

ORIGINAL ARTICLE

Genome assembly and transcriptomic analysis to elucidate the ability of *Nasonovia ribisnigri* to break host plant resistance

Dion Garrett^{1,2}  | Graham Teakle²  | Rosemary Collier² | James R. Bell^{1,3}  | Sergio Cerezo-Medina⁴  | Ramiro Morales-Hojas¹ 

¹Rothamsted Insect Survey, Rothamsted Research, Harpenden, UK

²Warwick Crop Centre, Wellesbourne Campus, University of Warwick, Warwick, UK

³Centre for Applied Entomology and Parasitology, School of Life Sciences, Keele University, Keele, UK

⁴Cereal Transformations, Rothamsted Research, Harpenden, UK

Correspondence

Dion Garrett, Rothamsted Insect Survey, Rothamsted Research, West Common, Harpenden, AL5 2JQ, UK.

Email: dion.garrett@rothamsted.ac.uk

Funding information

Biotechnology and Biological Sciences Research Council

Associate Editor: Elizabeth Duncan

Abstract

Aphid genomic resources enable the study of complex life history traits and provide information on vector biology, host adaption and speciation. The currant-lettuce aphid (*Nasonovia ribisnigri* (Hemiptera: Aphididae) (Mosley)) is a cosmopolitan pest of outdoor lettuce (*Lactuca sativa* (Asterales: Asteraceae) (Linnaeus)). Until recently, the use of resistant cultivars was an effective method for managing *N. ribisnigri*. A resistant cultivar containing a single gene (Nr-locus), introduced in the 1980s, conferred complete resistance to feeding. Overreliance of this Nr-locus in lettuce resulted in *N. ribisnigri*'s ability to break resistance mechanism, with first reports during 2003. Our work attempts to understand which candidate gene(s) are associated with this resistance-breaking mechanism. We present two de novo draft assemblies for *N. ribisnigri* genomes, corresponding to both avirulent (Nr-locus susceptible) and virulent (Nr-locus resistant) biotypes. Changes in gene expression of the two *N. ribisnigri* biotypes were investigated using transcriptomic analyses of RNA-sequencing (RNA-seq) data to understand the potential mechanisms of resistance to the Nr-locus in lettuce. The draft genome assemblies were 94.2% and 91.4% complete for the avirulent and virulent biotypes, respectively. Out of the 18,872 differentially expressed genes, a single gene/locus was identified in *N. ribisnigri* that was shared between two resistant-breaking biotypes. This locus was further explored and validated in Real-Time Quantitative Reverse Transcription PCR (qRT-PCR) experiments and has predicted localisations in both the cytoplasm and nucleus. This is the first study to provide evidence that a single gene/locus is likely responsible for the ability of *N. ribisnigri* to overcome the Nr-locus resistance in the lettuce host.

KEYWORDS

currant-lettuce aphid, gene-for-gene resistance, host plant resistance, insect pest, resistance-breaking mechanisms

INTRODUCTION

Aphids are an economically important insect group with over 450 species impacting major cropping systems worldwide (Blackman & Eastop, 2017). Over 100 species of the world's aphid fauna are of significant economic importance, transmitting 50% of all insect-borne

Graham Teakle, Rosemary Collier, James R. Bell, Sergio Cerezo-Medina, Ramiro Morales-Hojas contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Insect Molecular Biology* published by John Wiley & Sons Ltd on behalf of Royal Entomological Society.

plant viruses, and they are drawn largely from one family, the Aphididae (Blackman & Eastop 2017; Katis et al., 2007; Nault, 1997). Genome sequencing has provided several resources and tools that have facilitated the study of the genetic basis of the aphid–plant–virus interactions (Biello et al., 2020; Field et al., 2017; Jiang et al., 2019; Mathers et al., 2020; Tagu et al., 2008).

Over the last decade and with advances in sequencing technology, computational power and the reduction of associated costs, reference genomes have been easily generated. However, in 2010, the only aphid genomic model available to study aphid genetics and aphid–plant interactions was the pea aphid (*Acyrtosiphon pisum*) (Harris) (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., 2020; Li et al., 2019; Mathers et al., 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 located on the Bioinformatics Platform for Agroecosystem Arthropods (BIPAA) (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 those regarding 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).

