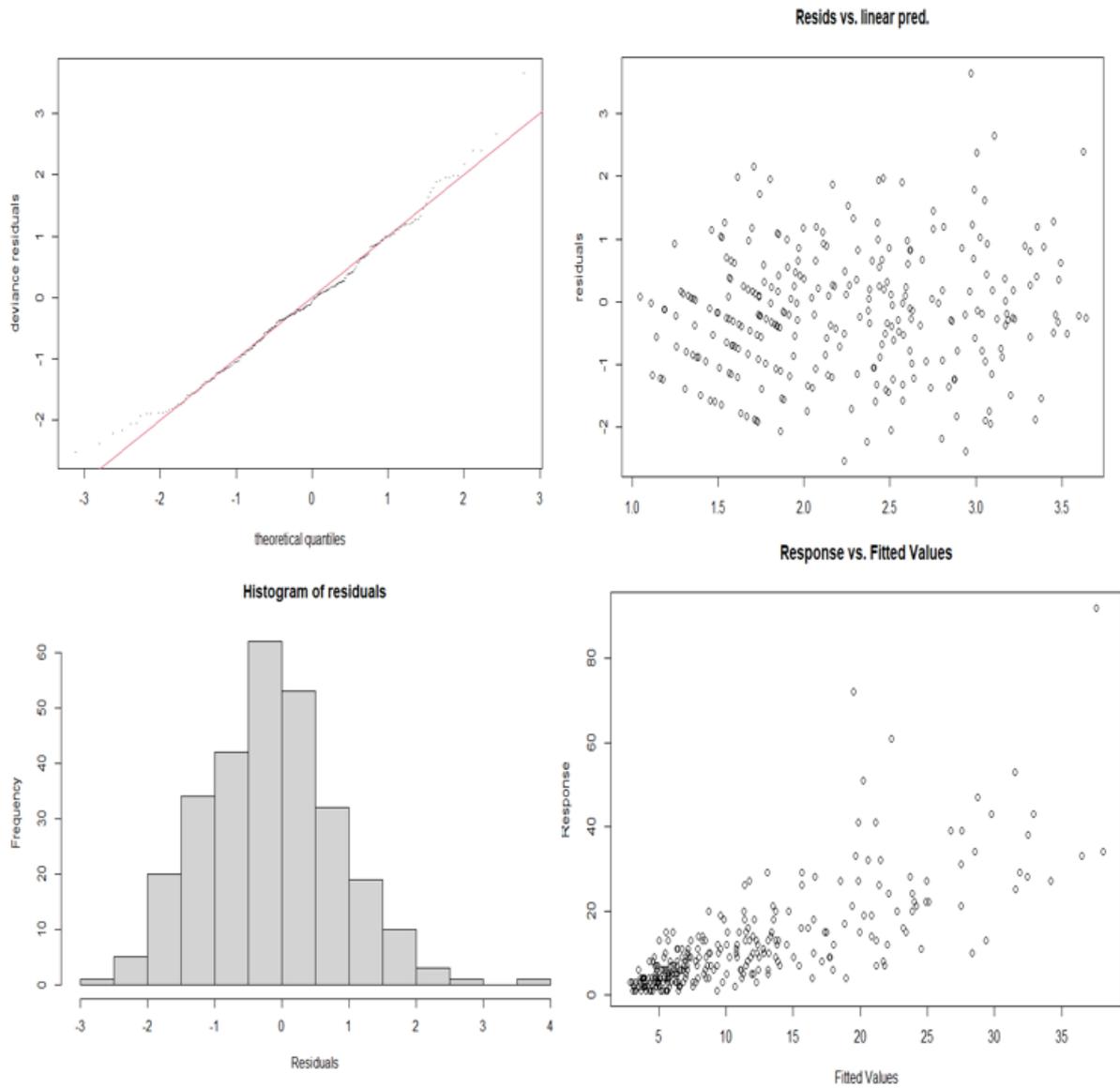
change(Linear_poptrend, 1965, 2020) # To get the percentage change over the
specified years. NOTE: The non-linear models are
# not appropriate for giving a good estimate of percentage change as they are too
dependent on the first and last year

#CheckFit
par(mfrow=c(2,2))

```

```
AphTrend.nb.loglinear<-ptrend (Count ~ trend(Year, tempRE = TRUE, type =  
"loglinear") + s(Trap, bs = "re"), family = nb(theta = 5.789227, link = "log"), data =  
NRIB_ANNUAL_EAST_03DEC21, gamModel=TRUE)
```

```
checkFit(AphTrend.nb.loglinear)
```



```
par(mfrow=c(1,1))
```

```

#Check percentage change from Year 1965 to 2020
plot(AphTrend.nb.loglinear, ciBase = 1965)

change(AphTrend.nb.loglinear,1965,2020)
summary(AphTrend.nb.loglinear$gam)

Family: Negative Binomial(5.789)
Link function: log

Formula:
Count ~ Year + s(Trap, bs = "re") + s(Year__Fac, bs = "re")

Parametric coefficients:

      Estimate Std. Error z value Pr(>|z|)
(Intercept) 70.767689  8.112874  8.723  <2e-16 ***
Year        -0.034364  0.004072 -8.439  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

      edf Ref.df Chi.sq p-value
s(Trap)    2.969    5  7.395 0.0382 *
s(Year__Fac) 40.221   54 165.983 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.47  Deviance explained = 64.3%
-REML = 908.88  Scale est. = 1      n = 283

# Plot the model
plot(Linear_poptrend)

ggplot(data = NRIB_ANNUAL_EAST_03DEC21, aes(x = Year, y = Count)) +
  geom_smooth(method = "lm") +
  geom_point() +
  facet_wrap(~Trap)

#without data points
ggplot(data = NRIB_ANNUAL_EAST_03DEC21, aes(x = Year, y = Count)) +
  geom_smooth(method = "lm") +
  facet_wrap(~Trap)

```

#NORTH analysis

```
# Make the factors into factors)
```

```
NRIB_ANNUAL_NORTH_03DEC21$fYear <-  
paste0("Y_",NRIB_ANNUAL_NORTH_03DEC21$Year)  
NRIB_ANNUAL_NORTH_03DEC21$Trap <-  
as.factor(NRIB_ANNUAL_NORTH_03DEC21$Trap)  
NRIB_ANNUAL_NORTH_03DEC21$fYear <-  
as.factor(NRIB_ANNUAL_NORTH_03DEC21$fYear)
```

```
# Models can be linear or non-linear. Poptrend models use the mgcv syntax
```

```
# Run a non-linear first
```

```
# Non linear (wiggly) #####
```

```
# Run a GAMM with a non-linear year effect
```

```
Wiggly_poptrend <- ptrend(Count ~ trend(Year,
```

```
tempRE = TRUE, # This adds a random effect for year  
(tempRE = "Temporal Random Effect")
```

```
type = "smooth", # This makes it smooth (wiggly) rather  
than linear
```

```
k = 16) + # This is the wiggleness of the curve. A rule of  
thumb is to set as one third of
```

```
# time series length - ie., 49 years divided by 3 = roughly 16.  
Can be reduced for a straighter curve or
```

```
# increased for a more wiggly one
```

```
s(Year, Trap, bs = "re"), # This adds a random slope within each  
year
```

```
family = nb, # Negative binomial error distribution is usually good  
for count data
```

```
data = NRIB_ANNUAL_NORTH_03DEC21) # Takes a little while...
```

```
change(Wiggly_poptrend, 1965, 2020) # To get the percentage change over the  
specified years plus 95% CIs.
```

```
# NOTE: The non-linear models are not appropriate for giving a good estimate of  
percentage change as they
```

```
# are too dependent on the first and last year - use log-linear model for that instead  
(see below)
```

```

# Plot the model
plot(Wiggly_poptrend)

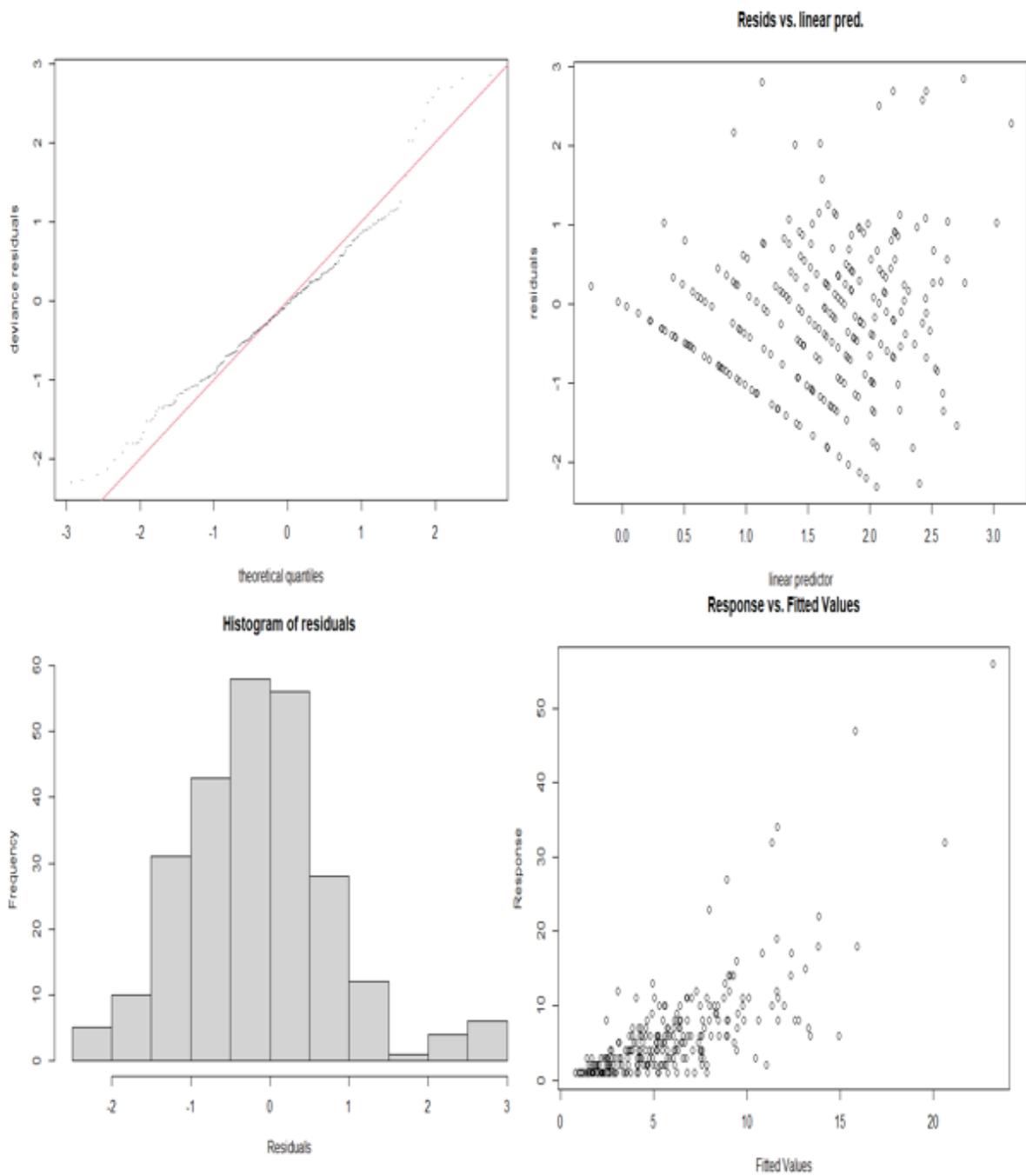
# Now do a linear version
# # # #
# Linear (log-linear) #####

# Run a GAMM with a linear year effect
Linear_poptrend <- ptrend(Count ~ trend(Year,
                                     tempRE = TRUE, # This adds a random effect for year
                                     (tempRE = "Temporal Random Effect")
                                     type = "loglinear") +
                        s(Year, Trap, bs = "re"), # This adds a random slope within each
year
                        family = nb, # Negative binomial error distribution is usually good
for count data
                        data = NRIB_ANNUAL_NORTH_03DEC21)

change(Linear_poptrend, 1965, 2020) # To get the percentage change over the
specified years. NOTE: The non-linear models are
# not appropriate for giving a good estimate of percentage change as they are too
dependent on the first and last year

#CheckFit
par(mfrow=c(2,2))
AphTrend.nb.loglinear<-ptrend (Count ~ trend(Year, tempRE = TRUE, type =
"loglinear") + s(Trap, bs = "re"), family = nb(theta = 5.789227, link = "log"), data =
NRIB_ANNUAL_NORTH_03DEC21, gamModel=TRUE)
checkFit(AphTrend.nb.loglinear)

```



`par(mfrow=c(1,1))`

```
#Check percentage change from Year 1965 to 2020
plot(AphTrend.nb.loglinear, ciBase = 1965)
change(AphTrend.nb.loglinear,1965,2020)
```

```
summary(AphTrend.nb.loglinear$gam)
Family: Negative Binomial(5.789)
```

```
Link function: log
```

```
Formula:
```

```
Count ~ Year + s(Trap, bs = "re") + s(Year__Fac, bs = "re")
```

```
Parametric coefficients:
```

```
Estimate Std. Error z value Pr(>|z|)
(Intercept) 40.441708  8.779893  4.606 4.10e-06 ***
Year        -0.019497  0.004406 -4.426 9.62e-06 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
```

```
edf Ref.df Chi.sq p-value
s(Trap)    6.666    7 158.39 <2e-16 ***
s(Year__Fac) 29.503   52  92.81 <2e-16 ***
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) = 0.402  Deviance explained = 59.7%
```

```
-REML = 670.12  Scale est. = 1      n = 254
```

```
# Plot the model
```

```
plot(Linear_poptrend)
```

```
ggplot(data = NRIB_ANNUAL_NORTH_03DEC21, aes(x = Year, y = Count)) +
  geom_smooth(method = "lm") +
  geom_point() +
  facet_wrap(~Trap)
```

```
#without data points
```

```
ggplot(data = NRIB_ANNUAL_NORTH_03DEC21, aes(x = Year, y = Count)) +
  geom_smooth(method = "lm") +
  facet_wrap(~Trap)
```

#SOUTH analysis

```
# Make the factors into factors)
```

```
NRIB_ANNUAL_SOUTH_03DEC21$fYear <-
paste0("Y_",NRIB_ANNUAL_SOUTH_03DEC21$Year)
```

```
NRIB_ANNUAL_SOUTH_03DEC21$Trap <-
as.factor(NRIB_ANNUAL_SOUTH_03DEC21$Trap)
```

```
NRIB_ANNUAL_SOUTH_03DEC21$fYear <-
as.factor(NRIB_ANNUAL_SOUTH_03DEC21$fYear)
```

```
# Models can be linear or non-linear. Poptrend models use the mgcv syntax
```

```
# Run a non-linear first
```

```
# Non linear (wiggly) #####
```

```
# Run a GAMM with a non-linear year effect
```

```
Wiggly_poptrend <- ptrend(Count ~ trend(Year,
```

```
tempRE = TRUE, # This adds a random effect for year
(tempRE = "Temporal Random Effect")
```

```
type = "smooth", # This makes it smooth (wiggly) rather
than linear
```

```
k = 16) + # This is the wiggleness of the curve. A rule of
thumb is to set as one third of
```

```
# time series length - ie., 49 years divided by 3 = roughly 16.
Can be reduced for a straighter curve or
```

```
# increased for a more wiggly one
```

```

                                s(Year, Trap, bs = "re"), # This adds a random slope within each
year
                                family = nb, # Negative binomial error distribution is usually good
for count data
                                data = NRIB_ANNUAL_SOUTH_03DEC21) # Takes a little while...

change(Wiggly_poptrend, 1965, 2020) # To get the percentage change over the
specified years plus 95% CIs.
# NOTE: The non-linear models are not appropriate for giving a good estimate of
percentage change as they
# are too dependent on the first and last year - use log-linear model for that instead
(see below)

# Plot the model
plot(Wiggly_poptrend)

# Now do a linear version
# # # #
# Linear (log-linear) ####
# Run a GAMM with a linear year effect
Linear_poptrend <- ptrend(Count ~ trend(Year,
                                tempRE = TRUE, # This adds a random effect for year
(tempRE = "Temporal Random Effect")
                                type = "loglinear") +
                                s(Year, Trap, bs = "re"), # This adds a random slope within each
year
                                family = nb, # Negative binomial error distribution is usually good
for count data
                                data = NRIB_ANNUAL_SOUTH_03DEC21)

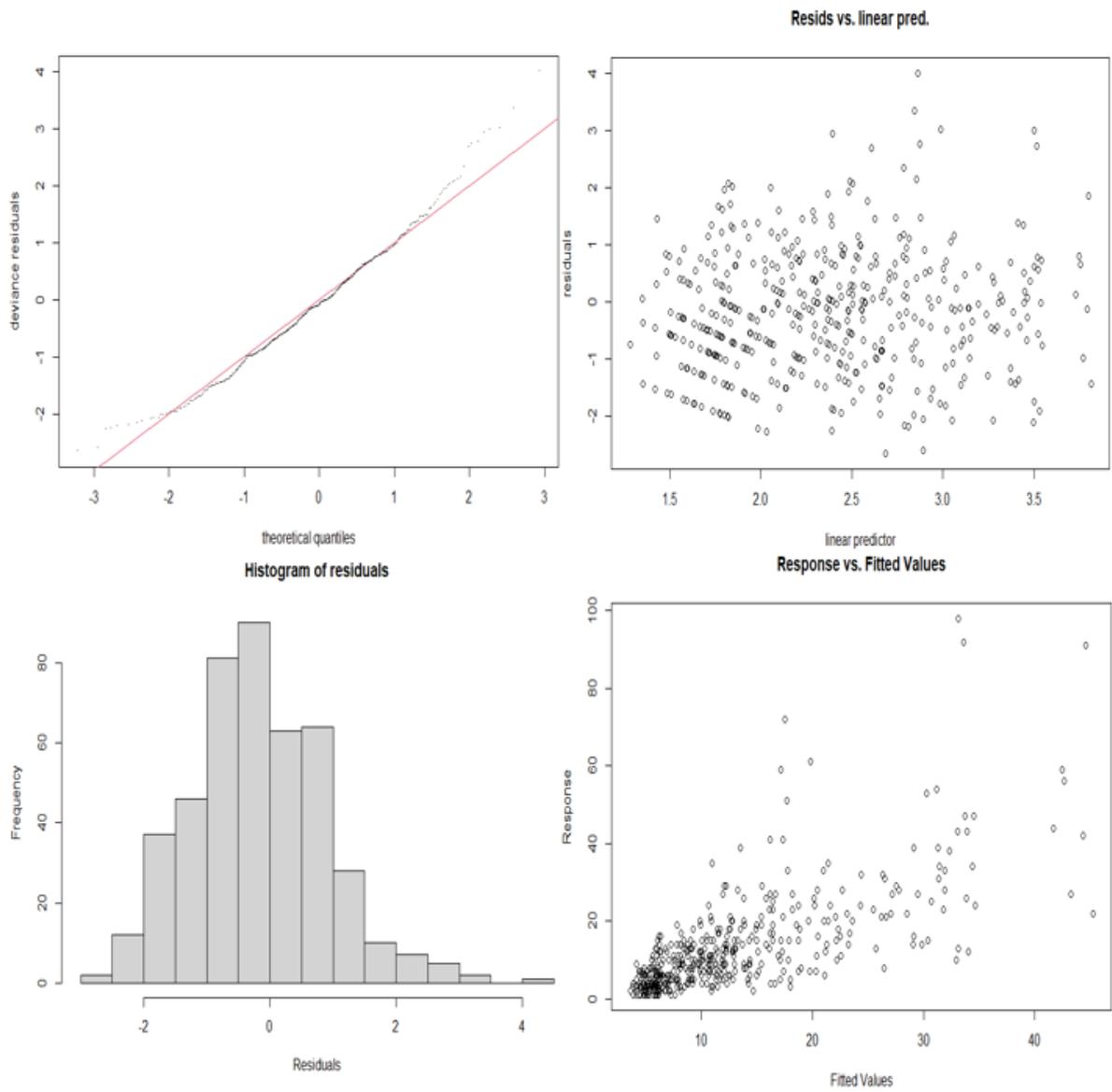
change(Linear_poptrend, 1965, 2020) # To get the percentage change over the
specified years. NOTE: The non-linear models are
# not appropriate for giving a good estimate of percentage change as they are too
dependent on the first and last year

#CheckFit
par(mfrow=c(2,2))

AphTrend.nb.loglinear<-ptrend (Count ~ trend(Year, tempRE = TRUE, type =
"loglinear") + s(Trap, bs = "re"), family = nb(theta = 5.789227, link = "log"), data =
NRIB_ANNUAL_SOUTH_03DEC21, gamModel=TRUE)

```

```
checkFit(AphTrend.nb.loglinear)
```



```
par(mfrow=c(1,1))
```

```
#Check percentage change from Year 1965 to 2020  
plot(AphTrend.nb.loglinear, ciBase = 1965)  
change(AphTrend.nb.loglinear,1965,2020)  
summary(AphTrend.nb.loglinear$gam)
```

Family: Negative Binomial(5.789)

Link function: log

Formula:

Count ~ Year + s(Trap, bs = "re") + s(Year__Fac, bs = "re")

Parametric coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 58.41377 7.67185 7.614 2.66e-14 ***

Year -0.02810 0.00385 -7.299 2.89e-13 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

edf Ref.df Chi.sq p-value

s(Trap) 7.832 10 44.23 <2e-16 ***

s(Year__Fac) 44.254 54 273.94 <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.43 Deviance explained = 57.6%

-REML = 1495.1 Scale est. = 1 n = 448

Plot the model

```
plot(Linear_poptrend)
```

```
ggplot(data = NRIB_ANNUAL_SOUTH_03DEC21, aes(x = Year, y = Count)) +
```

```
geom_smooth(method = "lm") +
```

```
geom_point() +
```

```
facet_wrap(~Trap)
```

```
#without data points
```

```
ggplot(data = NRIB_ANNUAL_SOUTH_03DEC21, aes(x = Year, y = Count)) +
```

```
geom_smooth(method = "lm") +
```

```
facet_wrap(~Trap)
```

R code – Experiment 2: Correlation of other aphid species presence for use as a 'proxy' for *N. ribisnigri* & Trend analysis of *N. ribisnigri* abundance (1965-2020) using 'poptrend'

Correlation of other aphid species presence or absence for use as a 'proxy' for *N. ribisnigri*

```
## Daily
data <- read.csv("data/aphids/community/allAphidsDaily.csv")
sp_names <- read.csv("data/aphids/taxonomy/spCodeName.csv")
rownames(sp_names) <- sp_names$sp

colnames(data)[5:18] <- na.omit(sp_names[colnames(data), "name"])
head(data)

library("corrplot")

cor_mat <- cor(na.omit(data[,5:18]))
corrplot(cor_mat)

cor_mat <- cor(log(na.omit(data[,5:18])+1))
corrplot(cor_mat)

## weekly
data$week <- data$day %/% 7
data_wk <- data %>% pivot_longer(cols=5:18) %>%
  group_by(year, x, y, week, name) %>% summarise(value=sum(value)) %>%
  pivot_wider(1:4, name)
```

```
cor_mat <- cor(log(na.omit(data_wk[,5:18])+1))
corrplot(cor_mat)
```

Forecasting first flight of *N. ribisnigri*

```
library("tidyr")
library("cartography")
library("ggplot2")
library("RColorBrewer")
library("dplyr")
library("mgcv")
library("sp")

myPalette <- colorRampPalette(brewer.pal(11, "Spectral"))

# -----
# load and arrange first flight data
# -----

data <- read.csv("data/N_ribisnigri.updated5.csv")
# convert to date format
for (x in 1:ncol(data)){data[,x] <- as.Date(data[,x], format="%d/%m/%Y")}

# convert to long format
data <- gather(data, key="trap", value="date")

# convert to year and julian day
data$day <- as.POSIXlt(data$date)$yday
data$year <- as.POSIXlt(data$date)$year + 1900
data <- na.omit(data[,-2])

# add coordinates to data
# -----
traps <- read.csv("data/traps_locations.csv")
traps$Site.Name <- as.character(traps$Site.Name )

# change some trap names to correspond to the trap data names
data$trap[data$trap=="Brooms.Barn"] <- "Brooms Barn"
data$trap[data$trap=="i..Askham.Bryan.S"] <- "Askham Bryan"
data$trap[data$trap=="Farm.One"] <- "Rothamsted Farm"
data$trap[data$trap=="Farm.Two"] <- "Rothamsted Farm II"
data$trap[data$trap=="Hereford"] <- "Rosemaund, Hereford"
data$trap[data$trap=="High.Mowthorpe"] <- "High Mowthorpe"
```

```

data$trap[data$trap=="Kirton.2"] <- "Kirton II"
data$trap[data$trap=="Long.Ashton.One"] <- "Long Ashton"
data$trap[data$trap=="Long.Ashton.Two"] <- "Long Ashton II"
data$trap[data$trap=="North.Farm"] <- "North Farm"
data$trap[data$trap=="Rainham.One"] <- "Rainham"
data$trap[data$trap=="Rainham.Two"] <- "Rainham II"
data$trap[data$trap=="Roseware"] <- "Rosewarne"
data$trap[data$trap=="Rothamsted"] <- "Rothamsted Tower"
data$trap[data$trap=="Silwood"] <- "Silwood Park"
data$trap[data$trap=="Wellesbourne.S"] <- "Wellesbourne"
data$trap[data$trap=="Writtle.S"] <- "Writtle"

# add spatial coordinates to data according to trap names
idx <- unlist(apply(data, 1, FUN=function(x) which(traps$Site.Name==x[1])))
data$lat <- traps$Latitude[idx]
data$lon <- traps$Longitude[idx]

# convert LatLon to planar projection, British National Grid
pts <- spTransform(SpatialPoints(cbind(data$lon, data$lat),
                                   proj4string = CRS("+proj=longlat +datum=WGS84 +no_defs
+ellps=WGS84")),
                  "+init=epsg:27700")

data$x <- pts@coords[,1] / 1000 # in km instead of meters
data$y <- pts@coords[,2] / 1000

# Representations
# -----
# discard samples where detection occurred past early june (day 160)
data <- data[data$day <= 160, ]

ggplot(data, aes(day)) + geom_histogram() + theme_bw() +
  scale_x_log10() + facet_wrap(~trap)

ggplot(data) + geom_segment(aes(y=year, yend=year, xend=day, x=0)) +
  theme_bw() +
  facet_wrap(~trap)

# import SPDF from cartography
UKspdf <- nuts0.spdf[nuts0.spdf$id == "UK",]
UKspdf <- spTransform(UKspdf, "+init=epsg:27700 +units=km" )
poly.df <- fortify(UKspdf) %>% setNames(., c("x", "y", colnames(fortify(UKspdf)))[-

```

```

(1:2]))
poly <- geom_path(data=poly.df, aes(x=x, y=y, group=group), alpha=.5, size=.8)

ggplot(data) + poly + coord_equal() + theme_minimal() + facet_wrap(~year,
ncol=14) +

  geom_point(aes(x=x, y=y, fill=day), shape=21, size=3) +

  scale_fill_gradientn("Day of\nfirst flight", colours=(myPalette(100))) +

  theme(axis.title = element_blank(), axis.text = element_blank())

```

```

# reduce huge meteo data to trap locations (NO NEED TO RUN AGAIN)
# -----
# meteo <- read.csv("data/meteo.csv") # ~ 1GB
# coords.met <- unique(meteo[,c("x", "y")])
# coords.dat <- unique(data[, c("x", "y", "trap")])
# idx <- sapply(1:nrow(coords.dat), function(x) which.min(sqrt((coords.met$x -
coords.dat[x,1])^2 +
#
#                               (coords.met$y - coords.dat[x,2])^2)))
# coords.met <- cbind(coords.met[idx,], trap=coords.dat$trap)
# meteo <- merge(meteo, coords.met)
# meteo <- meteo[order(meteo$year, meteo$day, meteo$x, meteo$y),]
# write.csv(meteo, file="data/small_meteo.csv", row.names = FALSE) # ~ 30 MB

```

```

# -----

```

```

# Analysis

```

```

# -----

```

```

meteo <- read.csv("data/small_meteo.csv")
meteo <- meteo[,c("trap", "year", "day", "temp")]

# degree day model function
degree.day <- function(temp, start=1, end, th.lo=0, th.hi=35){
  sum(pmin(pmax(temp - th.lo, 0), th.hi-th.lo)[start:end])
}

# divide data in training/testing sets
data$trained <- sample(c(TRUE, FALSE), nrow(data), prob=c(.7,.3), replace=TRUE)

# discard samples where detection occurred past early June (day 180)
data <- data[data$day <= 160, ]

# hyper parameters grid
grid <- expand.grid(start=seq(1, 160, by=5),
                   270

```

```

        end=seq(5, 160, by=5),
        MAE=NA, RMSE=NA, R2=NA,
        MAEtrn=NA, RMSEtrn=NA, R2trn=NA,
        MAEtst=NA, RMSEtst=NA, R2tst=NA)
grid <- grid[grid$end > grid$start,]
for (i in 1:nrow(grid)){
  # convert meteo to degree day depending on start and end
  c1 <- c2 <- c3 <- c()
  for (tr in unique(meteo$trap)){
    temp <- meteo[meteo$trap == tr,]
    for (yr in unique(temp$year)){
      temp2 <- temp[temp$year==yr,]
      temp2 <- temp2[order(temp2$day),]

      c1 <- c(c1, yr)
      c2 <- c(c2, tr)
      c3 <- c(c3, sum(pmin(pmax(temp2$temp - 6, 0), 26-6)[grid[i, "start"]:grid[i,
"end"]]))
    }
  }
  df <- data.frame(year=c1, trap=c2, dd=c3)
  ## same with cleaner syntax but longer to run...
  # df <- meteo %>%
  #   group_by(trap, year) %>%
  #   do(data.frame(., dd = degree.day(temp=.$temp, start=grid[i, "start"],
  #                                   end=grid[i, "end"])))

  # merge degree day and ff data
  df <- unique(df[,c("trap", "year", "dd")])
  df <- merge(data, df)

```

```

plot(df$day, x=df$dd)
lines(x=df$dd, y=predict(lm(data=df, day~dd)), col="red")

# iterate the CV steps for less noisy outputs
err <- c()

for (j in 1:30){
  m <- lm(data=df[df$trained,], log(day)~dd)
  err <- rbind(err, data.frame(exp(predict(m, df)), df[, "day"], trained=df$trained))
# m <- gam(data=df[df$trained,], log(day)~s(dd))
# err <- rbind(err, cbind(exp(predict(m, df[!df$trained, ])), df[!df$trained, "day"]))

df$trained <- sample(df$trained) # shuffle test/train
}

# plot(x=err[,1], y=err[,2]) ; abline(0,1, col="red")
grid[i, "MAE"] <- mean(abs(err[,1]-err[,2])) # MAE
grid[i, "RMSE"] <- sqrt(mean((err[,1]-err[,2])^2)) # RMSE
grid[i, "R2"] <- 1 - sum((err[,1]-err[,2])^2) / sum((err[,2]-mean(err[,2]))^2)
grid[i, "MAEtest"] <- mean(abs(err[!err$trained,1]-err[!err$trained,2])) # MAE
grid[i, "RMSEtest"] <- sqrt(mean((err[!err$trained,1]-err[!err$trained,2])^2)) #
RMSE
grid[i, "R2test"] <- 1 - sum((err[!err$trained,1]-err[!err$trained,2])^2) /
  sum((err[!err$trained,2]-mean(err[!err$trained,2]))^2)

grid[i, "MAEtrn"] <- mean(abs(err[err$trained,1]-err[err$trained,2])) # MAE
grid[i, "RMSEtrn"] <- sqrt(mean((err[err$trained,1]-err[err$trained,2])^2)) # RMSE
grid[i, "R2trn"] <- 1 - sum((err[err$trained,1]-err[err$trained,2])^2) /
sum((err[err$trained,2]-mean(err[err$trained,2]))^2)

  print(grid[i,1:5])
}

# plot
# -----
write.csv(grid, "data/start_end13.csv")

# check the results from here
grid <- read.csv("data/start_end13.csv")

```

```

ggplot(grid) + geom_raster(aes(x=start, y=end, fill=R2)) +
scale_fill_gradientn(colours=rev(myPalette(100)))

gather(grid, key="crit", value="error", MAE:R2tst) %>%
  group_by(crit) %>%
  do(data.frame(start=.$start, end=.$end, crit=.$crit,
                error=scale(.$error))) %>%
  ggplot + geom_raster(aes(x=start, y=end, fill=error), show.legend = FALSE) +
  scale_fill_gradientn(colours=rev(myPalette(100))) + facet_wrap(~crit) +
  theme_minimal()

# check which days are best start/end
grid[which.min(grid$RMSE),]

# compute degree-days according to these days
c1 <- c2 <- c3 <- c()

for (tr in unique(meteo$trap)){
  temp <- meteo[meteo$trap == tr,]
  for (yr in unique(temp$year)){
    temp2 <- temp[temp$year==yr,]
    temp2 <- temp2[order(temp2$day),]

    c1 <- c(c1, yr)
    c2 <- c(c2, tr)
    c3 <- c(c3, sum(pmin(pmax(temp2$temp - 6, 0), 26-6)[15:153]))
    ##### here
  }
}

df <- data.frame(year=c1, trap=c2, dd=c3)
df <- unique(df[,c("trap", "year", "dd")])
df <- merge(data, df)

# again, a negative relationship
filter(df, day <= 150) %>%

```

```
ggplot() + geom_point(aes(x=dd, y=day, colour=y)) +  
scale_colour_gradientn(colours=(myPalette(100))) +  
geom_smooth(method="lm", aes(x=dd, y=day)) + theme_bw()  
m <- lm(data=df, day~dd)  
summary(m)
```

Appendix 2: Chapter 3 supplementary material

Table 1: List of all broadleaf and grass species found in Experiment 1.

Hainey (organic) site (Cambridgeshire)

Grasses

Common name	latin	Family
Annual bluegrass	<i>Poa annua</i>	Poaceae
Barley	<i>Hordeum vulgare</i>	Poaceae
Barren brome	<i>Anisantha sterilis</i>	Poaceae
Cocks foot	<i>Dactylis glomerata</i>	Poaceae
Creeping bent	<i>Agrostis stolonifera</i>	Poaceae
Crested dogs tail	<i>Cynosurus cristatus</i>	Poaceae
False oat	<i>Arrhenatherum elatius</i>	Poaceae
fescue	<i>Festuca spp.</i>	Poaceae
Fox tail	<i>Alopecurus pratensis</i>	Poaceae
Italian rye	<i>Festuca perennis</i>	Poaceae
oat grass	<i>Arrhenatherum spp.</i>	Poaceae
Perrenial rye	<i>Lolium perenne</i>	Poaceae
Red fescue	<i>Festuca rubra</i>	Poaceae
Timothy	<i>Phleum pratense</i>	Poaceae
Yorkshire fog	<i>Holcus lanatus</i>	Poaceae

Broadleaves

Common name	latin	Family
Yarrow	<i>Achillea millefolium</i>	Asteraceae
Corn chamomile	<i>Anthemis arvensis</i>	Asteraceae
Burdock	<i>Arctium spp.</i>	Asteraceae
Mugwort	<i>Artemisia vulgaris</i>	Asteraceae
Knapweed	<i>Centaurea nigra</i>	Asteraceae
Wetted thistle	<i>Carduus crispus</i>	Asteraceae

Common thistle	<i>Cirsium vulgare</i>	Asteraceae
corn marigold	<i>Glebionis segetum</i>	Asteraceae
	<i>Helminthotheca</i>	
bristly ox tongue	<i>echioides</i>	Asteraceae
Prickly lettuce	<i>Lactuca serriola</i>	Asteraceae
Bitter lettuce	<i>Lactuca virosa</i>	Asteraceae
Nipplewort	<i>Lapsana communis</i>	Asteraceae
	<i>Leucanthemum</i>	
Dog Daisy	<i>vulgare</i>	Asteraceae
Hawkweed oxtongue	<i>Picris hieracioides</i>	Asteraceae
Prickly sowthistle	<i>Sonchus asper</i>	Asteraceae
Smooth sow thistle	<i>Sonchus oleraceus</i>	Asteraceae
Tansy	<i>Tanacetum vulgare</i>	Asteraceae
Dandelion	<i>Taraxacum spp.</i>	Asteraceae
Orache	<i>Atriplex hortensis</i>	Amaranthaceae
White goosefoot (fat hen)	<i>Chenopodium album</i>	Amaranthaceae
Cow parsley	<i>Anthriscus sylvestris</i>	Apiaceae
Hemlock	<i>Conium maculatum</i>	Apiaceae
Wild carrot	<i>Daucus carota</i>	Apiaceae
Hogweed	<i>Heracleum spp.</i>	Apiaceae
	<i>Capsella bursa-</i>	
Shepards purse	<i>pastoris</i>	Brassicaceae
Cabbage	<i>Brassica oleracea</i>	Brassicaceae
Phacelia (Lady)	<i>Phacelia tanacetifolia</i>	Boraginaceae
Hedge mustard	<i>Sisymbrium officinale</i>	Brassicaceae
Comfrey	<i>Symphytum spp.</i>	Boraginaceae
Red campion	<i>Silene dioica</i>	Caryophyllaceae
Chickweed	<i>Stellaria media</i>	Caryophyllaceae
Bindweed	<i>Calystegia sepium</i>	Convolvulaceae
Black bryony	<i>Dioscoreae communis</i>	Dioscoreaceae

Bird's-foot trefoil	<i>Lotus corniculatus</i>	Fabaceae
Black medick	<i>Medicago lupulina</i>	Fabaceae
Lucerne (alfalfa)	<i>Medicago sativa</i>	Fabaceae
White melilot	<i>Melilotus albus</i>	Fabaceae
Melilot	<i>Melilotus altissimus</i>	Fabaceae
Peafowl yellow clover	<i>Melilotus officinalis</i>	Fabaceae
lesser trefoil	<i>Trifolium dubium</i>	Fabaceae
White clover	<i>Trifolium repens</i>	Fabaceae
Red clover	<i>Trifolium pratense</i>	Fabaceae
Tufted vetch	<i>Vicia cracca</i>	Fabaceae
Cut-leaved crane's-bill	<i>Geranium dissectum</i>	Geraniaceae
Dove's-foot crane's-bill	<i>Geranium molle</i>	Geraniaceae
White deadnettle	<i>Lamium album</i>	Lamiaceae
Gypsywort	<i>Lycopus europaeus</i>	Lamiaceae
Self-heal	<i>Prunella vulgaris</i>	Lamiaceae
Mallow	<i>Malva sylvestris</i>	Malvaceae
Red poppy	<i>Papaver rhoeas</i>	Papaveraceae
Ribwort plantain	<i>Plantago lanceolata</i>	Plantaginaceae
Broadleaf plantain	<i>Plantago major</i>	Plantaginaceae
Seed head plantain type	<i>Plantago spp.</i>	Plantaginaceae
Knot grass	<i>Polygonum aviculare</i>	Polygonaceae
Dock	<i>Rumex obtusifolius</i>	Polygonaceae
salad burnet	<i>Sanguisorba minor</i>	Rosaceae
Stinging nettle	<i>Urtica dioica</i>	Rosaceae
Goosegrass (cleavers)	<i>Galium aparine</i>	Rubiaceae
hedge bedstraw	<i>Galium mollugo</i>	Rubiaceae
lady bedstraw	<i>Galium verum</i>	Rubiaceae

Wissington (conventional) site (Norfolk)

Grasses

Common name	latin	Family
Barren brome	<i>Anisantha sterilis</i>	Poaceae
Cocks foot	<i>Dactylis glomerata</i>	Poaceae
Creeping bent	<i>Agrostis stolonifera</i>	Poaceae
Crested dogs tail	<i>Cynosurus cristatus</i>	Poaceae
False oat	<i>Arrhenatherum elatius</i>	Poaceae
Italian rye	<i>Festuca perennis</i>	Poaceae
oat grass	<i>Arrhenatherum spp.</i>	Poaceae
Perrenial rye	<i>Lolium perenne</i>	Poaceae
Timothy	<i>Phleum pratense</i>	Poaceae
Yorkshire fog	<i>Holcus lanatus</i>	Poaceae
Bulrush	<i>Typha latifolia</i>	Typhaceae

Broadleafs

Common name	latin	Family
Yarrow	<i>Achillea millefolium</i>	Asteraceae
Burdock	<i>Arctium spp.</i>	Asteraceae
Mugwort	<i>Artemisia vulgaris</i>	Asteraceae
Knapweed	<i>Centaurea nigra</i>	Asteraceae
Common thistle	<i>Cirsium vulgare</i>	Asteraceae
Nipplewort	<i>Lapsana communis</i>	Asteraceae
Dog Daisy	<i>Leucanthemum vulgare</i>	Asteraceae
Smooth sow thistle	<i>Sonchus oleraceus</i>	Asteraceae
Dandelion	<i>Taraxacum spp.</i>	Asteraceae
Orache	<i>Atriplex hortensis</i>	Amaranthaceae
White goosefoot (fat hen)	<i>Chenopodium album</i>	Amaranthaceae

Cow parsley	<i>Anthriscus sylvestris</i>	Apiaceae
Hemlock	<i>Conium maculatum</i>	Apiaceae
Wild carrot	<i>Daucus carota</i>	Apiaceae
Hogweed	<i>Heracleum spp.</i>	Apiaceae
Comfrey	<i>Symphytum spp.</i>	Boraginales
Red campion	<i>Silene dioica</i>	Caryophyllaceae
Chickweed	<i>Stellaria media</i>	Caryophyllaceae
Bindweed	<i>Calystegia sepium</i>	Convolvulaceae
Black bryony	<i>Dioscoreae communis</i>	Dioscoreaceae
Birdsfoot trefoil	<i>Lotus corniculatus</i>	Fabaceae
Black medick	<i>Medicago lupulina</i>	Fabaceae
Lucerne (alfalfa)	<i>Medicago sativa</i>	Fabaceae
Melilot	<i>Melilotus altissimus</i>	Fabaceae
Peafowl yellow clover	<i>Melilotus officinalis</i>	Fabaceae
lesser trefoil	<i>Trifolium dubium</i>	Fabaceae
Red clover	<i>Trifolium pratense</i>	Fabaceae
White clover	<i>Trifolium repens</i>	Fabaceae
Cut-leaved Crane's-bill	<i>Geranium dissectum</i>	Geraniaceae
Gypsywort	<i>Lycopus europaeus</i>	Lamiaceae
Broadleaf plantain	<i>Plantago major</i>	Plantaginaceae
Knot grass	<i>Polygonum aviculare</i>	Polygonaceae
Dock	<i>Rumex obtusifolius</i>	Polygonaceae
Stinging nettle	<i>Urtica dioica</i>	Rosales
Goosegrass (cleavers)	<i>Galium aparine</i>	Rubiaceae

Table 2: Number of *N. ribisnigri* found at the Hainey (organic) site during experiment 3 (green bridge).

Hainey (organic)

Field Week	<i>L. sativa</i>				<i>H. pilosella</i>				<i>V. chamaedrys</i>				<i>L. hispidus</i>				<i>C. intybus</i>			
	G3	R54	R20	B(S)	G3	R54	R20	B(S)	G3	R54	R20	B(S)	G3	R54	R20	B(S)	G3	R54	R20	B(S)
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	2	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	27	0	0	
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38	19	0	0	
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	5	0	0	
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Table 3: Number of *N. ribisnigri* found at the Wissington (conventional) site during experiment 3 (green bridge).

Wissington (conventional)

Field	<i>L. sativa</i>				<i>H. pilosella</i>				<i>V. chamaedrys</i>				<i>L. hispidus</i>				<i>C. intybus</i>							
	G3	R54	R20	B(S)	G3	R54	R20	B(S)	G3	R54	R20	B(S)	G3	R54	R20	B(S)	G3	R54	R20	B(S)				
Week																								
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 4: Blastn results from NCBI nucleotide collection database. Sequence length is unclipped length from Sanger sequencing results. * Indicates additional search parameter refinement was needed for sample. No samples out of the 18 sequenced contained any unsure nucleotides (N).

Sample	Seq length	ID	Identity %
pos1	542	<i>ITS1 L. sativa</i>	92.6
pos2	542	<i>ITS1 L. sativa</i>	92.6
neg	0	-	0
0hr1	649	<i>ITS1 L. sativa</i>	90
0hr2	641	<i>ITS1 L. sativa</i>	91.4
0hr3	319	<i>ITS1 L. sativa</i>	91.4
3hr1	642	<i>ITS1 L. sativa</i>	91.6
3hr2	638	Fail	0
3hr3	634	<i>ITS1 L. sativa</i>	91.2
6hr1	640	<i>ITS1 L. sativa</i>	92.3
6hr2	685	<i>ITS1 L. sativa</i>	92.6
6hr3	651	<i>ITS1 L. sativa</i>	91.3
24hr1	646	<i>ITS1 L. sativa</i> *	92
24hr2	646	<i>ITS1 L. sativa</i>	91.6
24hr3	641	<i>ITS1 L. sativa</i>	89
48hr1	632	<i>ITS1 L. sativa</i>	89
48hr2	647	<i>ITS1 L. sativa</i>	87.6
48hr3	691	<i>ITS1 L. sativa</i>	91.3

Appendix 3: Chapter 4 supplementary material

9.1 Supplementary 1: A step-by-step guide to hybrid de novo genome assembly in the command line (Bash script)

MaSuRCA was used for *de novo* assembly

What is required for assembly:

- Short read (Illumina) sequence data (forward and reverse)
- Long read (Nanopore) sequence data (single end)
- PuTTY (software centre)
- Notepad++ (software centre)

Command line:

PuTTY (software centre)

Host name: lily (open – leave on default settings)

Log in (password doesn't show up for some reason, but it is working!)

pwd (present working directory)

#change directory location to correct genome folder (e.g. Kent_CL_1)

cd

/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/Novogene_assembly/N_ribisnigri_assembly_R_Novogene/Assembly/Kent_CL_1

ll and **ls** #lists what is in that folder location

#connect command line to MaSuRCA pathway for *de novo* assembly

export PATH=/home/data/bioinf_resources/programming_tools/MaSuRCA-3.2.9b/bin:\$PATH

Now before you start to assembly we need a configuration file (config file) for MaSuRCA to be able to run. This essentially is a set of instructions which tell MaSuRCA what parameters it is working with and the location of your short and long read sequence data for it to *de novo* assemble into a genome.

MaSuRCA configuration file