The currant–lettuce aphid, *Nasonovia ribisnigri* (Mosley), is a major pest of outdoor lettuce crops (*Lactuca sativa* L. (Linnaeus)) and is responsible for large economic losses (Diaz et al., 2012; Liu & McCreight, 2006; Niemann et al., 2022). The nymphs and adults preferentially feed on young developing leaves of lettuce, so they predominantly reside in the centre of the lettuce head, making the application of foliar insecticides with contact action ineffective (Aarts et al., 1999). This is particularly a problem in well-developed lettuce plants, in which the older leaves surround feeding aphids, leading to contamination of the harvested crop and thus economic loss. Biological control by both predatory insects and parasitoids is also hindered by their concealment in the centre of the lettuce, and even when successful, contamination by beneficial insects remains an issue (Tatchell et al., 2017).

To tackle these complex issues concerning insecticide use and the changing legislation pressures (e.g., the banning of systemic neonicotinoids on both flowering and non-flowering crop in the United Kingdom), a lettuce cultivar resistant to *N. ribisnigri* (Nr) was introduced into the growers' recommended list during the early 1980s. The single-gene resistance (Nr-locus) used originated from a wild species of lettuce (*L. virosa* (Galsso, Banfi, Bartolucci and Ardenghi, 1753)) that showed natural resistance to *N. ribisnigri*, and it was introgressed into lettuce via an interfertile bridging species (*L. 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 (Tatchell et al., 2017). The resistance in *N. ribisnigri* associated with the Nr-locus is unlikely to be mechanical or due to a change in host plant quality

(antibiosis), as it is expected that other aphid species would also be affected, so the resistance mechanism(s) involved are most likely species-specific (Tjallingii & Esch, 1993). Research has shown that the Nr-locus has little effect on other species of aphids that feed on lettuce, with complete susceptibility to *Macrosiphum euphorbiae* and partial resistance to *M. persicae* (Reinink & Dieleman, 1989). The variation in partial resistance observed in three lines of *M. persicae* is most likely a result of the interaction of the Nr-locus with additional genes, as opposed to the Nr-locus itself (Reinink & Dieleman, 1989).

As the Nr resistance in lettuce cultivars is conferred by a single gene/locus, and since there are few other effective control strategies, the overuse and reliance on this mechanism resulted in the first reports of resistance breakdown in early 2000 (van der Arend, 2003). This led to the identification of a resistance-breaking (Rb) biotypes of *N. ribisnigri*, which could overcome the Nr-locus. In both France and Germany during 2007, these new Rb biotypes were identified on resistant cultivars and subsequently an Rb biotype was identified in Kent, United Kingdom during 2009 (Hough, 2013). In the absence of lettuce cultivars that exploit new resistance mechanisms, with 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). A selection of these avirulent and virulent lineages from the United Kingdom has been collected and cultured to study the potential mechanisms of resistance within the aphid, as they are still unknown.

Gene-for-gene resistance has been highlighted in several insects such as the brown plant hopper (*Nilaparvata lugens*), Asian rice gall midge (*Orseolia oryzae*) (Wood-Mason) and the barley midge (*Mayetiola destructor*) (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 several times, with the first example being between the barley midge (*M. destructor*) and wheat (*Triticum* spp.) (Hatchett & Gallun, 1970). More recent examples have identified several *R* genes (Gm1–Gm11) in rice (*Oryza sativa*) against the Asian rice gall midge (*O. oryzae*) which have been mapped in the rice genome (Himabindu et al., 2010) and a gene (*BPH15*) in the brown planthopper (*N. lugens*) that mediates a unique defence mechanism in rice (Lv et al., 2014). With the presence of a single resistance gene in the lettuce host, and the strong selection pressure to overcome it in the currant–lettuce aphid, it is hypothesised that a single gene or locus is responsible for the breakdown of resistance.

The genome for *N. ribisnigri* has not been sequenced to date, and only the mitochondrial cytochrome c oxidase subunit I barcode and a selection of odorant binding proteins have been sequenced. This project aimed to generate two de novo assembled reference genomes for *N. ribisnigri*, corresponding to both a biotype avirulent to host plant resistance and a virulent biotype. Second, gene expression of the two *N. ribisnigri* biotypes were investigated using RNA-sequencing (RNA-seq) to study the potential mechanisms of resistance to the

TABLE 1 Summary statistics of the currant-lettuce aphid, *Nasonovia ribisnigri* genomes.

Genome	Genome size (mb)	N50	Scaffolds	Max scaffold size	GC content (combined %)	BUSCO (%)	Accession no.
Nr8 (123)	423	295,999	4778	3391,359	29.81	94.2	SAMN29667034
WT_Kent (123)	423.2	259,518	5225	2,467,012	29.8	92.5	SAMN29667035
UK631 (1)	367.9	105,098	8130	967,723	29.79	91.6	SAMN29667036
UK631 (2)	359.5	71,978	10,488	721,085	29.87	82.9	SAMN29667037
UK631 (12)	367.9	107,002	8047	1,394,296	29.8	91.3	SAMN29667038
Kent_CL (1)	384.2	163,106	6264	1,500,737	29.76	92.9	SAMN29667039
Kent_CL (2)	384	162,012	6305	1,147,043	29.74	92.4	SAMN29667040
Kent_CL (12)	382.3	99,972	8234	1,161,853	29.8	91.4	SAMN29667041
Ely (1)	366.9	107,464	8035	956,705	29.75	91.1	SAMN29667042
Ely (2)	365.2	101,694	8289	903,197	29.77	91.5	SAMN29667043
Ely (12)	366.9	109,458	7890	1,394,296	29.7	91.5	SAMN29667044

Note: Nr8 and WT_Kent are avirulent biotypes and UK631, Kent CL and Ely are virulent biotypes. Only the most complete genome assemblies for the avirulent and virulent were selected and used for downstream annotation and analysis, highlighted in red. All genomes generated and presented in the article are available at the National Center for Biotechnology Information (NCBI) (BioProject ID: PRJNA857679). (1), (2), (12) and (123) denote different sequencing read data from the same *N. ribisnigri* line that was used in the genome assembly and annotation process.

Abbreviation: BUSCO, Benchmarking Universal Single-Copy Orthologs.

Nr-locus in lettuce. By providing two draft genome assemblies of *N. ribisnigri* and a transcriptomics analysis of an avirulent and virulent biotype, this article provides additional insight and knowledge to understand the genetic basis of host plant resistance breaking by *N. ribisnigri*, bringing science closer to elucidating the resistance mechanisms involved.

RESULTS

Genome de novo assembly and annotation

Altogether, 45.2 Gb Illumina short read data were generated for the avirulent biotype *N. ribisnigri* genome, and 20.6 Gb was produced for the virulent biotype *N. ribisnigri* genome. From the two SpotON flow cells, a total of 2,966,975 (average sequence length; 2155) and 5,167,875 (average sequence length; 2401) long-read nanopore sequences were generated for the Kent CL (virulent) and Nr8 (avirulent) genomes, respectively. This produced an additional 30.3 Gb of long-read Nanopore data for the hybrid genome assembly. The hybrid de novo genome assembly of the avirulent (Nr8_123) *N. ribisnigri* genome was assembled into 4778 scaffolds with an N50 of 295 Kb, whereas, the resistant Nr:1 (Kent CL 1) was assembled into 6264 scaffolds with an N50 of 163 Kb (Table 1). Multiple *N. ribisnigri* lineages from the sequence read data were assembled but only two most complete genomes of the avirulent (Nr8_123) and virulent (Kent CL 1) were annotated and used for further downstream analysis. The guanine and cytosine (GC) contents of the *N. ribisnigri* genomes ranged between 29.75% and 30.1%. After thorough contaminant removal using MEGAN, no further contaminant sequences could be identified using this contaminant removal pipeline and suggesting that most contaminant sequences were removed.

The genome size of *N. ribisnigri* was estimated to be between 384 and 423 Mbp, for the virulent and avirulent strains, respectively. The genome size of *N. ribisnigri* is comparable to other previously sequenced aphid genomes (Chen et al., 2019; Nicholson et al., 2015; Wenger et al., 2020). The number of scaffolds in the genome assemblies of the different strains sequenced in this study ranged between 4778 and 12,004 and the maximum scaffold size between 721 and 3391 Kbp (Table 1). The k-mer spectra analyses (k-mer length of 21) on both the Rb and wild-type (WT) genome indicates that the WT genome has a k-mer coverage of 24.9 with a 0.41% error, whereas the Rb genome has a k-mer coverage of 14.4 and an error of 2.37% (Figure S1).