```
# example configuration file
# DATA is specified as type {PE,JUMP,OTHER,PACBIO} and 5 fields:
# 1)two_letter_prefix 2)mean 3)stdev 4)fastq(.gz)_fwd_reads
# 5)fastq(.gz)_rev_reads. The PE reads are always assumed to be
# innies, i.e. --->.<---, and JUMP are assumed to be outties
# <--->.<--->. If there are any jump libraries that are innies, such as
# longjump, specify them as JUMP and specify NEGATIVE mean. Reverse reads
# are optional for PE libraries and mandatory for JUMP libraries. Any
# OTHER sequence data (454, Sanger, Ion torrent, etc) must be first
# converted into Celera Assembler compatible .frg files (see
# http://wgs-assembler.sourceforge.com)

DATA
#Illumina paired end reads supplied as <two-character prefix> <fragment mean> <fragment stdev> <forward_reads> <reverse_reads>
#if single-end, do not specify <reverse_reads>
#MUST HAVE Illumina paired end reads to use MaSuRCA
PE= pe 150 15
/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/N_ribisnigri_assembly_sequences/SeqData_DNA/Kent_CL_1_1.fq.gz
/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/N_ribisnigri_assembly_sequences/SeqData_DNA/Kent_CL_1_2.fq.gz
PE= se 125 50
/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/N_ribisnigri_assembly_sequences/SeqData_DNA/out_AphidFinalMappe
d_0.fq.gz
#Illumina mate pair reads supplied as <two-character prefix> <fragment mean> <fragment stdev> <forward_reads> <reverse_reads>
#JUMP= sh 3600 200 /FULL_PATH/short_1.fastq /FULL_PATH/short_2.fastq
#pacbio OR nanopore reads must be in a single fasta or fastq file with absolute path, can be gzipped
#if you have both types of reads supply them both as NANOPORE type
#PACBIO=/FULL_PATH/pacbio.fa
NANOPORE=/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/N_ribisnigri_assembly_sequences/MinION_reads/merged.f
q
#Other reads (Sanger, 454, etc) one frg file, concatenate your frg files into one if you have many
#OTHER=/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/N_ribisnigri_assembly_sequences/SeqData_DNA/out_AphidFi
nalMapped_0.fq.gz
END

PARAMETERS
#set this to 1 if your Illumina jumping library reads are shorter than 100bp
EXTEND_JUMP_READS=0
#this is k-mer size for deBruijn graph values between 25 and 127 are supported, auto will compute the optimal size based on the read data and GC
content
GRAPH_KMER_SIZE = auto
#set this to 1 for all Illumina-only assemblies
#set this to 0 if you have more than 15x coverage by long reads (Pacbio or Nanopore) or any other long reads/mate pairs (Illumina MP, Sanger, 454, etc)
USE_LINKING_MATES = 0
#specifies whether to run mega-reads correction on the grid
USE_GRID=0
#specifies grid engine to use SGE or SLURM
#specifies queue (for SGE) or partition (for SLURM) to use when running on the grid MANDATORY
#GRID_QUEUE=standard
#GRID_ENGINE=SLURM
#NUM_THREADS=20
#batch size in the amount of long read sequence for each batch on the grid
GRID_BATCH_SIZE=300000000
#use at most this much coverage by the longest Pacbio or Nanopore reads, discard the rest of the reads
LHE_COVERAGE=30
#set to 1 to only do one pass of mega-reads, for faster but worse quality assembly
MEGA_READS_ONE_PASS=0
#this parameter is useful if you have too many Illumina jumping library mates. Typically set it to 60 for bacteria and 300 for the other organisms
LIMIT_JUMP_COVERAGE = 300
#these are the additional parameters to Celera Assembler. do not worry about performance, number of processors or batch sizes -- these are
computed automatically.
#set cgwErrorRate=0.25 for bacteria and 0.1<=cgwErrorRate<=0.15 for other organisms.
CA_PARAMETERS= cgwErrorRate=0.15
#minimum count k-mers used in error correction 1 means all k-mers are used. one can increase to 2 if Illumina coverage >100
```

```
KMER_COUNT_THRESHOLD = 1
#whether to attempt to close gaps in scaffolds with Illumina data
CLOSE_GAPS=1
#auto-detected number of cpus to use
NUM_THREADS = 99
#this is mandatory jellyfish hash size -- a safe value is estimated_genome_size*estimated_coverage
JF_SIZE = 10500000000
#set this to 1 to use SOAPdenovo contigging/scaffolding module. Assembly will be worse but will run faster. Useful for very large (>5Gbp) genomes
from Illumina-only data
SOAP_ASSEMBLY=0
END
```

Main areas to pay attention to are highlighted. Everything that started with '#' is ignored – just like R script. This can be copied and pasted into Notepad++ for editing and requires to be in your set file location to run.

So once the PE (paired end) locations for the Illumina sequence data (short reads) is correct, along with the nanopore (long reads). You can go ahead with assembly command.

masurca Aphid_Nasonovia_ribisnigri_Kent_CL_1.txt

providing no errors, this will create an 'assemble' file in your folder location. To execute command:

./assemble

this will take a day or two depending on genome size.
once finished, you can run some **summary statistics** on the genome.
change folder location to that of where the genome is now located.

cd

/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/Novogene_assembly/N_ribisnigri_assembly_R_Novogene/Assembly/Kent_CL_1/CA.mr.41.15.15.02

ll or **ls #** to make sure your genome is there

#CEGMA

```
export PERL5LIB=$PERL5LIB:/home/data/bioinf_resources/CEGMA/CEGMA_v2.5/lib/  
/home/data/bioinf_resources/scripts/Perl_Rob/N50.pl final.genome.scf.fasta
```

#if happy with results (Less scaffolds the better and N50 above 100,000 is good) you can continue to contaminant removal

#before removing contaminants, it is a good idea to rename your scaffolds to avoid annoyance/confusion later

```
awk '/^>/{print ">Scaffold_" ++i; next}{print}' < Kent_CL_1_final.updated.18.11.19.fasta >  
Kent_CL_1_final.updated.18.11.19.scaffold_names.fasta.fa
```

We can use BUSCO to assess the 'completeness' of the genome

#MAKE SURE TO **dos2unix** YOUR FASTA DATA WHEN EXPORTING FROM GENEIOUS OR THE PROTEINS WON'T WORK!

```
export
```

```
AUGUSTUS_CONFIG_PATH="/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/Novogene_assembly/Busco/config"
```

```
source /home/data/bioinf_resources/programming_tools/busco/my36project/bin/activate
```

```
/home/data/bioinf_resources/programming_tools/busco-master/scripts/run_BUSCO.py -i  
/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/Novogene_assembly/N_ribisnigri_assembly_R_Novogene/Assembly/Kent_CL_1/CA.mr.41.15.15.0.02/  
Kent_CL_1_final.updated.18.11.19.fasta -o BUSCO_final18.11.19 -l  
/home/data/bioinf_resources/programming_tools/BUSCO_v1.1b1/insecta_odb9/ -m genome  
-c 20 -sp fly
```

This produces a nice easy to interpret output as a .txt file. E.g. C:92.7% [S:88.0%,D:4.7%],F:2.2%,M:5.1%,n:1658 – So **92.7%** complete – which isn't too bad! **Takes ~ 40 minutes to run**

#Now we need to use Decypher to generate a .txt file which it generates from running the genome against the NCBI database of all known nucleotide sequences.

Adds a database to DeCypher is needed

```
dc_run -parameters format_aa_into_aa -query Kent_CL_1_final.updated.18.11.19.fasta -  
database test_db -description "Insect"
```

show database present

dc_show -database a

#Run Decypher using raw DNA reads against database. Remember you are using nt and not nr!

#For future reference, if you run the command 'dc_show -database all' on lily, it gives you the list of all available databases you can use and classifies them by AA / NT

BEFORE YOU RUN: Make sure the **blast2go-tera-blastn** file is in the same folder as your genome and where you are running decypher or IT WILL NOT WORK!

dc_run -parameters blast2go-tera-blastn -query Kent_CL_1_final.updated.18.11.19.fasta -database NCBI_nt_010519 -description "test" > decypher.txt

this takes 2 days to run – which is not surprising!

Now open MEGAN

>File

>Import from BLAST

First box 'specify the BLAST'

folder open and find the decypher.txt file

Apply

decypher.txt output – highlighting the majority of 'hits' are situated in the Opisthokonta (aphids) which is good but there are some proteobacteria and viridiplantae that require attention and removing.

#Now open Geneious (software centre) alongside MEGAN (with loaded decypher.txt)

File

Import from file

find your genome and import

import sequences in a list for the scaffolds

Decypher output breakdown of the viridiplantae showing some lettuce contamination that be be removed in Geneious. Look for the right scaffold in Geneious and edit away – make sure to blast some sequences if they look dubious.

Once finished removing contaminants in Geneious, export edited genome

> File
> Export
> Selected documents # Make sure to save as a .FASTA

now we can run BUSCO again to check for completeness and whether removal has improved or got worse. If much worse, consider going back and see what was removed and

Before we run it again, make sure to remove annoying windows characters that exporting from Geneious creates – or BUSCO will not like it and will stop running on the ‘predicting proteins bit!’!

Command line:

dos2unix Kent_CL_1_final.updated.fasta

Now we can run Decypher again to confirm whether all contaminants have been removed.
This will then again need to go through the Geneious process if any contaminants are still present.

export from Geneious again – remembering to **dos2unix** before running BUSCO

Once there are no contaminants remaining and are happy with completeness etc. – we can go onto polishing.

Polishing – Galaxy and Pilon

Galaxy

Open Galaxy online

Import original Illumina forward and reverse PE reads (fasta format)

Note: This needs to be .fasta format and not gzipped fasta (GZ file). If it is, use **gunzip forward/reverse.fq (IN COMMAND LINE)**

> Get Data

> Upload File

drag and drop or ‘choose file/s’

Now we need the newly *de novo* assembled genome (contaminant free)

Import the same way as Illumina data

> NGS: Mapping

> Map with BWA

Use a genome from history and build index –

```
> Select forward.fasta
> Select reverse.fasta
Execute
```

This will create a BAM file that can now be exported into the genome folder

```
> Download #both formats
> Download dataset (.bam) # rename
> Download bam_index (.bai) # rename
```

Pilon

```
#to convert .fq to .fasta
sed -n '1~4s/^@/>/p;2~4p' in.fastq > out.fasta
```

```
#Now in the command line
#Change directory to genome location and
```

```
export JRE_HOME=/home/data/bioinf_resources/java/bin
export JAVA_HOME=/home/data/bioinf_resources/java/bin/
export PATH=/home/data/bioinf_resources/java/bin/:$PATH
```

```
java -Xmx54G -jar pilon-1.23.jar --genome.fasta --frags out_sort.bam --fix gaps,bases --changes
--output pilon_corrected
```

NOTE: If you get out.of.memory error: INCREASE memory e.g. -Xmx54G to -Xmx108G (G = GB)

```
java.lang.illegalargumentexception invalid reference index -1
```

going to try and run it in the older version of pilon (1.22) and see if this makes a difference.

Worked 😊 No improvement to genome though! still 92.4% complete.

How to use Pilon in command line # USED THIS AS IT'S LESS FRUSTRATING!

```
dos2unix Dionv3.fasta #only need to use this if exported from Geneious but doesn't do any
harm making sure no annoying characters have slipped into the genome!
```

```
bwa index Dionv3.fasta
```

```
bwa mem -t 40 Dionv3.fasta Kent_CL_1_1.fq.gz Kent_CL_1_2.fq.gz > out.sam
```

```
samtools view -bS out.sam > out.bam
```

```
/home/data/bioinf_resources/programming_tools/samtools-1.8/samtools sort -@ 30 -o  
out_sort.bam out.bam
```

```
samtools index out_sort.bam
```

```
java -Xmx108G -jar pilon-1.23.jar --genome Dionv3.fasta --frags out_sort.bam --fix gaps,bases -  
-changes --output pilon_corrected
```

#Run BUSCO to check completeness – below is code just for program and pathway. You may to need to run the export codes above to map pathway.

```
/home/data/bioinf_resources/programming_tools/busco-master/scripts/run_BUSCO.py -i  
pilon_corrected.fasta -o pilon_correctedv2 -l  
/home/data/bioinf_resources/programming_tools/BUSCO_v1.1b1/insecta_odb9/ -m genome  
-c 20 -sp fly
```

9.1.1 Annotation of the *N. ribisnigri* genomes

So now we have a pilon-corrected genome and all contaminants have been removed, we can now start with annotating!

Run repeatMasker (on Galaxy) to screen DNA sequences for interspersed repeats and low complexity DNA sequences. Output is a detailed annotation of the repeats that are present in the query sequence as well as a modified version of the query sequence in which all the annotated repeats have been masked (default: replaced by Ns).

RepeatMasker doesn't seem to work on Galaxy atm (surprise surprise) – running on command line instead:

```
#I am using this code to generate repeatmasker file
RepeatMasker -engine ncbi -species sternorrhyncha genome.fasta
#note: make sure to reduce file name to a minimum for it to work – current limitation of 50
characters! Also change scaffold names! Code is on page 3.
```

Now using Blast2Go for annotation (software centre), seems to be going okay atm.

Nice guide to annotation on BIOSTARS: <https://www.biostars.org/p/261203/>

Galaxy

Make sure to change scaffold names (page 3)

Import newly masked genome into Galaxy EU

Can then use MAKER to annotate genome.

Settings I used are as follows:

Organism type: Eukaryotic

re-annotate using existing MAKER annotation: No (if it's the first run) Yes (and add the previous gff file).

Infer gene predictions directly from all ESTs: No

ESTs or assembled cDNA: Transcriptome (all RNA-seq data)

Infer gene predictions directly from all protein alignments: No

Protein sequences: Myzus persicae proteins

Ab-initio gene prediction

Prediction with AUGUSTUS: run with a predefined prediction model (*Acyrtosiphon pisum*)

Run ab-initio prediction programs on unmasked sequence: No
Repeat Masking: Disable repeat masking (not recommended) – The genome has already been masked and therefore redundant.

Leave everything else default

Execute. Will take a good few days to run – even though it has been split into 10!

Once annotation has been completed, export all annotations and merge (concatenate) them using cat tool in lily.

Import file back into Galaxy and re-run MAKER using the previous annotation (.gff) to see if the annotation can be improved.

Once finished load the annotations, along with the genome in IGV to visualise the genes etc.

Re-run annotation if necessary to see if further improvements can be made but they are likely to only be minor now.

OmicsBox

Upload annotation into OmicsBox and see how many CDS, exons, genes, mRNA, 3'/5' UTRs are present and export in a table format. E.g. the Nr_8_123 genome had a list of 408,000 for everything, with 38,000 genes. A lot of these genes will be crap predictions and rubbish (with no evidence) and it is expected that there are ~16,000 genes for insects.

Now to extract the proteins from the annotations.

PuTTY

In lily/pear:

```
Gffread maker.gff -g genome.fasta -y protein.fa
```

This creates a .fai and protein.fa file and extracts all proteins from the annotation.gff

Then:

```
sed -i 's/\.$//g' protein.fa
```

not certain what this does but likely removes weird characters and reformats so it can be read by OmicsBox.

Import this lovely protein.fa file into OmicsBox and run InterproScan to identify the CYP genes and which ones have no evidence.

Now to remove all genes that have just one transcript, missing start/stop codon (even if have full length transcript)

#on lily

```
gffread annotation.gff -g genome.fasta -x cds.fa
```

#get fasta sequences on one line as below or import and export of geneious and untick 80 char option

```
/home/data/bioinf_resources/scripts/Perl_Rob/fa2oneline.pl cds.fa >cds2.fa
```

```
cp /home/data/bioinf_resources/scripts/Perl_Rob/find_sequence_no_stop.pl ./
```

```
cp /home/data/bioinf_resources/scripts/Perl_Rob/find_sequence_no_start.pl ./
```

```
sed -i 's/eva\.tran\.fasta/cds2\.fa/g' find_sequence_no_start.pl
```

```

sed -i 's/eva\.tran\.fasta/cds2\.fa/g' find_sequence_no_stop.pl

#no start fasta file

./find_sequence_no_start.pl > start.fa

#no stop fasta file

./find_sequence_no_stop.pl > stop.fa

grep '>' start.fa >start2.txt

grep '>' stop.fa >stop2.txt

sed 's/-mRNA-.*-/mRNA-/g' start2.txt > 1.txt

sed 's/-mRNA-.*;/Na/g' start2.txt > 2.txt

sed 's/-mRNA-.*-/mRNA-/g' stop2.txt > 3.txt

sed 's/-mRNA-.*;/Na/g' stop2.txt > 4.txt

cat 1.txt 2.txt 3.txt 4.txt > all.txt

sed -i 's/>//g' all.txt

cat 1.txt 2.txt 3.txt 4.txt > all.txt

sed -i 's/>//g' all.txt

# note this will remove all genes that have just one transcript missing a start and stop even if
have a full length transcript.

grep -v -f all.txt annotation.gff > annotation2.gff

```

9.1.2 Constructing the transcriptome using RNA-Seq data

Essentially you need to map the RNA-Seq data onto the constructed genome (in this case Nr_8_123). The RNA-Seq data from all conditions were concatenated and merged into 2 files (forward and reverse). Need to use Hisat2 to do this.

If you're not on the latest version of Java:

```
export JRE_HOME=/home/data/bioinf_resources/java/bin
export JAVA_HOME=/home/data/bioinf_resources/java/bin/
export PATH=/home/data/bioinf_resources/java/bin/:$PATH
```

In the command line:

```
#map RNA-Seq using hisat
/home/data/bioinf_resources/programming_tools/hisat2-2.0.5/hisat2-build genome.fa
genome
/home/data/bioinf_resources/programming_tools/hisat2-2.0.5/hisat2 -x genome -p 40 -1
/home/data/pest_genomics/CSFB/CSFB_RNAseq/data/FASTQ/C89HYANXX/R1_all.fastq
.gz -2
/home/data/pest_genomics/CSFB/CSFB_RNAseq/data/FASTQ/C89HYANXX/R2_all.fastq
.gz > RNA_CSFB627all_stringtie_main.sam
samtools view -F 4 -bS SPv23.sam > SPv23.bam
/home/data/bioinf_resources/programming_tools/samtools-1.8/bin/samtools sort -o
SPv23.sort.bam SPv23.bam
```

We then use Trinity to *de novo* assembly the transcriptome.

```
#trinity
export PATH=/home/data/bioinf_resources/programming_tools/bowtie2-2.3.4.1-linux-
x86_64:$PATH
/home/data/bioinf_resources/programming_tools/trinityrnaseq-
Trinity-v2.5.1/Trinity --max_memory 50G --inchworm_cpu 2 --CPU 12
--bflyGCThreads 2 --bflyCPU 10 --output Trinity_Nr_8_123 --
genome_guided_bam SPv23.sort.bam SPv23.bam --genome_guided_max_intron 20000
```

```
/home/data/bioinf_resources/programming_tools/hisat2-2.0.5/hisat2-build
Nr_8_123.final.pilon_corrected.masked.03.03.20.fasta genome
/home/data/bioinf_resources/programming_tools/hisat2-2.0.5/hisat2 -x genome -p 40 -1
/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/N_ribisnigri_ass
```

```
embly_sequences/SeqData_RNA/RNA_with_aphid_IDs_conditions/Nr_8_008_1_RNA.fastq -2  
/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/N_ribisnigri_ass  
embly_sequences/SeqData_RNA/RNA_with_aphid_IDs_conditions/Nr_8_008_2_RNA.fastq -2 >  
RNA_Nr_8_123_all_stringtie_main.sam
```

#add this onto the end and specify folder for the tmp storage to work in.

--java_opts -

Djava.io.tmpdir="/home/data/aphid_s2156/Rothamsted_Research/Laboratory_work/Sequences/N_ribisnigri_assembly_sequences/Hisat2"

#also for memory issues, add this on the end:

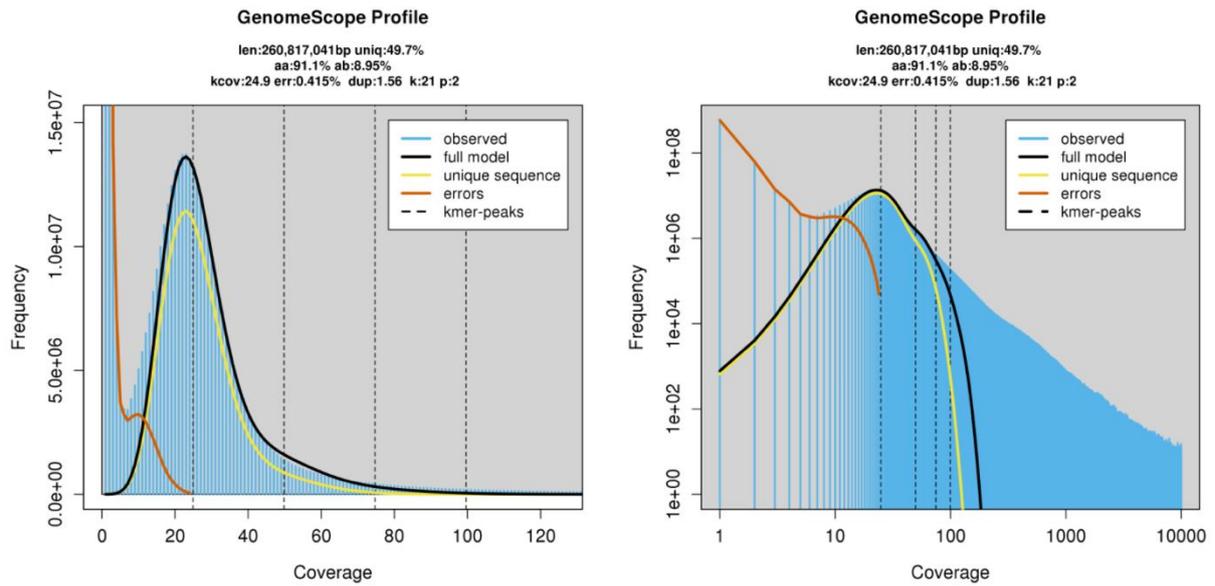
--max_memory 200G --inchworm_cpu 6 --CPU 16 --bflyCalculateCPU

X:\

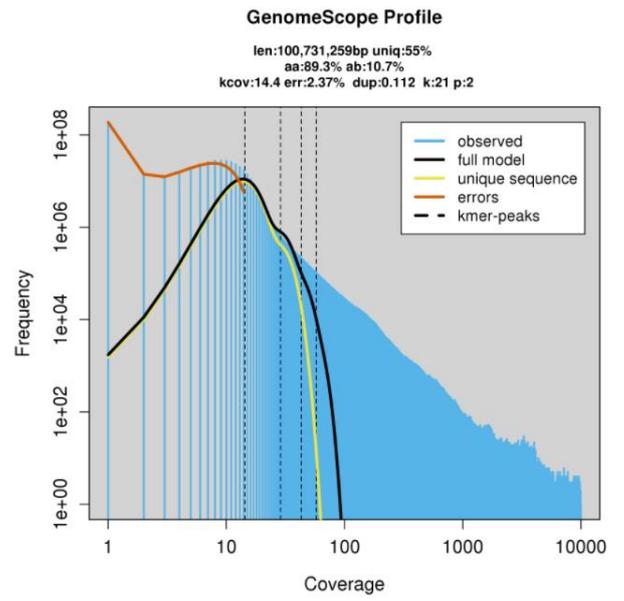
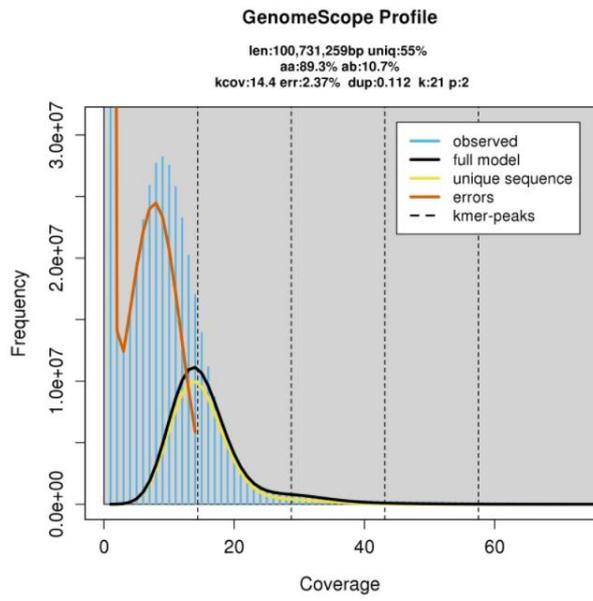
There are generally two possible routes for comparing the transcript between samples, or genomes in this case.

9.2 Supplementary 2: k-mer graphs for the *N. ribisnigri* genomes

Wild-type *N. ribisnigri* (Nr8_123) k-mer plots (Illumina data only)



GenomeScope v2.0 profile plots of A: a transformed linear plot of k=mer frequency and B: a transformed log plot of k-mer coverage, at a k-mer length of 21 and a maximum k-mer coverage of 10000.



Resistant *N. ribisnigri* k-mer plots (Illumina data only)

9.3 Supplementary 3: RNA-seq analysis using DeSeq2 in RStudio

RNA-seq analysis used herein was adapted from the course material provided by Rothamsted Research bioinformatics team - Planning & analysing an RNA-seq experiment - 2018). Cheers guys.

Data available on request.

DE - differentially expressed.

Required libraries

```
library(DESeq2)
library(ggplot2)
library(pheatmap)
```

Working directory

```
setwd("//salt/aphid_s2156/RRes/Rothamsted_Research/Laboratory_work/Sequences/Novogene_assembly/RNA-seq_analysis")
getwd()
## [1] "\\salt/aphid_s2156/RRes/Rothamsted_Research/Laboratory_work/Sequences/Novogene_assembly/RNA-seq_analysis"
```

Import count matrix

```
data <- read.csv("combined_feature_counts_table_ALL2.tab", row.names=1, header=TRUE, stringsAsFactors=FALSE, sep='\t',
               colClasses=c("character",rep("numeric",9)))
```

Header

```
head(data) ; dim(data)
##           SU1 SU2 SU3 SU4 SU5 RE1 RE2 RE3 RE4
## augustus-Scaffold_1-processed-gene-0.0  0  0  0  0  0  0  0  0  0
## augustus-Scaffold_1-processed-gene-1.0  0  0  0  0  0  0  0  0  0
## augustus-Scaffold_1-processed-gene-1.1  0  0  0  0  0  0  0  0  0
## augustus-Scaffold_1-processed-gene-1.7  0  0  0  0  0  0  0  0  0
## augustus-Scaffold_1-processed-gene-1.2 20  7  8  1 25  2  7  1  3
## augustus-Scaffold_1-processed-gene-3.0 25  8  6  0 33  3  7  7  4
```

⌄⌄⌄

```
## [1] 38389 9
```

Assigning susceptible and resistant *N. ribisnigri* biotype columns

```
SUS <-1:5 ; RES <-6:9  
data <-data[,c(SUS, RES)]  
colnames(data)<- c("SUS1","SUS2","SUS3","SUS4","SUS5","RES1","RES2","RES3","RES4")
```

Checking to make sure its worked

```
head(data) ; dim(data)  
##           SUS1 SUS2 SUS3 SUS4 SUS5 RES1 RES2 RES3  
## augustus-Scaffold_1-processed-gene-0.0 0 0 0 0 0 0 0 0  
## augustus-Scaffold_1-processed-gene-1.0 0 0 0 0 0 0 0 0  
## augustus-Scaffold_1-processed-gene-1.1 0 0 0 0 0 0 0 0  
## augustus-Scaffold_1-processed-gene-1.7 0 0 0 0 0 0 0 0  
## augustus-Scaffold_1-processed-gene-1.2 20 7 8 1 25 2 7 1  
## augustus-Scaffold_1-processed-gene-3.0 25 8 6 0 33 3 7 7  
##           RES4  
## augustus-Scaffold_1-processed-gene-0.0 0  
## augustus-Scaffold_1-processed-gene-1.0 0  
## augustus-Scaffold_1-processed-gene-1.1 0  
## augustus-Scaffold_1-processed-gene-1.7 0  
## augustus-Scaffold_1-processed-gene-1.2 3  
## augustus-Scaffold_1-processed-gene-3.0 4  
## [1] 38389 9
```

Making a dataframe that describes the data

```
rbtype <- c(rep("SUS",5),rep("RES",4))  
colData <- data.frame(rbtype, row.names=colnames(data))
```

Checking data

```
colData
```

```
##  rbtype  
## SUS1  SUS  
## SUS2  SUS  
## SUS3  SUS  
## SUS4  SUS  
## SUS5  SUS  
## RES1  RES  
## RES2  RES  
## RES3  RES  
## RES4  RES
```

Constructing a DESeq dataset and removing rows with 0 or 1 count in total

```
dds <- DESeqDataSetFromMatrix(countData=data, colData=colData, design=~rbtype)  
dds <- dds[rowSums(counts(dds))>1,]
```

Checking dataset

```
dds  
## class: DESeqDataSet  
## dim: 18872 9  
## metadata(1): version  
## assays(1): counts  
## rownames(18872): augustus-Scaffold_1-processed-gene-1.2  
##  augustus-Scaffold_1-processed-gene-3.0 ...  
##  maker-Scaffold_998-augustus-gene-0.0  
##  augustus-Scaffold_999-processed-gene-0.3  
## rowData names(0):  
## colnames(9): SUS1 SUS2 ... RES3 RES4  
## colData names(1): rbtype
```

Log transform data for ease of viewing

```
rld <- rlog(dds, blind=FALSE)
rld
## class: DESeqTransform
## dim: 18872 9
## metadata(1): version
## assays(1): "
## rownames(18872): augustus-Scaffold_1-processed-gene-1.2
## augustus-Scaffold_1-processed-gene-3.0 ...
## maker-Scaffold_998-augustus-gene-0.0
## augustus-Scaffold_999-processed-gene-0.3
## rowData names(7): baseMean baseVar ... dispFit rlogIntercept
## colnames(9): SUS1 SUS2 ... RES3 RES4
## colData names(2): rdtype sizeFactor
head(assay(rld))
##           SUS1  SUS2  SUS3
## augustus-Scaffold_1-processed-gene-1.2 2.755547 2.7595630 2.7906589
## augustus-Scaffold_1-processed-gene-3.0 3.033816 3.0283387 3.0157852
## augustus-Scaffold_1-processed-gene-3.7 5.418481 5.5533021 5.4976565
## augustus-Scaffold_1-processed-gene-3.6 -0.953468 -0.9259157 -0.9495109
## augustus-Scaffold_1-processed-gene-3.9 -1.801843 -1.8033043 -1.8029351
## maker-Scaffold_1-augustus-gene-3.2    7.965819 7.9562191 7.8821325
##           SUS4  SUS5  RES1
## augustus-Scaffold_1-processed-gene-1.2 2.6677391 2.782008 2.6778780
## augustus-Scaffold_1-processed-gene-3.0 2.9104657 3.073921 2.9446747
## augustus-Scaffold_1-processed-gene-3.7 5.2168133 5.571040 5.2786479
## augustus-Scaffold_1-processed-gene-3.6 -0.9519402 -0.953463 -0.9400544
## augustus-Scaffold_1-processed-gene-3.9 -1.8037713 -1.805309 -1.8039255
## maker-Scaffold_1-augustus-gene-3.2    8.0416578 7.904672 8.2774507
```

502

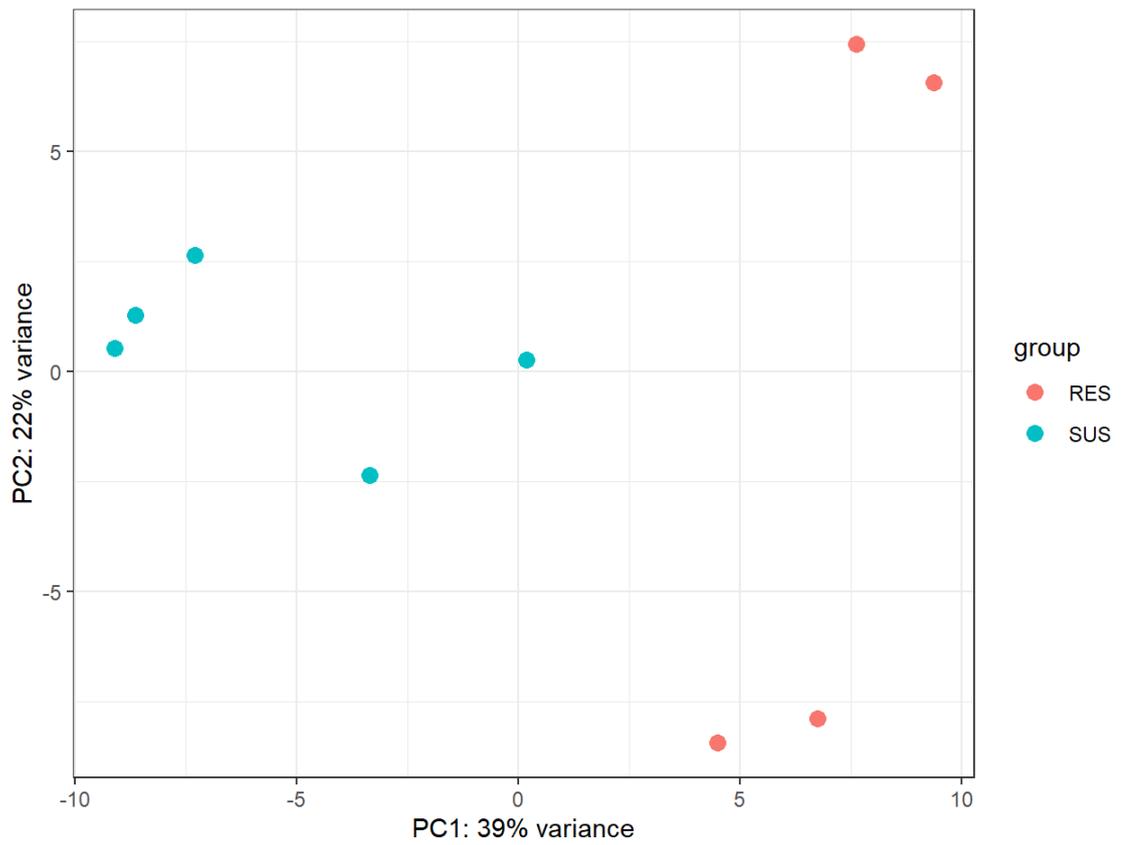
```

##          RES2    RES3    RES4
## augustus-Scaffold_1-processed-gene-1.2 2.7323701 2.6640431 2.701823
## augustus-Scaffold_1-processed-gene-3.0 2.9868966 2.9840345 2.971527
## augustus-Scaffold_1-processed-gene-3.7 5.2488767 5.3186358 5.343757
## augustus-Scaffold_1-processed-gene-3.6 -0.9403563 -0.9526872 -0.950746
## augustus-Scaffold_1-processed-gene-3.9 -1.8040135 -1.7998480 -1.803360
## maker-Scaffold_1-augustus-gene-3.2    8.0900516 8.0803817 8.054818
head(assays(dds)[["counts"]])
##          SUS1 SUS2 SUS3 SUS4 SUS5 RES1 RES2 RES3
## augustus-Scaffold_1-processed-gene-1.2 20  7  8  1 25  2  7  1
## augustus-Scaffold_1-processed-gene-3.0 25  8  6  0 33  3  7  7
## augustus-Scaffold_1-processed-gene-3.7 100 46 34 19 143 27 25 34
## augustus-Scaffold_1-processed-gene-3.6  0  2  0  0  0  1  1  0
## augustus-Scaffold_1-processed-gene-3.9  1  0  0  0  0  0  0  1
## maker-Scaffold_1-augustus-gene-3.2    525 164 124 229 482 330 273 280
##          RES4
## augustus-Scaffold_1-processed-gene-1.2  3
## augustus-Scaffold_1-processed-gene-3.0  4
## augustus-Scaffold_1-processed-gene-3.7 26
## augustus-Scaffold_1-processed-gene-3.6  0
## augustus-Scaffold_1-processed-gene-3.9  0
## maker-Scaffold_1-augustus-gene-3.2   192

```

Using ggplot2 to make PCA

```
plotPCA(rld, intgroup="rbtype") + theme_bw()
```



```
df.PCA <- plotPCA(rld, intgroup="rbtype", returnData=TRUE)
```

```
df.PCA
```

```
##      PC1    PC2 group rbtype name
## SUS1 -8.6451925  1.2837997  SUS  SUS SUS1
## SUS2 -7.2914072  2.6477619  SUS  SUS SUS2
## SUS3 -9.1130319  0.5148459  SUS  SUS SUS3
## SUS4  0.1759198  0.2642074  SUS  SUS SUS4
## SUS5 -3.3585805 -2.3691001  SUS  SUS SUS5
## RES1  6.7496877 -7.8950568  RES  RES RES1
## RES2  4.4838921 -8.4494928  RES  RES RES2
## RES3  9.3650327  6.5695534  RES  RES RES3
```

```
## RES4 7.6336799 7.4334815 RES RES RES4
```

DeSeq2 Analysis First the size factors need to be estimated

```
dds <- estimateSizeFactors(dds)
```

```
dds
```

```
## class: DESeqDataSet
```

```
## dim: 18872 9
```

```
## metadata(1): version
```

```
## assays(1): counts
```

```
## rownames(18872): augustus-Scaffold_1-processed-gene-1.2
```

```
## augustus-Scaffold_1-processed-gene-3.0 ...
```

```
## maker-Scaffold_998-augustus-gene-0.0
```

```
## augustus-Scaffold_999-processed-gene-0.3
```

```
## rowData names(0):
```

```
## colnames(9): SUS1 SUS2 ... RES3 RES4
```

```
## colData names(2): rbtype sizeFactor
```

```
sizeFactors(dds)
```

```
## SUS1 SUS2 SUS3 SUS4 SUS5 RES1 RES2 RES3
```

```
## 2.1762529 0.6917661 0.5813352 0.8619969 2.1620382 0.9269670 0.9661969 1.0033169
```

```
## RES4
```

```
## 0.7102451
```

Secondly, the dispersion is estimated

```
dds <- estimateDispersions(dds)
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
```

```
## final dispersion estimates
```

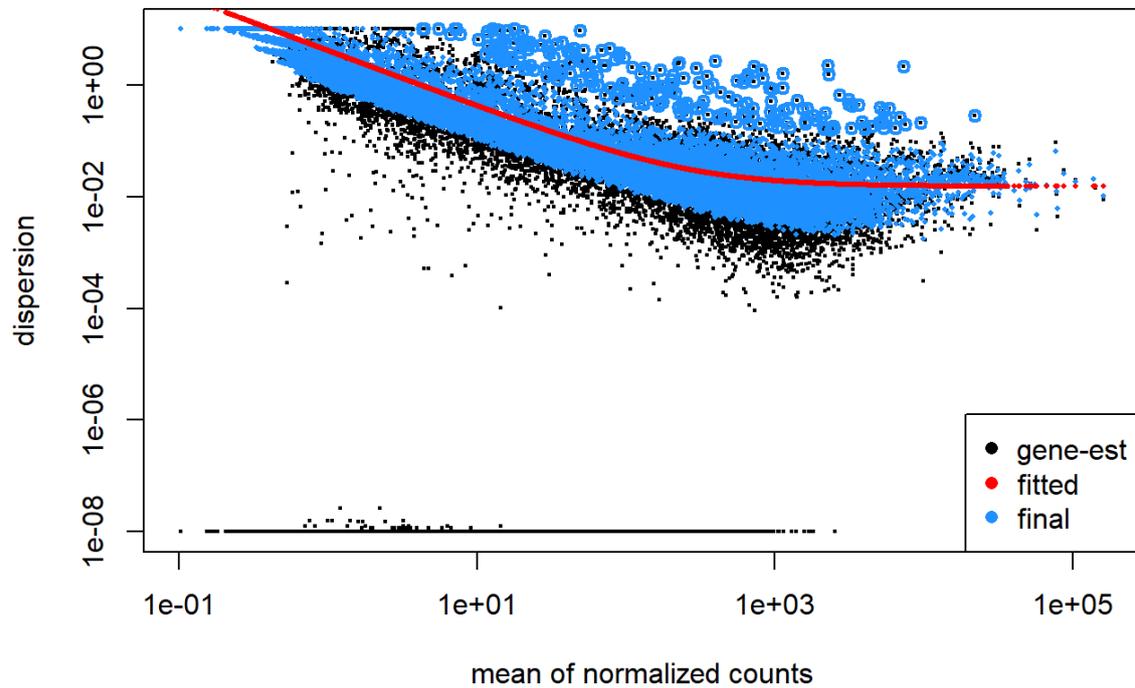
```
dds
```

```
## class: DESeqDataSet
```

```
## dim: 18872 9
## metadata(1): version
## assays(2): counts mu
## rownames(18872): augustus-Scaffold_1-processed-gene-1.2
## augustus-Scaffold_1-processed-gene-3.0 ...
## maker-Scaffold_998-augustus-gene-0.0
## augustus-Scaffold_999-processed-gene-0.3
## rowData names(10): baseMean baseVar ... dispOutlier dispMAP
## colnames(9): SUS1 SUS2 ... RES3 RES4
## colData names(2): rbtype sizeFactor
```

Plotting the dispersion to visualise data

```
plotDispEsts(dds)
```



Now to fit a negative binomial GLM and calculate Wald statistic. The Wald test was incorporated to test for differential expression between resistant-breaking (Nr:1) and susceptible (Nr:0) *N. ribisnigri* biotypes.

Null hypothesis = no differential expression between Nr:1 and Nr:0.