The completeness of the genome assemblies as determined using Benchmarking Universal Single-Copy Orthologs (BUSCO) (Insecta_odb9), indicated that the *N. ribisnigri* genomes of the avirulent and virulent (resistance-breaking) aphids were 94.2% (complete: 90.5%, duplicated: 3.6%, missing: 4.5%) and 91.4% (complete: 88.1%, duplicated: 4.7%, missing: 5%) complete, respectively (Figure 1). A de novo hybrid of multiple *N. ribisnigri* biotypes (PAN) genome assembly was attempted using the combination of reads from all biotypes to determine whether this would improve overall assembly statistics but was unsuccessful. 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 for the WT genome, ~8000 genes had no IPS matches, ~23,000 genes had no gene ontology (GO) terms but had IPS matches and ~7000 had both GO terms and IPS matches. Out of these, a further 3044 genes had either a start or stop codon missing and were subsequently removed. For the Rb genome, IPS identified 36,018 predicted genes, ~7500 genes had no IPS match,

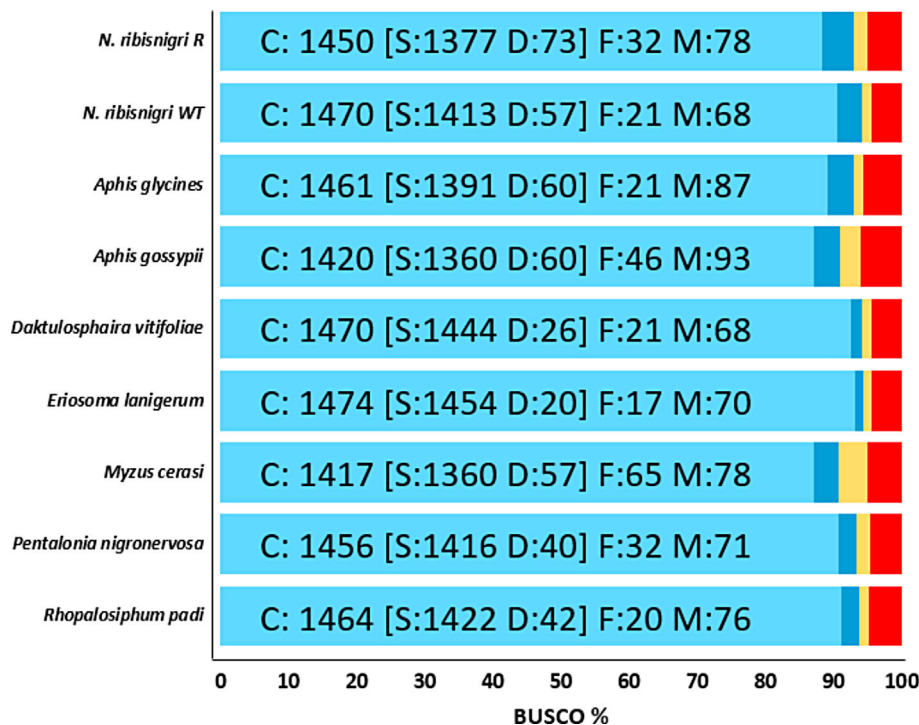


FIGURE 1 Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis score of both *Nasonovia ribisnigri* genomes (WT = Nr:0, avirulent; R = Nr:1, virulent (resistance-breaking) and previously published aphid genomes (obtained from the Bioinformatics Platform for Agroecosystem Arthropods (BIPAA) genomic resources) using an arthropod gene set (Insecta_odb9) ($n = 1562$). C, complete; D, duplicate (dark blue); F, fragmented (yellow); M, missing (red); S, single copy (light blue).

~22,000 genes had no GO terms but had IPS matches and ~6800 had both GO terms and IPS matches. There were several species-specific genes identified from the BLAST results, with over 12,000 to be associated with *A. pisum* and *M. persicae* (Figure S2A,B).

Transcriptome de novo assembly and annotation

A total of 90 Gb of strand-specific paired-end RNA-seq data from isogenic lines of *N. ribisnigri* were generated and provided by Warwick Crop Centre, and this 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 1093 (Figure S3A). The transcriptome assembly N50 was 2341, in which 50% of all bases in the assembly are covered by sequences equal or larger than 2341 and is comparatively high for a non-model organism (Francis et al., 2013). The transcriptome completeness (BUSCO) for the avirulent *N. ribisnigri* biotype was 94.8% and is similar to other available aphid transcriptomes, such as the pea aphid (*A. pisum*), which was 96.2% (Figure S3B).

RNA-seq analysis

In total, there were 183 million RNA-seq reads for *N. ribisnigri* for all conditions (Figure S4). Over 115 million RNA-seq reads were obtained

for the *N. ribisnigri*, which were avirulent to the *Nr-locus* in the lettuce host plant, with the remaining 68 million RNA-seq reads obtained for *N. ribisnigri*, which were resistant to the *Nr-locus* in the lettuce host plant. The power analysis highlighted that the number of replicates used in the analysis has a marginal power of 0.68–0.76 (Supporting Information S16). Doubling the RNA-seq depth would have only increased this to 0.7–0.78 and 10 replicates would have provided a power of 0.88–0.95. In the case of genes with very low read counts (i.e., 0–10), replicates below five had low power (0–0.2) but, when read counts were above 10, the power increased to >0.4 for four or more replicates, and >0.7 with read counts >320.