```
dds <- nbinomWaldTest(dds)
dds
## class: DESeqDataSet
## dim: 18872 9
## metadata(1): version
## assays(4): counts mu H cooks
## rownames(18872): augustus-Scaffold_1-processed-gene-1.2
```

```
## augustus-Scaffold_1-processed-gene-3.0 ...
## maker-Scaffold_998-augustus-gene-0.0
## augustus-Scaffold_999-processed-gene-0.3
## rowData names(22): baseMean baseVar ... deviance maxCooks
## colnames(9): SUS1 SUS2 ... RES3 RES4
## colData names(2): rbtype sizeFactor
```

View results

```
resultsNames(dds)
## [1] "Intercept"      "rbtype_SUS_vs_RES"
```

To generate more accurate log₂ fold change estimates, the `apegglm` package was utilised to generate LFC estimates (Zhu et al., 2018). The more recent versions of DESeq2 package does not perform shrinkage of the LFC estimates by default. When information of a gene is low, `apegglm` allows for the shrinkage of the LFC estimates towards zero when the information of a gene is low (i.e. low counts or high dispersion) (Zhu et al., 2018).

Alpha is *P*-value

```
res <- lfcShrink(dds=dds, coef = 2, type = "apeglm")
alpha <- 0.05
```

View results

```
summary(res) ; mcols(res)
##
## out of 18872 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)      : 598, 3.2%
## LFC < 0 (down)    : 554, 2.9%
## outliers [1]     : 60, 0.32%
## low counts [2]   : 6204, 33%
## (mean count < 3)
## [1] see 'cooksCutoff' argument of ?results
```

```
## [2] see 'independentFiltering' argument of ?results
## DataFrame with 5 rows and 2 columns
##           type      description
##           <character> <character>
## baseMean   intermediate mean of normalized c..
## log2FoldChange results log2 fold change (MA..
## lfcSE      results posterior SD: rbtype..
## pvalue     results Wald test p-value: r..
## padj       results BH adjusted p-values
```

For ease of viewing, significantly DE genes are ordered by *P*-value

```
res.ordered <- res[order(res$padj),]
res.ordered ; dim(res.ordered)
## log2 fold change (MAP): rbtype SUS vs RES
## Wald test p-value: rbtype SUS vs RES
## DataFrame with 18872 rows and 5 columns
##           baseMean log2FoldChange lfcSE
##           <numeric> <numeric> <numeric>
## maker-Scaffold_2698-augustus-gene-0.8  655.0535  -3.527614  0.119132
## augustus-Scaffold_463-processed-gene-0.8  481.5691  -2.560592  0.269206
## augustus-Scaffold_4753-processed-gene-1.4  44.6054   2.876173  0.342276
## augustus-Scaffold_784-processed-gene-2.7  22.7633  -2.439800  0.367568
## maker-Scaffold_4214-augustus-gene-0.27  1405.3243  -0.731487  0.121986
## ...           ...           ...           ...
## augustus-Scaffold_995-processed-gene-0.4  0.102448  0.00218592  0.122421
## maker-Scaffold_997-augustus-gene-0.1    1.458628  0.00712607  0.122125
## maker-Scaffold_998-augustus-gene-0.2    0.263067  0.00331658  0.122453
## maker-Scaffold_998-augustus-gene-0.0    0.632388  -0.00505025  0.122404
## augustus-Scaffold_999-processed-gene-0.3  0.312881  -0.00387592  0.122487
```

```

##           pvalue   padj
##           <numeric> <numeric>
## maker-Scaffold_2698-augustus-gene-0.8  2.82053e-195 3.55613e-191
## augustus-Scaffold_463-processed-gene-0.8  1.15147e-22 7.25889e-19
## augustus-Scaffold_4753-processed-gene-1.4  3.70325e-18 1.55635e-14
## augustus-Scaffold_784-processed-gene-2.7  1.91550e-12 6.03766e-09
## maker-Scaffold_4214-augustus-gene-0.27  1.22059e-10 2.88096e-07
## ...           ...     ...
## augustus-Scaffold_995-processed-gene-0.4  0.829947    NA
## maker-Scaffold_997-augustus-gene-0.1      0.561868    NA
## maker-Scaffold_998-augustus-gene-0.2      0.682352    NA
## maker-Scaffold_998-augustus-gene-0.0      0.491657    NA
## augustus-Scaffold_999-processed-gene-0.3  0.560801    NA
## [1] 18872  5

```

```
table(res.ordered$padj < alpha)
```

```

##
## FALSE TRUE
## 11919 689

```

Removing genes which DeSeq2 has filtered out

```
res.filtered <- res.ordered[!is.na(res.ordered$padj),]
res.filtered ; dim(res.filtered)
```

```

## log2 fold change (MAP): rbtype SUS vs RES
## Wald test p-value: rbtype SUS vs RES
## DataFrame with 12608 rows and 5 columns
##           baseMean log2FoldChange  lfcSE
##           <numeric> <numeric> <numeric>
## maker-Scaffold_2698-augustus-gene-0.8  655.0535  -3.527614 0.119132
## augustus-Scaffold_463-processed-gene-0.8  481.5691  -2.560592 0.269206

```

```

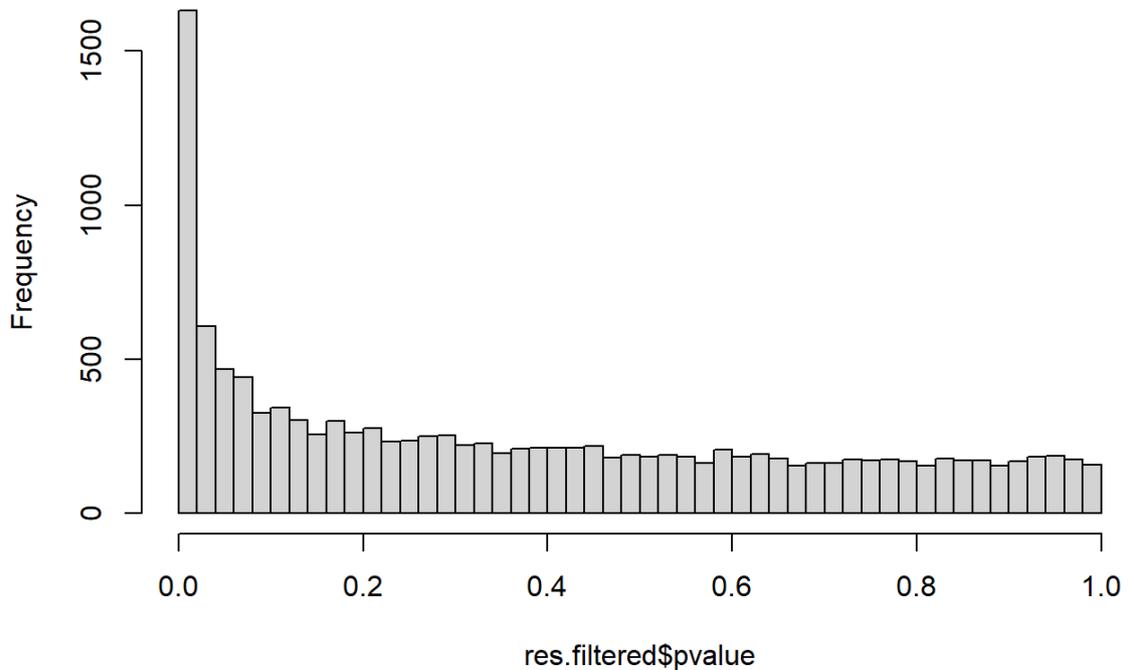
## augustus-Scaffold_4753-processed-gene-1.4 44.6054 2.876173 0.342276
## augustus-Scaffold_784-processed-gene-2.7 22.7633 -2.439800 0.367568
## maker-Scaffold_4214-augustus-gene-0.27 1405.3243 -0.731487 0.121986
## ...
## maker-Scaffold_4257-augustus-gene-4.3 67.5184 9.23453e-05 0.1082653
## augustus-Scaffold_10-processed-gene-0.3 508.1105 -8.58747e-05 0.0986863
## maker-Scaffold_475-augustus-gene-7.6 303.0801 1.10025e-04 0.0785460
## maker-Scaffold_650-augustus-gene-1.4 1675.3757 1.10061e-04 0.0686087
## maker-Scaffold_772-augustus-gene-1.27 3221.5519 -1.56398e-04 0.1008540
##
## pvalue padj
## <numeric> <numeric>
## maker-Scaffold_2698-augustus-gene-0.8 2.82053e-195 3.55613e-191
## augustus-Scaffold_463-processed-gene-0.8 1.15147e-22 7.25889e-19
## augustus-Scaffold_4753-processed-gene-1.4 3.70325e-18 1.55635e-14
## augustus-Scaffold_784-processed-gene-2.7 1.91550e-12 6.03766e-09
## maker-Scaffold_4214-augustus-gene-0.27 1.22059e-10 2.88096e-07
## ...
## maker-Scaffold_4257-augustus-gene-4.3 0.999209 0.999548
## augustus-Scaffold_10-processed-gene-0.3 0.999517 0.999755
## maker-Scaffold_475-augustus-gene-7.6 0.999910 0.999989
## maker-Scaffold_650-augustus-gene-1.4 0.999989 0.999989
## maker-Scaffold_772-augustus-gene-1.27 0.999889 0.999989
## [1] 12608 5
table(res.filtered$padj < alpha)
##
## FALSE TRUE
## 11919 689

```

Plot of adjusted P values

```
hist(res.filtered$pvalue, main="Frequencies of pvalues", breaks=50)
```

Frequencies of pvalues



Significance abundance plot (SA) for both filtered and unfiltered data

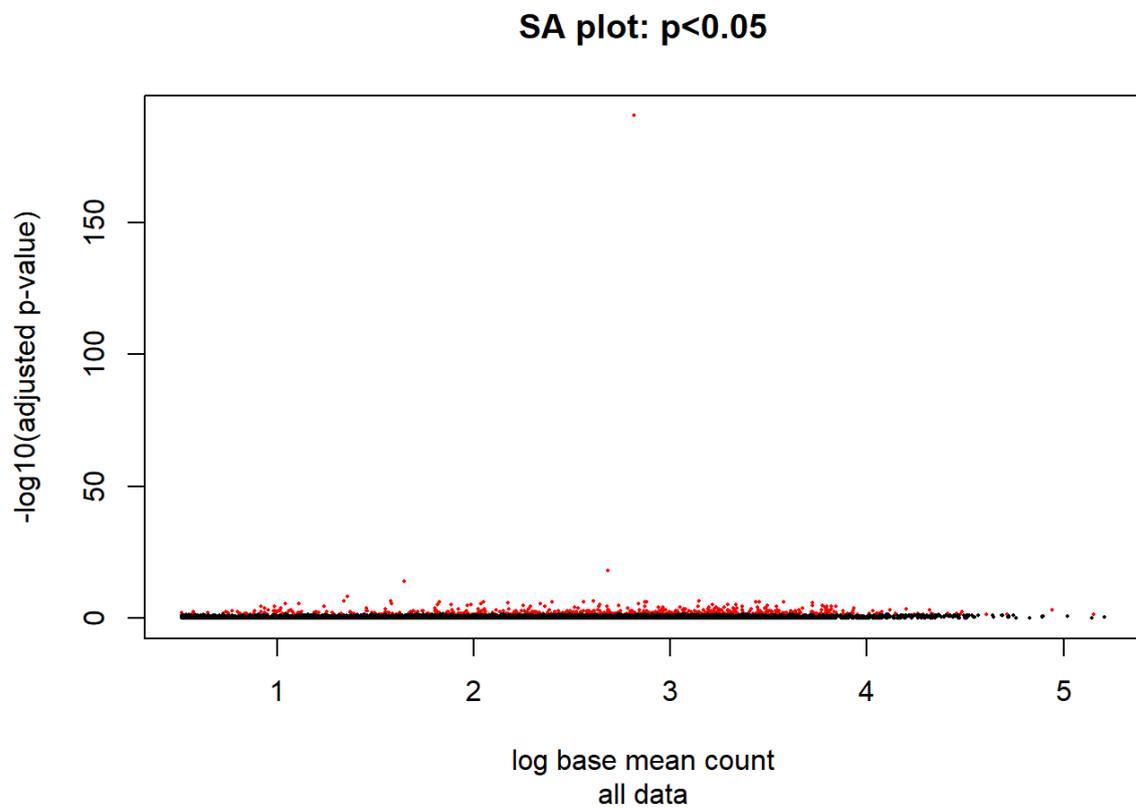
Used this to create SAplot2

```
SAplot.2 <- function(res, alpha=0.05,...){  
  #  
  # SA plot for DESEQ2 results  
  #  
  data <- data.frame(gene=row.names(res), pvalue=-log10(res$padj), bm=log10(res$baseMean))  
  data <- na.omit(data)  
  data <- transform(data, colour=ifelse(pvalue>-log10(alpha), "red", "black"))  
  #head(data) ; dim(data)
```

```
#  
title <- paste("SA plot: p<", alpha, sep="")  
plot(data$bm, data$pvalue, main=title, pch=16, cex=0.3, col=data$colour,  
      xlab="log base mean count", ylab="-log10(adjusted p-value)", ...)  
return()  
}
```

SAplot

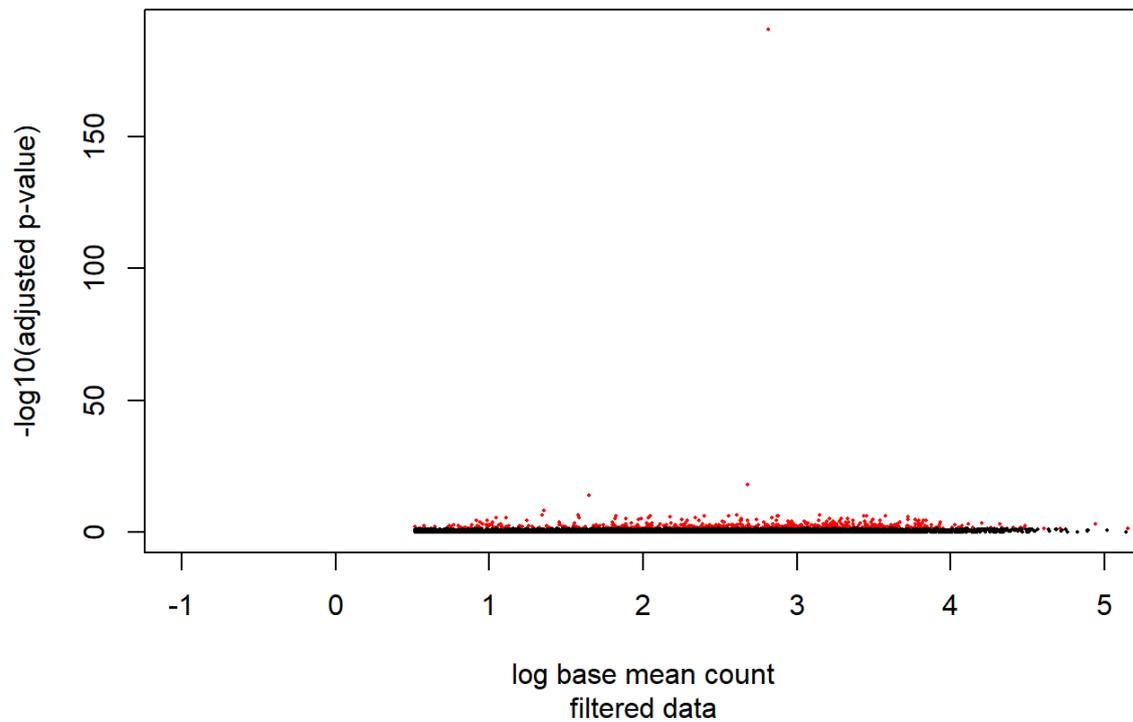
```
alpha <- 0.05  
SAplot.2(res.ordered, alpha=alpha, sub="all data")
```



```
## NULL
```

```
SAPlot.2(res.filtered, alpha=alpha, xlim=c(-1,5), sub="filtered data")
```

SA plot: $p < 0.05$



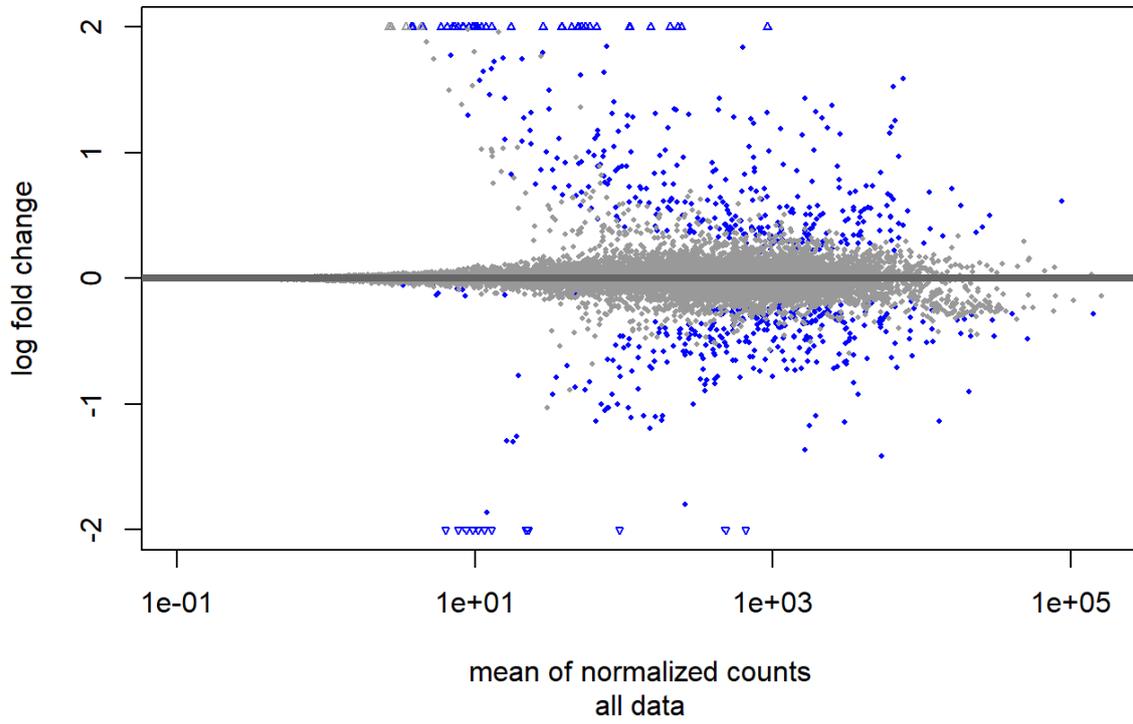
```
## NULL
```

MA plot: log ratio(M) vs. mean average (A)

```
title <- paste("MA plot: p<", alpha, sep="")
```

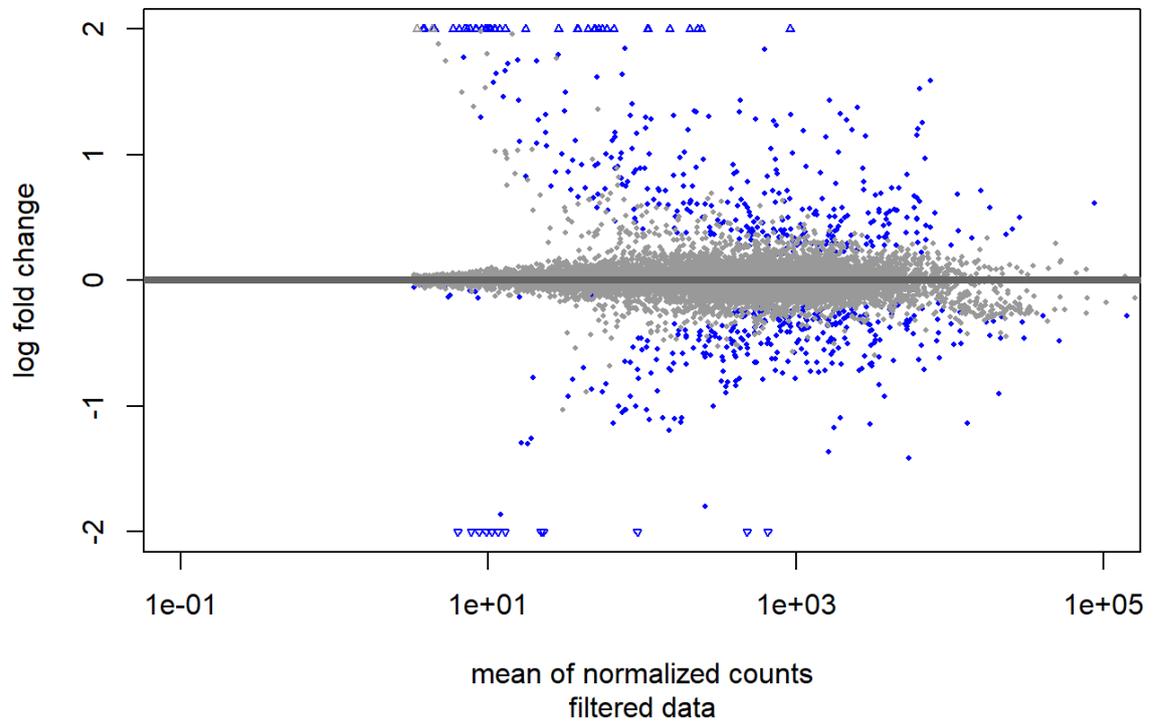
```
plotMA(res.ordered, ylim=c(-2,2), alpha=alpha, main=title, sub="all data")
```

MA plot: $p < 0.05$



```
plotMA(res.filtered, ylim=c(-2,2), xlim=c(0.1,100000), alpha=alpha, main=title, sub="filtered data")
```

MA plot: $p < 0.05$



Visualising DE genes using a volcano plot

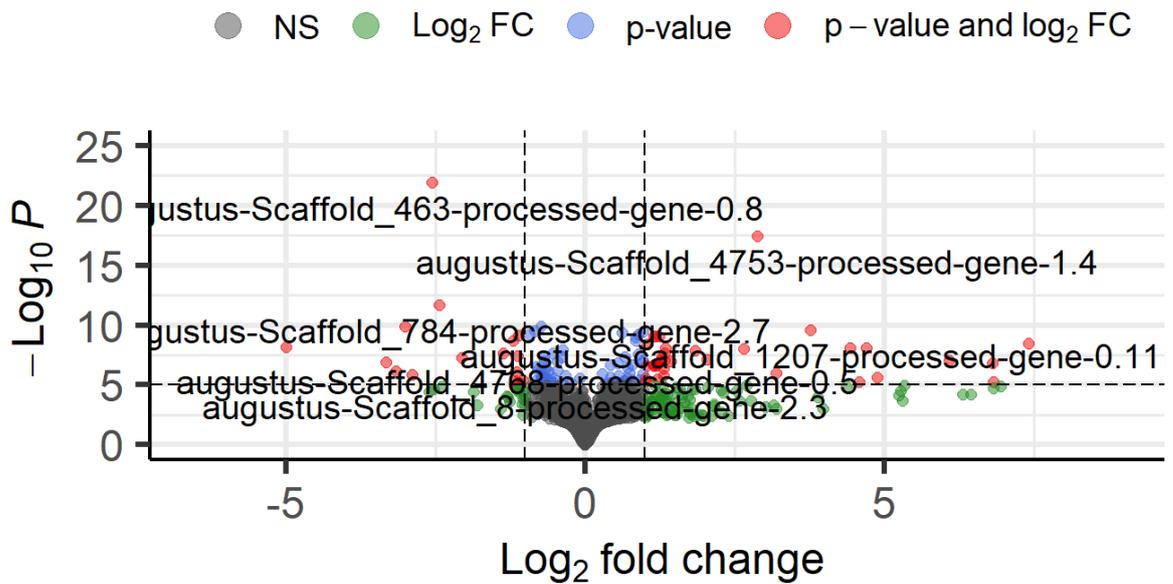
```
library(EnhancedVolcano)
## Loading required package: ggrepel
## Registered S3 methods overwritten by 'ggalt':
## method      from
## grid.draw.absoluteGrob ggplot2
## grobHeight.absoluteGrob ggplot2
## grobWidth.absoluteGrob ggplot2
## grobX.absoluteGrob    ggplot2
## grobY.absoluteGrob    ggplot2
```

Now for the volcano plot

```
EnhancedVolcano(res.filtered,
  lab = rownames(res.filtered),
  x = 'log2FoldChange',
  y = 'pvalue', ylim=c(0,25))
```

Volcano plot

EnhancedVolcano



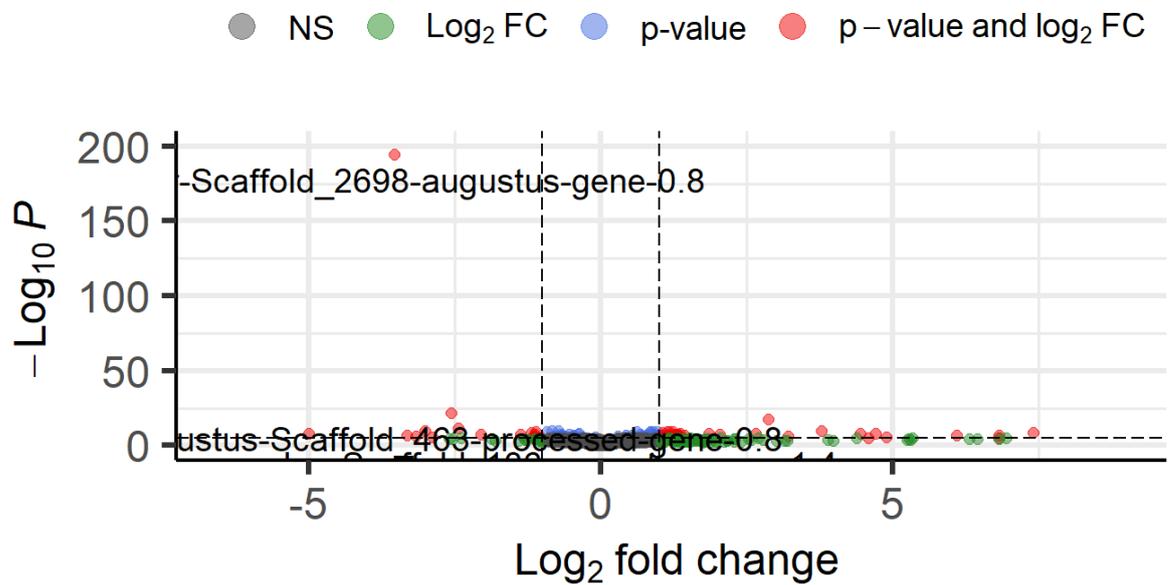
total = 12608 variables

It was found that Y-lim was too strict and cut-off an important DE gene

```
EnhancedVolcano(res.filtered,
  lab = rownames(res.filtered),
  x = 'log2FoldChange',
  y = 'pvalue', ylim=c(0,200))
```

Volcano plot

EnhancedVolcano



total = 12608 variables

Still not a great view, as all other DE genes are compressed. Adapted this code for better visuals and to scale the Y axis to show the DE gene of interest along with the remaining DE genes:

```
volcanoPlot.2 <- function(res, alpha=0.05){  
  #  
  # Volcano plot for DESEQ2 results  
  # based on https://twbattaglia.github.io/2016/12/17/volcano-plot/  
  #  
  library(ggplot2)  
  library(magrittr)  
  library(plyr)  
  library(dplyr)  
  library(ggrepel)
```

```

#
data <- data.frame(gene = row.names(res), pvalue = -log10(res$padj), lfc = res$log2FoldChange)
data <- na.omit(data)
#
# colour significant genes, depending whether up or down regulated)
threshold = -log10(alpha)
data <- data %>%
  mutate(color = ifelse(data$lfc > 0 & data$pvalue > threshold,
    yes = "Up",
    no = ifelse(data$lfc < 0 & data$pvalue > threshold,
      yes = "Down",
      no = "none")))
#
return.plot <- ggplot(data, aes(x = lfc, y = pvalue)) +
  coord_cartesian(xlim = rev(c(-5, 5))) +
  geom_point(aes(color = factor(color)), size = 1.75, alpha = 0.8, na.rm = T) + # add gene points
  theme_bw(base_size = 16) + # clean up theme
  theme(legend.position = "none") + # remove legend
  ggtitle(label = "Volcano Plot", subtitle = paste("p<", alpha, sep="")) + # add title
  xlab(expression(log[2]("Down" / "Up"))) + # x-axis label
  ylab(expression(-log[10]("adjusted p-value"))) + # y-axis label
  geom_vline(xintercept = 0, colour = "black") + # add line at 0
  geom_hline(yintercept = threshold, colour = "black") + # -log10(alpha)
  annotate(geom = "text", label = "Up", x = -2, y = 85, size = 7, colour = "black") + # add Down text
  annotate(geom = "text", label = "Down", x = 2, y = 85, size = 7, colour = "black") + # add Up text
  scale_color_manual(values = c("Down" = "#E64B35", "Up" = "#3182bd", "none" = "#636363")) # change
colors

# Scaled Y-axis with log1p function

```

```

return.plot <- return.plot + scale_y_continuous(trans = "log1p")

# Subset table to show top and bottom gene labels
labelled <- rbind(top_n(data, n = -2, wt = lfc) , top_n(data, n = 0, wt = lfc))

# Add layer of text annotation
return.plot <- return.plot +
  geom_text_repel(data = labelled,
    mapping = aes(label = gene),
    size = 3,
    fontface = 'bold',
    color = 'black',
    box.padding = unit(0.5, "lines"),
    point.padding = unit(0.5, "lines"))
return(return.plot)
}

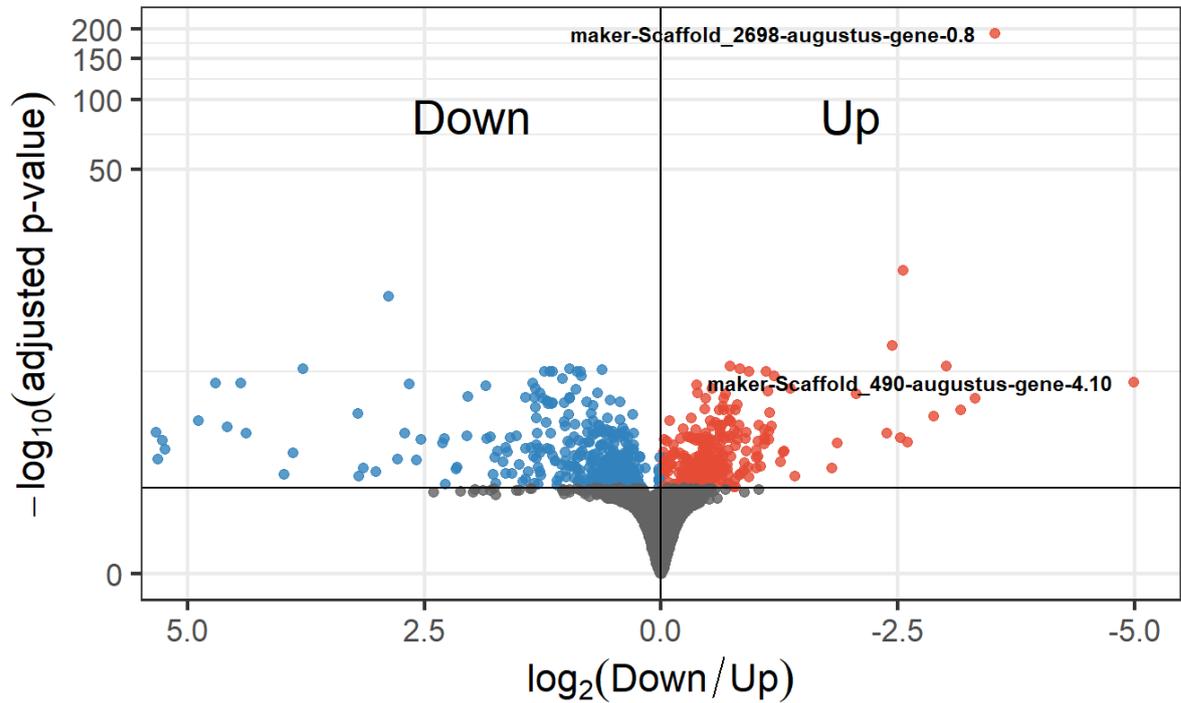
```

Viewing new volcano plot

```
volcanoPlot.2(res.filtered, alpha=alpha)
```

Volcano Plot

$p < 0.05$



A total of 689 DE genes were identified with a P -value of 0.05. A stricter P -value was incorporated to provide more informative results.

```
alpha <- 0.00001
```

Confirming P -value adjustment

```
DEgenes <- rownames(subset(res.filtered, padj < alpha)) ; length(DEgenes)
```

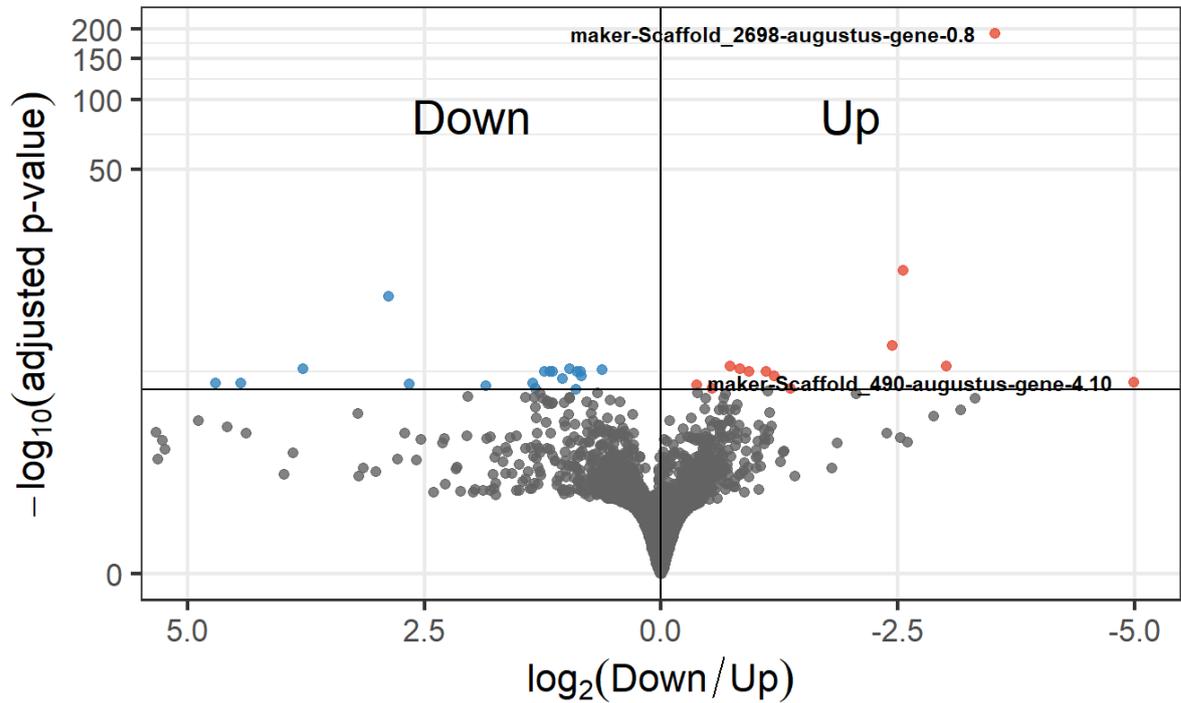
```
## [1] 32
```

New volcano plot with adjusted P -value (0.00001)

```
volcanoPlot.2(res.filtered, alpha=alpha)
```

Volcano Plot

$p < 1e-05$



To plot RNA counts for identified 32 DE genes for both resistant-breaking and susceptible *N. ribisnigri* biotypes. This function lists the top 32 DE genes in order of *P*-value

```
DEgenes <- rownames(subset(res.filtered, padj<alpha)) ; length(DEgenes)
```

```
## [1] 32
```

```
head(DEgenes, n=32)
```

```
## [1] "maker-Scaffold_2698-augustus-gene-0.8"
```

```
## [2] "augustus-Scaffold_463-processed-gene-0.8"
```

```
## [3] "augustus-Scaffold_4753-processed-gene-1.4"
```

```
## [4] "augustus-Scaffold_784-processed-gene-2.7"
```

```
## [5] "maker-Scaffold_4214-augustus-gene-0.27"
```

```
## [6] "maker-Scaffold_910-augustus-gene-0.6"
```

```
## [7] "augustus-Scaffold_1207-processed-gene-0.11"
```

```
## [8] "maker-Scaffold_821-augustus-gene-1.13"  
## [9] "maker-Scaffold_4762-augustus-gene-30.11"  
## [10] "augustus-Scaffold_1052-processed-gene-1.15"  
## [11] "maker-Scaffold_4732-augustus-gene-5.24"  
## [12] "augustus-Scaffold_634-processed-gene-0.9"  
## [13] "augustus-Scaffold_2666-processed-gene-0.1"  
## [14] "maker-Scaffold_200-augustus-gene-0.17"  
## [15] "maker-Scaffold_2061-augustus-gene-1.1"  
## [16] "augustus-Scaffold_2128-processed-gene-0.25"  
## [17] "maker-Scaffold_647-augustus-gene-1.3"  
## [18] "augustus-Scaffold_134-processed-gene-6.11"  
## [19] "augustus-Scaffold_82-processed-gene-1.8"  
## [20] "augustus-Scaffold_1158-processed-gene-0.1"  
## [21] "augustus-Scaffold_2243-processed-gene-0.15"  
## [22] "maker-Scaffold_490-augustus-gene-4.10"  
## [23] "maker-Scaffold_200-augustus-gene-0.18"  
## [24] "maker-Scaffold_922-augustus-gene-0.7"  
## [25] "augustus-Scaffold_1043-processed-gene-0.0"  
## [26] "augustus-Scaffold_2727-processed-gene-0.1"  
## [27] "augustus-Scaffold_1068-processed-gene-4.10"  
## [28] "augustus-Scaffold_4330-processed-gene-1.6"  
## [29] "maker-Scaffold_119-augustus-gene-0.4"  
## [30] "maker-Scaffold_4299-augustus-gene-0.56"  
## [31] "augustus-Scaffold_4700-processed-gene-2.10"  
## [32] "augustus-Scaffold_49-processed-gene-4.9"
```

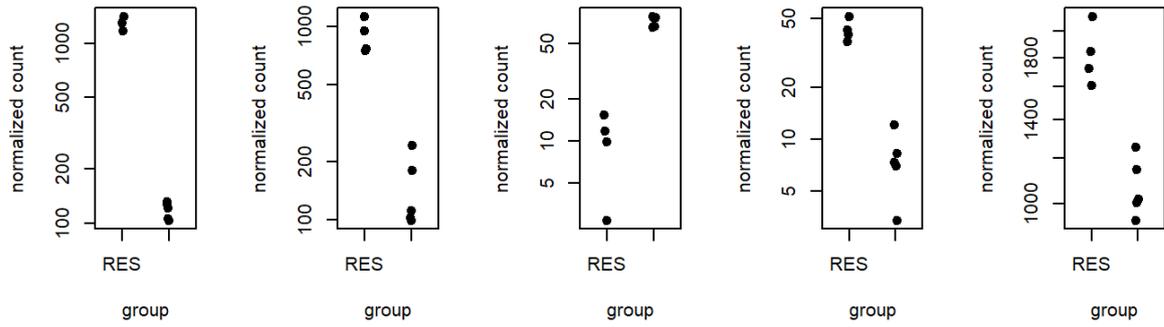
All 32 DE genes were plotted to view RNA count data between both resistant-breaking and susceptible *N. ribisnigri* biotypes

DE genes 1-10

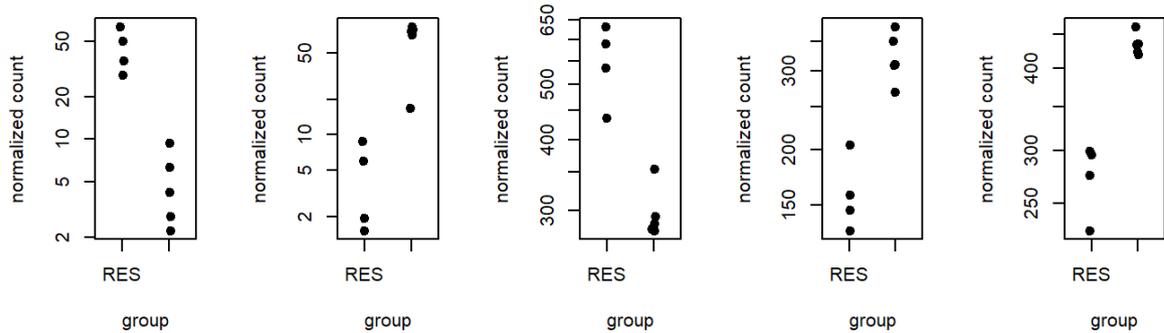
```
par(mfrow = c(2, 5))
```

```
mygene1="maker-Scaffold_2698-augustus-gene-0.8"  
mygene2="augustus-Scaffold_463-processed-gene-0.8"  
mygene3="augustus-Scaffold_4753-processed-gene-1.4"  
mygene4="augustus-Scaffold_784-processed-gene-2.7"  
mygene5="maker-Scaffold_4214-augustus-gene-0.27"  
mygene6="maker-Scaffold_910-augustus-gene-0.6"  
mygene7="augustus-Scaffold_1207-processed-gene-0.11"  
mygene8="maker-Scaffold_821-augustus-gene-1.13"  
mygene9="maker-Scaffold_4762-augustus-gene-30.11"  
mygene10="augustus-Scaffold_1052-processed-gene-1.15"  
plotCounts(dds, gene=mygene1, intgroup="rbtype", pch=19, main=mygene1)  
plotCounts(dds, gene=mygene2, intgroup="rbtype", pch=19, main=mygene2)  
plotCounts(dds, gene=mygene3, intgroup="rbtype", pch=19, main=mygene3)  
plotCounts(dds, gene=mygene4, intgroup="rbtype", pch=19, main=mygene4)  
plotCounts(dds, gene=mygene5, intgroup="rbtype", pch=19, main=mygene5)  
plotCounts(dds, gene=mygene6, intgroup="rbtype", pch=19, main=mygene6)  
plotCounts(dds, gene=mygene7, intgroup="rbtype", pch=19, main=mygene7)  
plotCounts(dds, gene=mygene8, intgroup="rbtype", pch=19, main=mygene8)  
plotCounts(dds, gene=mygene9, intgroup="rbtype", pch=19, main=mygene9)  
plotCounts(dds, gene=mygene10, intgroup="rbtype", pch=19, main=mygene10)
```

Scaffold_2698-augustus-Scaffold_463-process-Scaffold_4753-process-Scaffold_784-procesScaffold_4214-augustu



-Scaffold_910-augustu-Scaffold_1207-procesScaffold_821-augustuscaffold_4762-augustusScaffold_1052-proces:



DE genes 11-20

```
par(mfrow = c(2, 5))
```

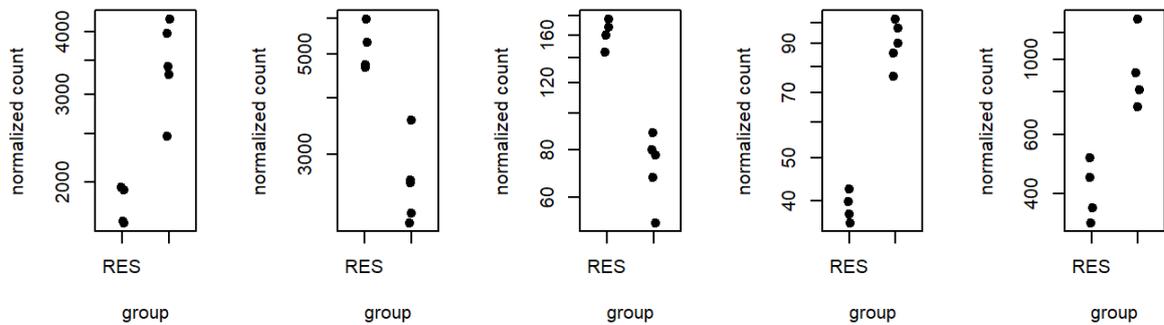
```
mygene11="maker-Scaffold_4732-augustus-gene-5.24"
mygene12="augustus-Scaffold_634-processed-gene-0.9"
mygene13="augustus-Scaffold_2666-processed-gene-0.1"
mygene14="maker-Scaffold_200-augustus-gene-0.17"
mygene15="maker-Scaffold_2061-augustus-gene-1.1"
mygene16="augustus-Scaffold_2128-processed-gene-0.25"
mygene17="maker-Scaffold_647-augustus-gene-1.3"
mygene18="augustus-Scaffold_134-processed-gene-6.11"
mygene19="augustus-Scaffold_82-processed-gene-1.8"
```

```

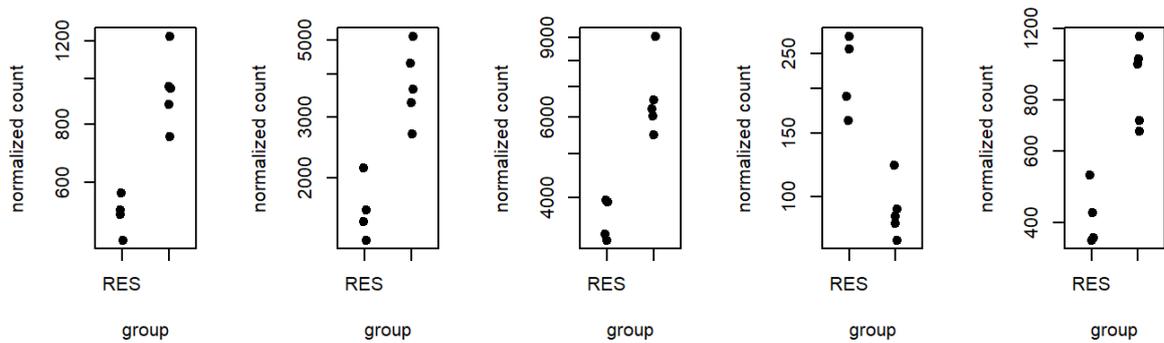
mygene20="augustus-Scaffold_1158-processed-gene-0.1"
plotCounts(dds, gene=mygene11, intgroup="rbtype", pch=19, main=mygene11)
plotCounts(dds, gene=mygene12, intgroup="rbtype", pch=19, main=mygene12)
plotCounts(dds, gene=mygene13, intgroup="rbtype", pch=19, main=mygene13)
plotCounts(dds, gene=mygene14, intgroup="rbtype", pch=19, main=mygene14)
plotCounts(dds, gene=mygene15, intgroup="rbtype", pch=19, main=mygene15)
plotCounts(dds, gene=mygene16, intgroup="rbtype", pch=19, main=mygene16)
plotCounts(dds, gene=mygene17, intgroup="rbtype", pch=19, main=mygene17)
plotCounts(dds, gene=mygene18, intgroup="rbtype", pch=19, main=mygene18)
plotCounts(dds, gene=mygene19, intgroup="rbtype", pch=19, main=mygene19)
plotCounts(dds, gene=mygene20, intgroup="rbtype", pch=19, main=mygene20)

```

Scaffold_4732-augustus-Scaffold_634-process-Scaffold_2666-processScaffold_200-augustusScaffold_2061-augustu



Scaffold_2128-process-Scaffold_647-augustus-Scaffold_134-processScaffold_82-process-Scaffold_1158-proces



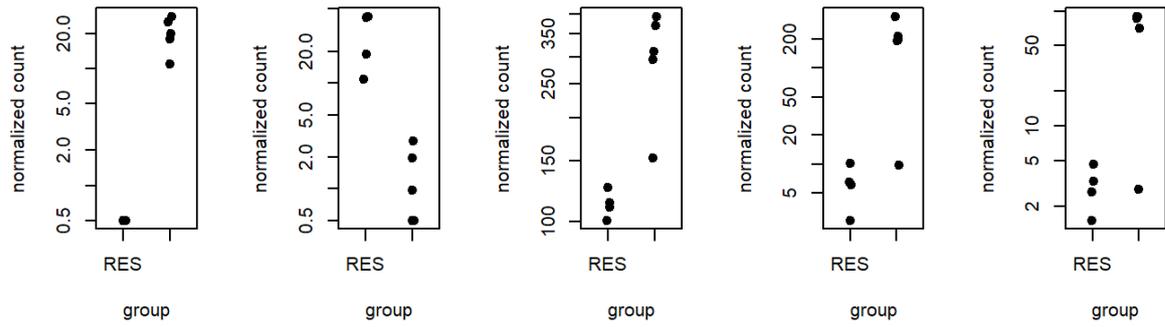
DE genes 21-30

```
par(mfrow = c(2, 5))

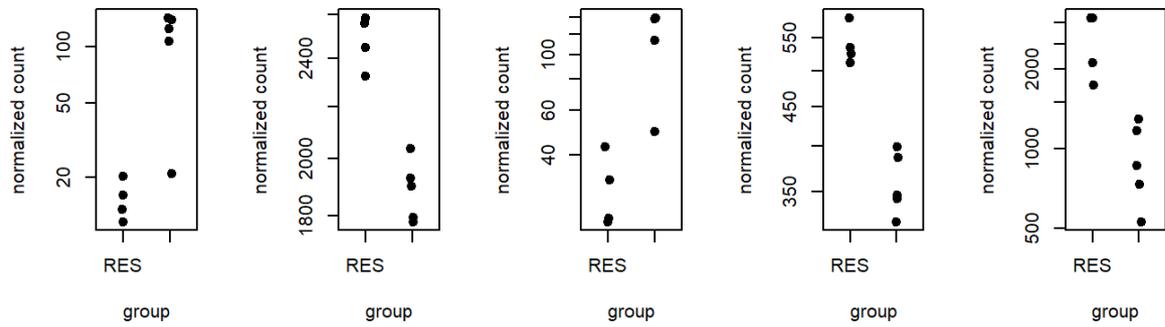
mygene21="augustus-Scaffold_2243-processed-gene-0.15"
mygene22="maker-Scaffold_490-augustus-gene-4.10"
mygene23="maker-Scaffold_200-augustus-gene-0.18"
mygene24="maker-Scaffold_922-augustus-gene-0.7"
mygene25="augustus-Scaffold_1043-processed-gene-0.0"
mygene26="augustus-Scaffold_2727-processed-gene-0.1"
mygene27="augustus-Scaffold_1068-processed-gene-4.10"
mygene28="augustus-Scaffold_4330-processed-gene-1.6"
mygene29="maker-Scaffold_119-augustus-gene-0.4"
mygene30="maker-Scaffold_4299-augustus-gene-0.56"

plotCounts(dds, gene=mygene21, intgroup="rbtype", pch=19, main=mygene21)
plotCounts(dds, gene=mygene22, intgroup="rbtype", pch=19, main=mygene22)
plotCounts(dds, gene=mygene23, intgroup="rbtype", pch=19, main=mygene23)
plotCounts(dds, gene=mygene24, intgroup="rbtype", pch=19, main=mygene24)
plotCounts(dds, gene=mygene25, intgroup="rbtype", pch=19, main=mygene25)
plotCounts(dds, gene=mygene26, intgroup="rbtype", pch=19, main=mygene26)
plotCounts(dds, gene=mygene27, intgroup="rbtype", pch=19, main=mygene27)
plotCounts(dds, gene=mygene28, intgroup="rbtype", pch=19, main=mygene28)
plotCounts(dds, gene=mygene29, intgroup="rbtype", pch=19, main=mygene29)
plotCounts(dds, gene=mygene30, intgroup="rbtype", pch=19, main=mygene30)
```

-Scaffold_2243-proces;Scaffold_490-augustus;Scaffold_200-augustus-Scaffold_922-augustu;-Scaffold_1043-proces



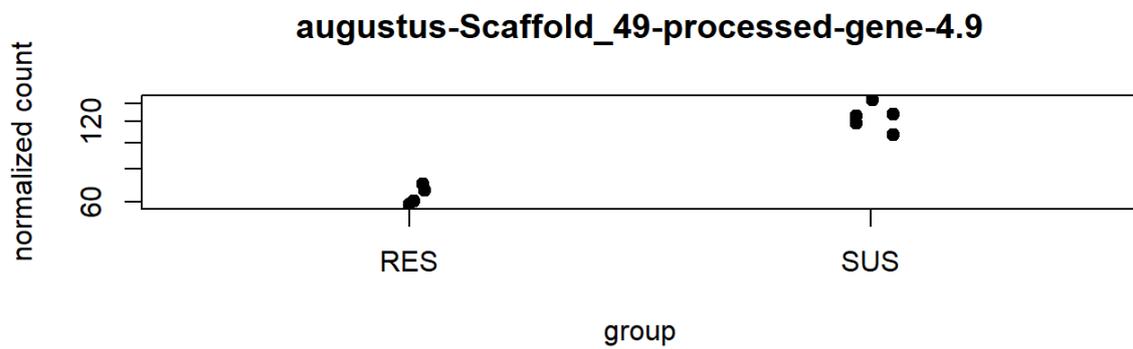
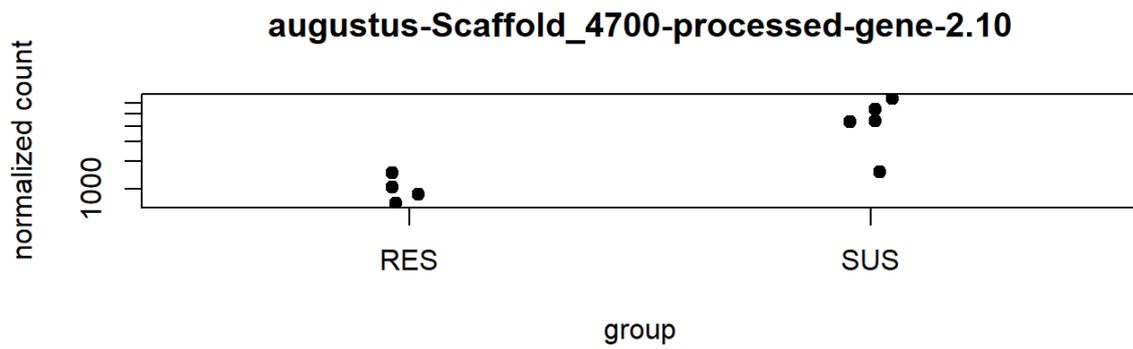
-Scaffold_2727-proces;Scaffold_1068-process;Scaffold_4330-proces;Scaffold_119-augustu;Scaffold_4299-augustu



DE genes 31-32

```
par(mfrow = c(2, 1))

mygene31="augustus-Scaffold_4700-processed-gene-2.10"
mygene32="augustus-Scaffold_49-processed-gene-4.9"
plotCounts(dds, gene=mygene31, intgroup="rbtype", pch=19, main=mygene31)
plotCounts(dds, gene=mygene32, intgroup="rbtype", pch=19, main=mygene32)
```



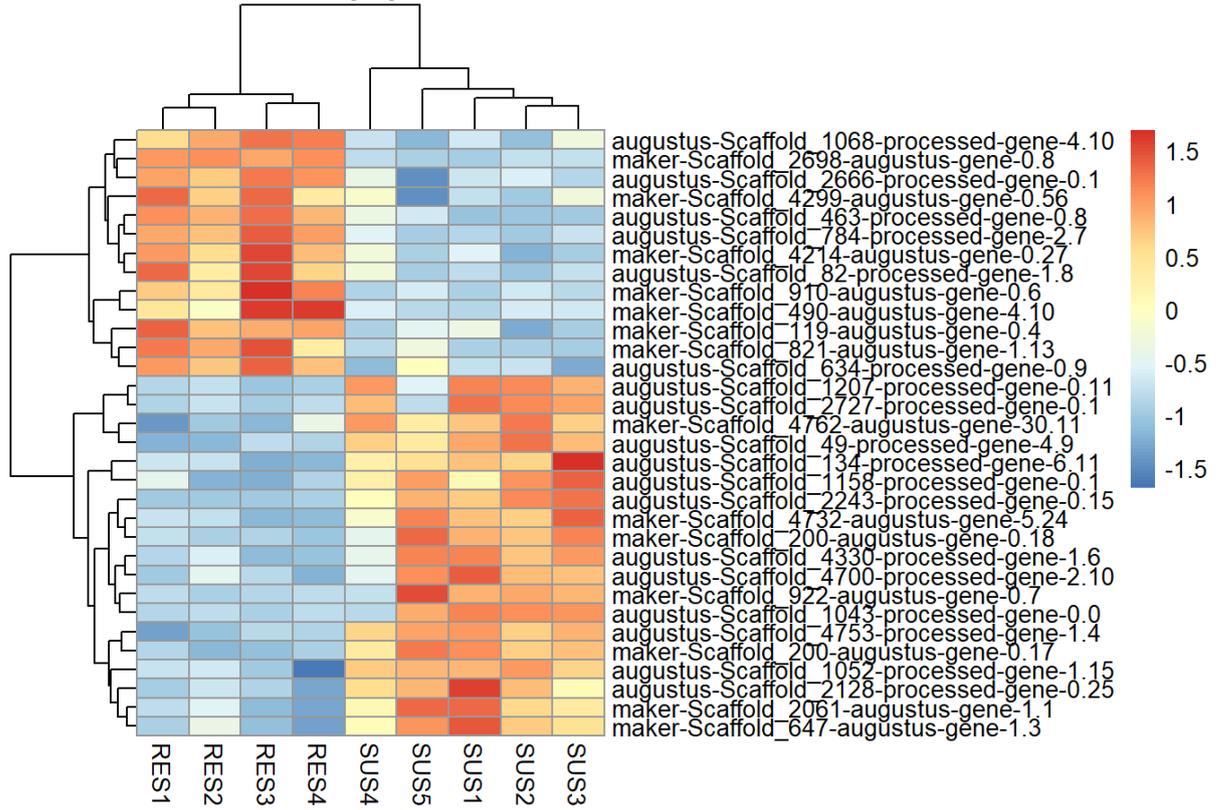
```
par(mfrow = c(1, 1))
par(mfrow = c(1, 1))
```

Producing a Heatmap for the top 32 DE genes between the resistant-breaking and susceptible *N. ribisnigri* biotypes

```
top <- 32
mygenes <- head(DEgenes, top)
hm <- assay(rld)[mygenes,]
hm <- hm - rowMeans(hm)

pheatmap(hm, scale="row", main=paste("Heatmap: p<", alpha, sep=""))
```

Heatmap: $p < 1e-05$



9.4 Supplementary 4: PCR and qRT-PCR primers designed from RNA-seq validation of gene 2698 and 3 house-keeping genes (HKGs)

Housekeeping genes (HKGs)

- **RPS9 – Ribosomal Protein S9**

N. ribisnigri transcript location: TRINITY_GG_1605_c0_g1_i1

- **RPS18 – Ribosomal Protein S18**

N. ribisnigri transcript location: TRINITY_GG_12566_c6_g1_i1

- **RPL13 – Ribosomal Protein L13a**

N. ribisnigri transcript location: TRINITY_GG_4490_c3_g1_i1

Resistant gene of interest – gene-2698

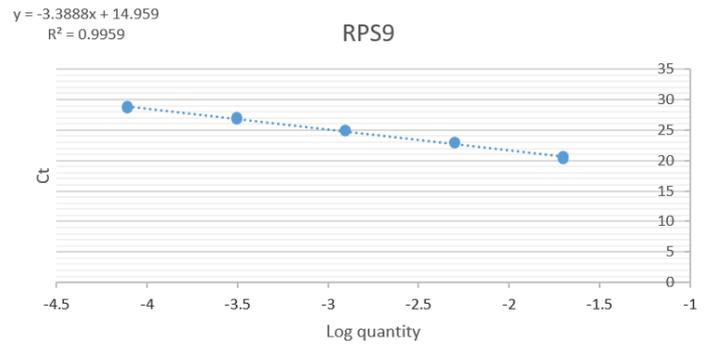
- **PP1**
- **PP2**

qRT-PCR efficiencies (E)

- **RPS9 – Ribosomal Protein S9**

C _T	Dilution
20.52	0.02
20.7	0.02
20.34	0.02
22.82	0.005
22.94	0.005
23.01	0.005
24.8	0.0013
24.96	0.0013
24.97	0.0013
26.84	0.0003
26.8	0.0003
27.04	0.0003
28.61	8E-05
28.64	8E-05
28.96	8E-05

Log Quantity	C _T
-1.69897	20.52
-1.69897	20.7
-1.69897	20.34
-2.30103	22.82
-2.30103	22.94
-2.30103	23.01
-2.90309	24.8
-2.90309	24.96
-2.90309	24.97
-3.50515	26.84
-3.50515	26.8
-3.50515	27.04
-4.10728	28.61
-4.10728	28.64
-4.10728	28.96

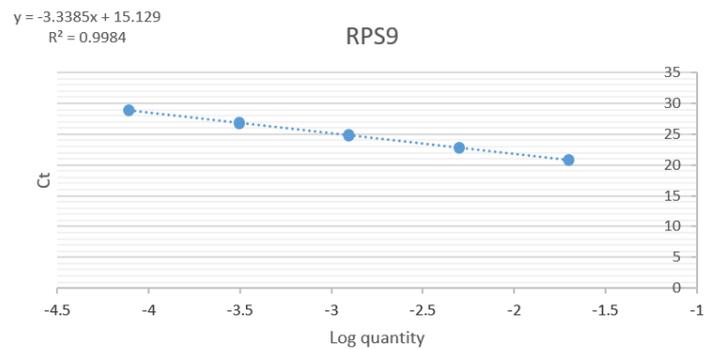


Slope	-1/Slope		Efficiency
-3.3888	0.2950897	1.9728302	97.28 %

- RPS18 – Ribosomal Protein S18**

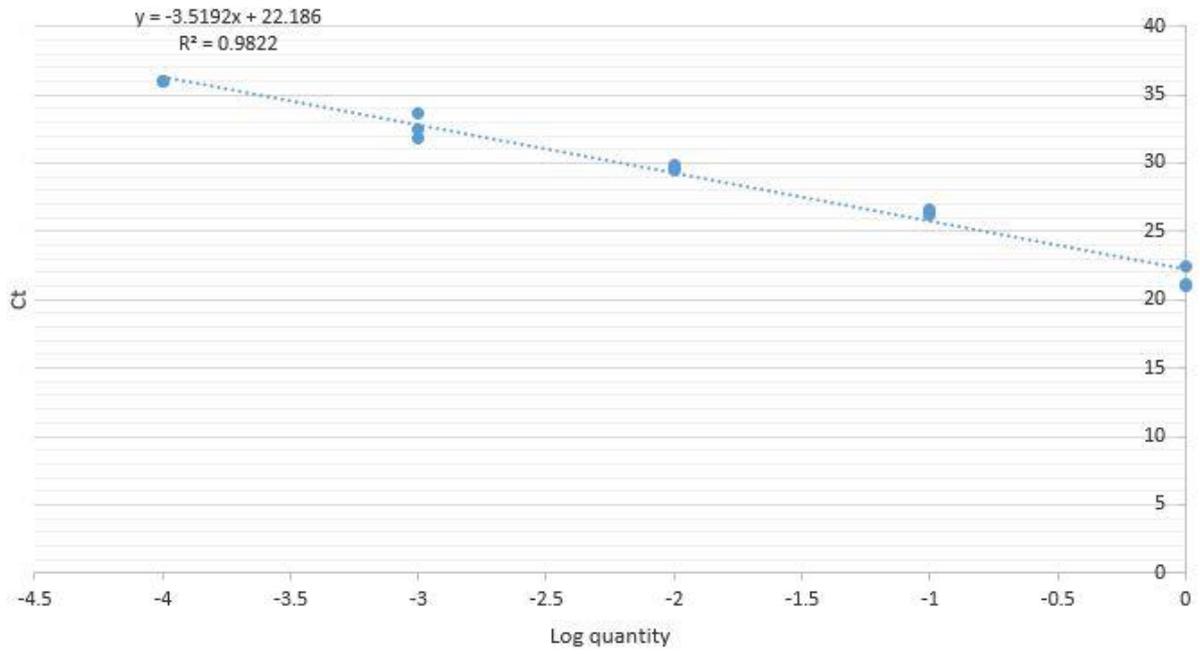
C _T	Dilution
20.79	0.02
20.85	0.02
20.92	0.02
22.64	0.005
22.78	0.005
22.89	0.005
24.68	0.0013
24.81	0.0013
24.91	0.0013
26.6	0.0003
26.81	0.0003
26.96	0.0003
28.75	8E-05
28.92	8E-05
29.01	8E-05

Log Quantity	C _T
-1.69897	20.79
-1.69897	20.85
-1.69897	20.92
-2.30103	22.64
-2.30103	22.78
-2.30103	22.89
-2.90309	24.68
-2.90309	24.81
-2.90309	24.91
-3.50515	26.6
-3.50515	26.81
-3.50515	26.96
-4.10728	28.75
-4.10728	28.92
-4.10728	29.01



Slope	-1/Slope		Efficiency
-3.3385	0.2995357	1.9931304	99.31 %

• **RPL13 – Ribosomal Protein L13a**

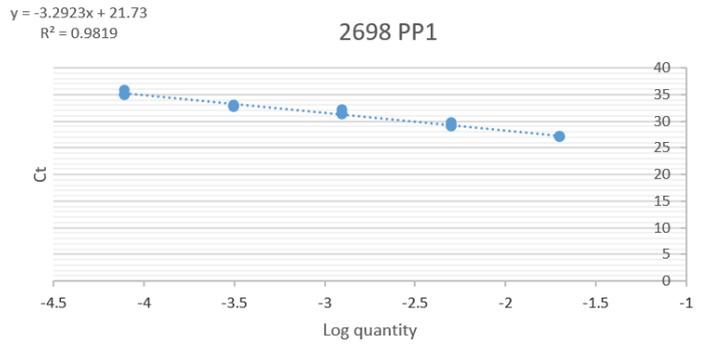


Slope	-1/Slope		Efficiency
-3.5192	0.2841555	1.9237804	92.38 %

• **PP1**

C _T	Dilution
27.1	0.02
27.23	0.02
27.12	0.02
29.11	0.005
29.32	0.005
29.78	0.005
31.45	0.0013
32.14	0.0013
31.37	0.0013
32.9	0.0003
32.76	0.0003
33.156	0.0003
34.94	8E-05
35.94	8E-05
35	8E-05

Log Quantity	C _T
-1.69897	27.1
-1.69897	27.23
-1.69897	27.12
-2.30103	29.11
-2.30103	29.32
-2.30103	29.78
-2.90309	31.45
-2.90309	32.14
-2.90309	31.37
-3.50515	32.9
-3.50515	32.76
-3.50515	33.156
-4.10728	34.94
-4.10728	35.94
-4.10728	35

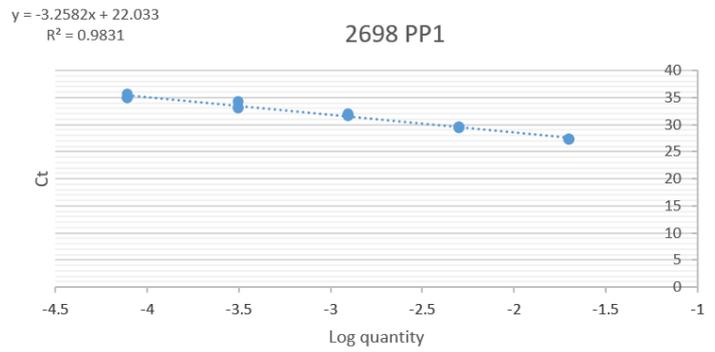


Slope	-1/Slope		Efficiency
-3.2923	0.303739	2.0125145	101.25 %

• **PP2**

C _T	Dilution
27.29	0.02
27.38	0.02
27.41	0.02
29.68	0.005
29.41	0.005
29.67	0.005
32.01	0.0013
31.63	0.0013
31.67	0.0013
33.07	0.0003
33.23	0.0003
34.37	0.0003
35.61	8E-05
34.94	8E-05
35	8E-05

Log Quantity	C _T
-1.69897	27.29
-1.69897	27.38
-1.69897	27.41
-2.30103	29.68
-2.30103	29.41
-2.30103	29.67
-2.90309	32.01
-2.90309	31.63
-2.90309	31.67
-3.50515	33.07
-3.50515	33.23
-3.50515	34.37
-4.10728	35.61
-4.10728	34.94
-4.10728	35



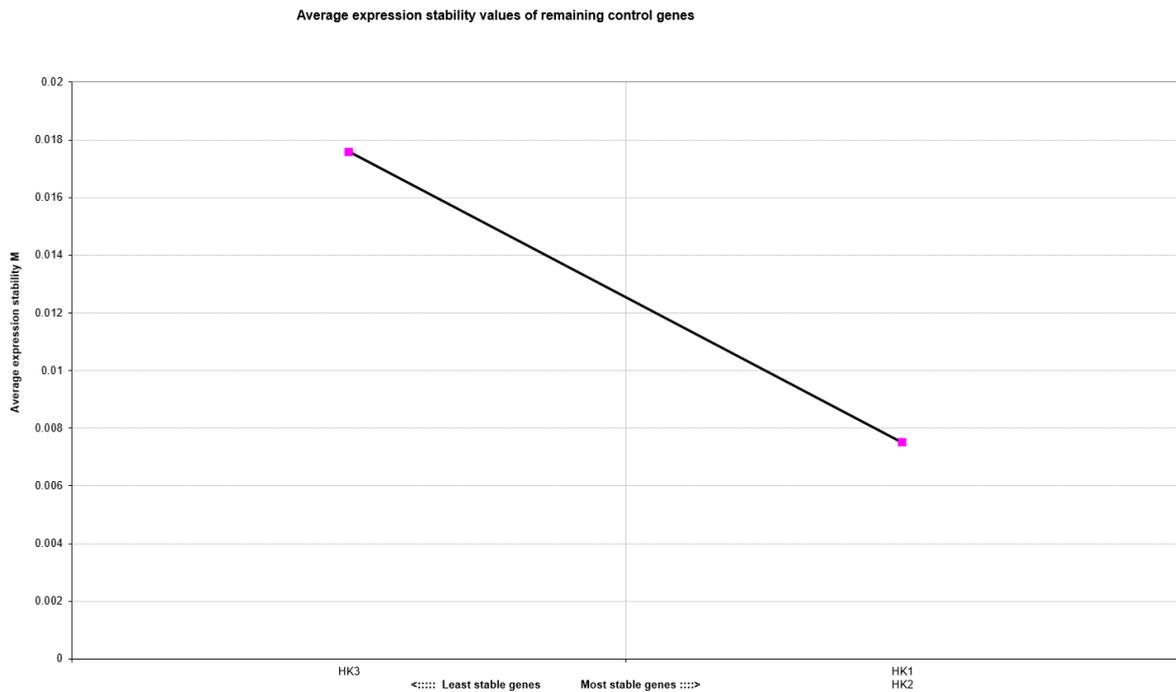
Slope	-1/Slope		Efficiency
-3.2582	0.3069179	2.0272996	102.73 %

Housekeeping (HKGs) stability (GeNorm)

Normalised expression levels for *RPS9* (HK1), *RPS18* (HK2) and *RPL13* (HK3). A M value of < 1.5 suggests a more stable gene expression or low variation, and an M-score > 1.5 indicating high variation and therefore not a suitable HKG. Most stable and least stable HKG's are highlighted in green and red, respectively.

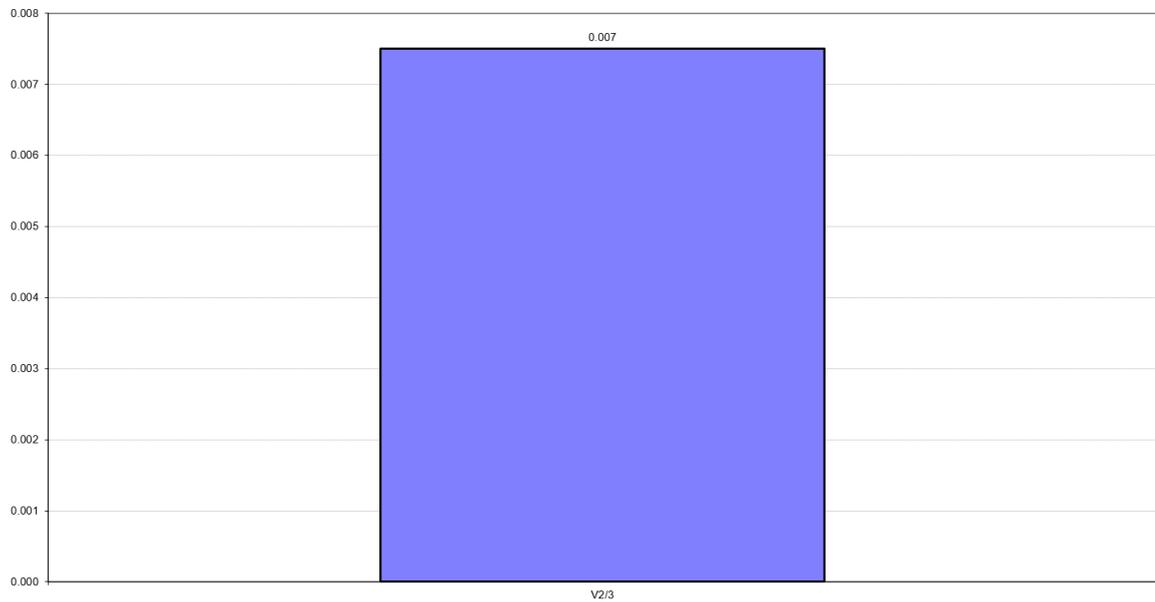
Change	HK1	HK2	HK3	Normalisation Factor
sample A	2.68E+01	2.68E+01	3.14E+01	0.9942
sample b	2.69E+01	2.69E+01	3.21E+01	1.0027
sample c	2.69E+01	2.69E+01	3.19E+01	1.0013
sample d	2.68E+01	2.69E+01	3.18E+01	0.9986
sample e	2.69E+01	2.66E+01	3.28E+01	1.0058
sample f	2.67E+01	2.68E+01	3.19E+01	0.9975

M < 1.5 **0.014** **0.017** **0.023**



Average stability of HK1, HK2 and HK3. Least stable on the left (HK3) and most stable on the right (HK1 and HK2).

Determination of the optimal number of control genes for normalization



Optimal number of HKGs for the qPCR analysis is identified to be between 2 to 3. Two HKGs (HK1

– *RPS9* and HK2 *RPS18* were used to validate gene-2698 using qPCR.

Appendix 4: Chapter 5 supplementary material

9.5 Supplementary 1: Chapter 5 – Population genetics DNA yield analysis between sites and years

Analysing the DNA yields from 225 *Nasonovia ribisnigri* samples between 2003-2020 collected in the Rothamsted Insect Survey suction traps.

First, these are the required libraries:

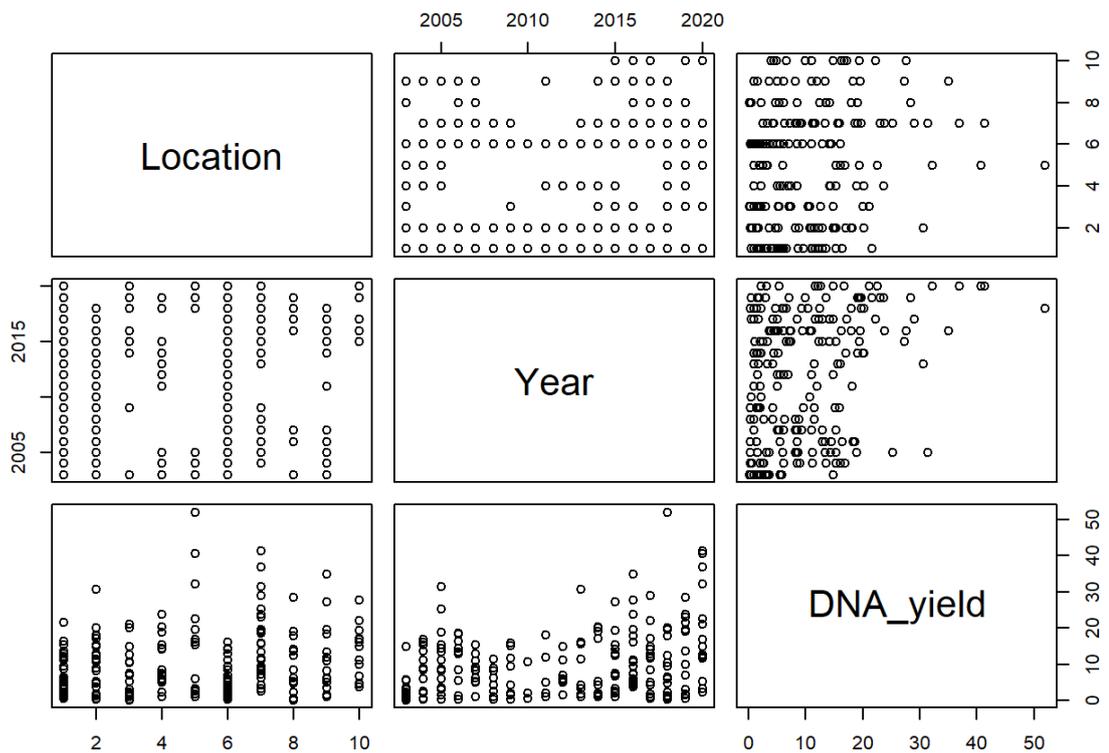
```
library("tidyverse")
library("hrbrthemes")
library("viridis")
library("viridisLite")
library("ggplot2")
library("dplyr")
library("hrbrthemes")
library("readxl")
```

Including the necessary file (available on request).

```
DNA_quantities_qubit_microsatellites_violin <- read_excel("DNA_quantities_qubit_microsatellit
es_violin.xlsx")
attach(DNA_quantities_qubit_microsatellites_violin)
data <- DNA_quantities_qubit_microsatellites_violin
```

Plot of DNA yields and years.

```
plot(data)
```



```
summary(data)
```

```
## Location      Year      DNA_yield
## Length:224    Min. :2003  Min. : 0.100
## Class :character 1st Qu.:2006  1st Qu.: 2.100
## Mode :character  Median :2013  Median : 6.650
##              Mean  :2012  Mean  : 9.445
##              3rd Qu.:2017  3rd Qu.:14.425
##              Max. :2020  Max. :52.000
```

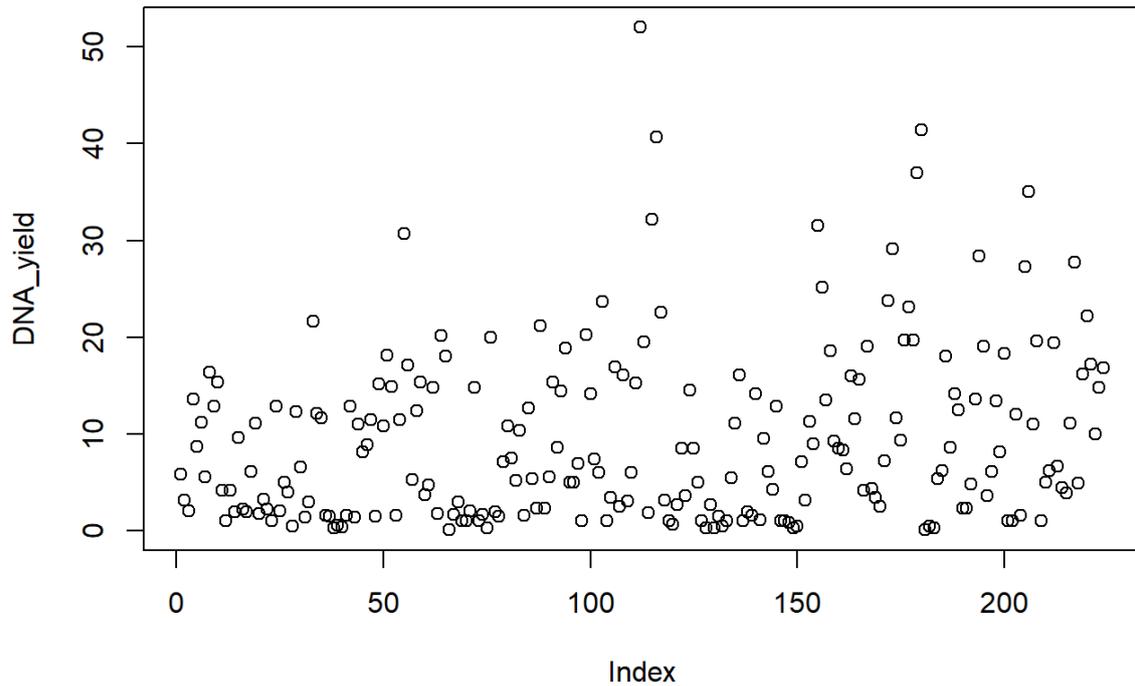
Need to change year and location to factor.

```
DNA_quantities_qubit_microsatellites_violin$Year<-factor(DNA_quantities_qubit_microsatellite
s_violin$Year)
```

```
DNA_quantities_qubit_microsatellites_violin$Location<-factor(DNA_quantities_qubit_microsatellites_violin$Location)
```

Plot to view data.

```
plot(DNA_yield)
```



Create a dataset with new parameters.

```
data <- data.frame(  
  name=c(Location ),  
  value=c(DNA_yield ))
```

Plotting the data using ggplot2 and plotting jitter boxplots for each location.

```
data %>%
```

```

ggplot( aes(x=name, y=value, fill=name)) +
geom_boxplot() +
scale_fill_viridis(discrete = TRUE, alpha=0.6) +
geom_jitter(color="black", size=2, alpha=0.9) +
theme_ipsum() +
theme(
  legend.position="none",
  plot.title = element_text(size=11)
) +
ggtitle("DNA yield from RIS archive (Nasonovia ribisnigri)") +
xlab("") +
ylab("DNA yield (ng/ul) ")

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font family not
## found in Windows font database

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font family not
## found in Windows font database

## Warning in grid.Call(C_stringMetric, as.graphicsAnnot(x$label)): font family not
## found in Windows font database

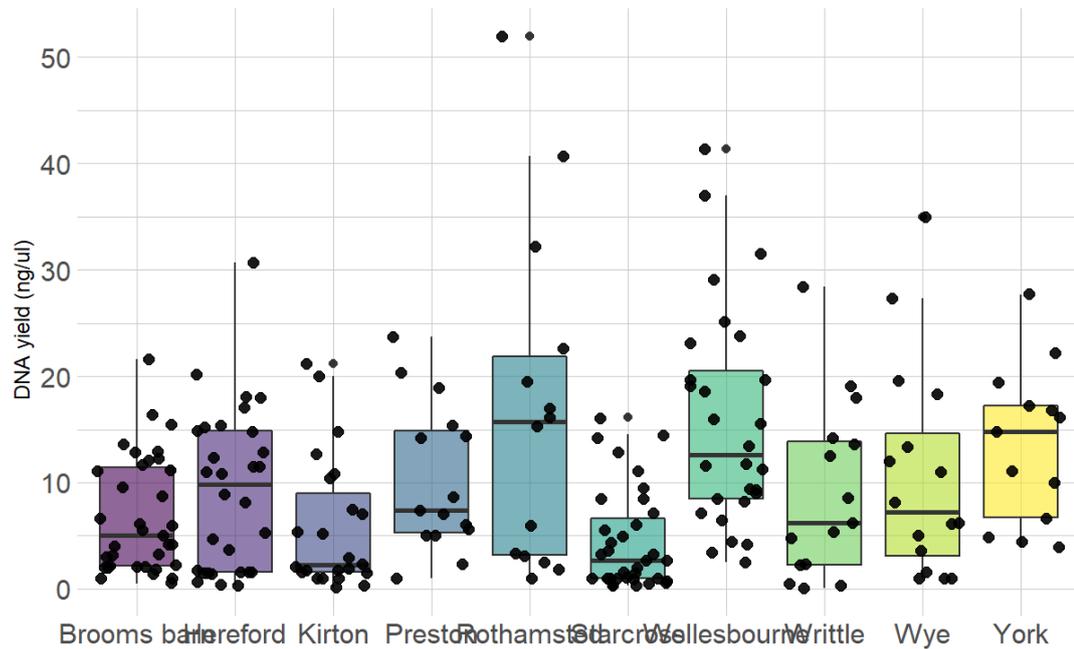
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## family not found in Windows font database

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font

```


DNA yield from RIS archive (*Nasonovia ribisnigri*)



Statistical analysis

Before any statistical test can be conducted, we need to plot the data to check that it meets the necessary assumptions before using the ANOVA function. The ANOVA will be used to determine if DNA yields significantly differs by Location between pooled samples 2003-2020.

Homogeneity of variances.

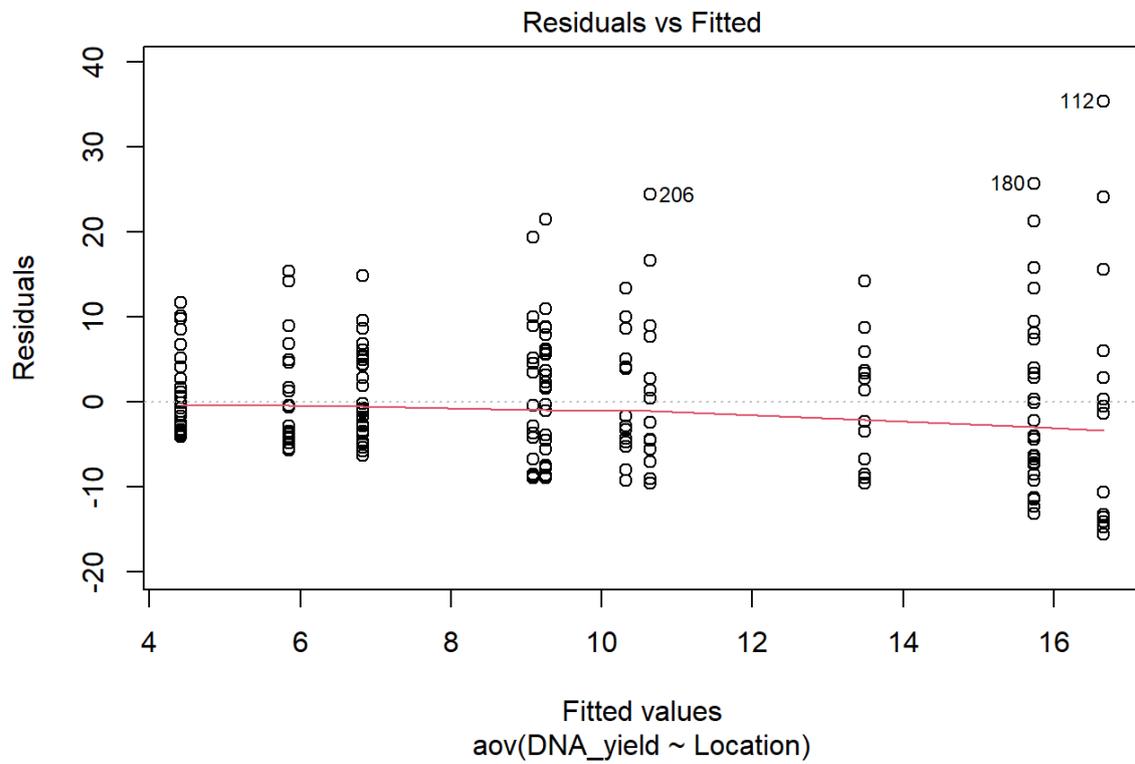
```
res.aov <- aov(DNA_yield ~ Location, data = data)
```

```
summary(res.aov)
```

```
##      Df Sum Sq Mean Sq F value  Pr(>F)
## Location    9  3511   390.1  5.876 2.53e-07 ***
## Residuals 214 14205    66.4
## ---
```

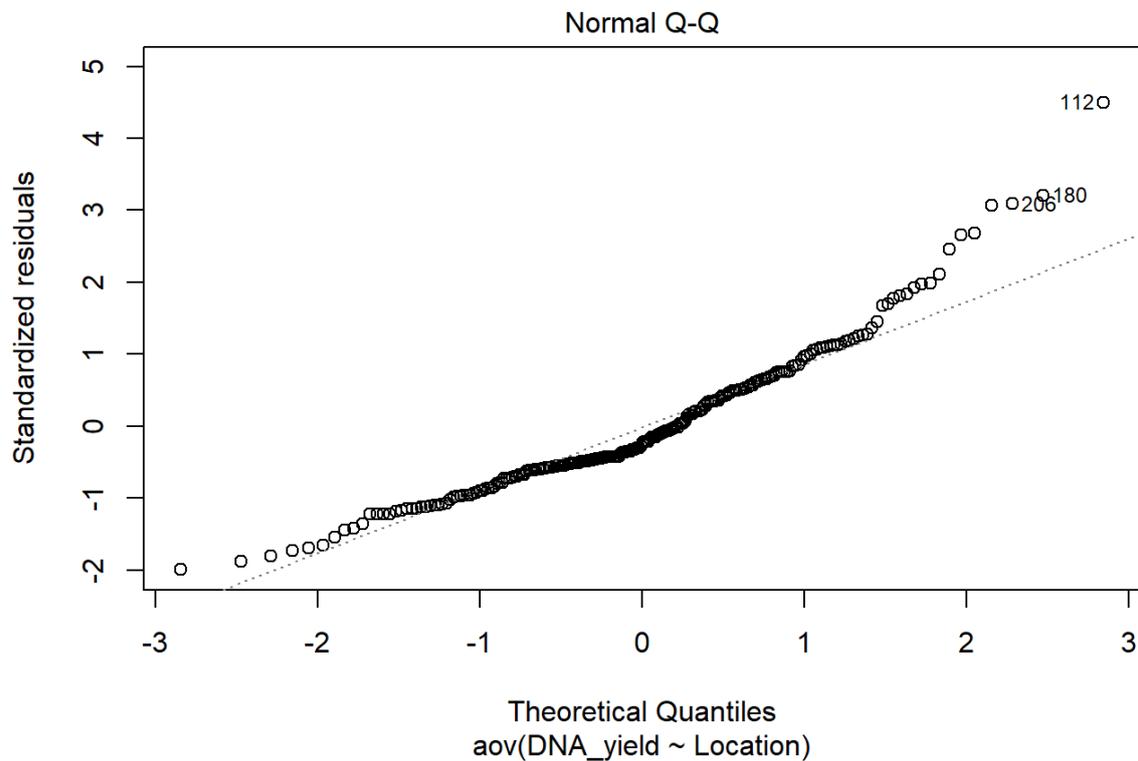
```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(res.aov, 1)
```



Checking normality.

```
plot(res.aov, 2)
```



Q-Q plot starts to tail off at the end, so need to make sure it is normally distributed before using the parametric ANOVA opposed to the non-parametric equivalent (e.g. Kruskal-Wallis).

A Shapiro-Wilk Test was used to determine whether the data conforms to a parametric assumption. First we extract the residuals

```
aov_residuals <- residuals(object = res.aov )
```

Then run the Shapiro-Wilk Test:

```
shapiro.test(x = aov_residuals )
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: aov_residuals
```

```
## W = 0.93993, p-value = 5.688e-08
```

A *P* value above 0.05, the data is considered 'normal' and below 0.05 is considered to be significantly deviated from a normal distribution.

The Shapiro-Wilk Test suggests the data is not normally distributed and therefore a non-parametric equivalent was used for further statistical analysis. In this case a Kruskal-Wallis Test.

Log transformation was attempted to normalise data but this failed (R code not shown).

```
kruskal.test(DNA_yield ~ Location, data = data)
```

```
##
```

```
## Kruskal-Wallis rank sum test
```

```
##
```

```
## data: DNA_yield by Location
```

```
## Kruskal-Wallis chi-squared = 44.806, df = 9, p-value = 1.002e-06
```

The Kruskal-Wallis test suggests that DNA yields significantly differs by location. This test does not show which years are significantly different from one another, so a multiple pairwise-comparison between groups was conducted.

Using a `pairwise.wilcox.test()` enable you to calculate pairwise comparisons between group levels, with corrections for multiple testing.

```
pairwise.wilcox.test(Year, Location, p.adjust = "BH")
```

```
##
```

```
## Pairwise comparisons using Wilcoxon rank sum test with continuity correction
```

```
##
```

```
## data: Year and Location
```

```
##
```

```
##      Brooms barn Hereford Kirton Preston Rothamsted Starcross
```

```
## Hereford  0.8459  -    -    -    -    -
```

```
## Kirton    0.9885  0.9062  -    -    -    -
```

```
## Preston   0.8459  0.9188  0.9062  -    -    -
```

```
## Rothamsted 0.9062  0.9885  0.9062  0.9062  -    -
```

```

## Starcross 0.9885 0.8305 0.9885 0.8305 0.9062 -
## Wellesbourne 0.9062 0.6203 0.9062 0.6203 0.8459 0.9062
## Writtle 0.6203 0.2067 0.8305 0.3138 0.9062 0.6203
## Wye 0.9188 0.9062 0.9885 0.8459 0.9885 0.9062
## York 0.0317 0.0056 0.1982 0.0123 0.7470 0.0437
## Wellesbourne Writtle Wye
## Hereford - - -
## Kirton - - -
## Preston - - -
## Rothamsted - - -
## Starcross - - -
## Wellesbourne - - -
## Writtle 0.8305 - -
## Wye 0.8459 0.3138 -
## York 0.1111 0.9885 0.0715
##
## P value adjustment method: BH

```

A *P* value below 0.05 is considered significant.

Overall the pairwise Wilcox test highlighted that only York significantly differed between Brooms Barn, Hereford, Preston and Starcross. This is likely due to York having no DNA yields below 4ng/ul and a high proportion above 10ng/ul. Sample size for York in particular was quite low due to few *N. ribisnigri* being captured in the York suction trap during the study period. Therefore it is difficult to ascertain whether samples of *N. ribisnigri* in York yield higher DNA concentrations compared to the other 9 sites.

Starcross *N. ribisnigri* samples were particularly difficult to get high DNA yields from, with many falling below the 4ng/ul threshold. These however were still not significantly different between the other 9 sites. Low DNA yields from Starcross could be a result of collection/handling methods prior to receipt at the Rothamsted Insect Survey for processing but further information is needed to clarify this.