The DESeq2 data identified 689 differentially expressed (DE) genes between Rb and WT strains with an alpha of $p < 0.05$ out of the 18,872 genes (Figure S5). Of these 689 DE genes, 351 up- and 338 down-regulated genes were identified between the avirulent and virulent groups of *N. ribisnigri*. Further stringent refinement (alpha = $p < 0.00001$), identified 32 DE genes out of the 18,872 genes. Out of the DE genes, maker-Scaffold_2698-augustus-gene-0.8 (2698) showed the highest level of log2 fold change (LFC) of 3.52, being up-regulated in the resistance-breaking biotype (Figure S6).

Differential gene expression of avirulent and virulent *N. ribisnigri* biotypes

A principal component analysis (PCA) highlighted some differences between the *N. ribisnigri* biotypes (Figure 2). The PCA indicated that

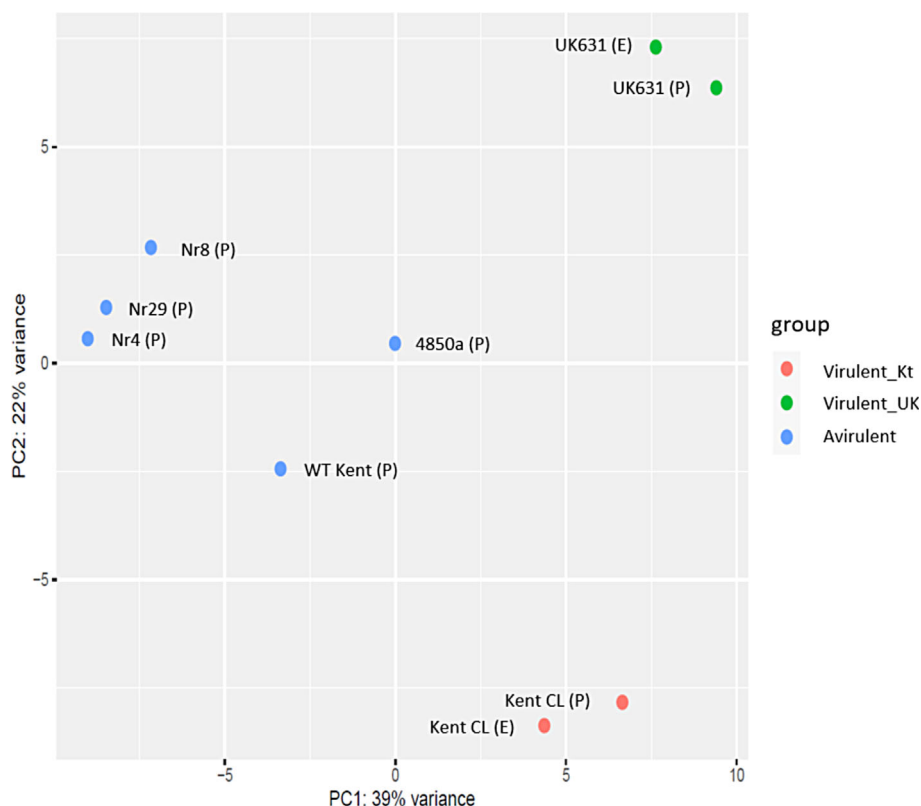


FIGURE 2 Principal component analysis of the RNA-read count data (*featurecounts*) for each *Nasonovia ribisnigri* biotype grouped into three groups (avirulent (blue points), virulent (red points) (Kent CL) and virulent (green points) (UK631). Susceptible denotes all five *N. ribisnigri* cultures that are unable to feed on lettuce containing the *Nr*-locus (avirulent). (E), cv. Eluarde, virulent; (P), cv. Pinokkio, avirulent.

there was little variation within the virulent biotypes feeding on either lettuce containing the *Nr*-locus (Eluarde) or lettuce without the gene (Pinokkio) but with a notable variation (particularly in PC2) between different resistance-breaking biotypes. A close grouping of three avirulent biotypes (Nr4, Nr8 and Nr29) was apparent, which are all *N. ribisnigri* cultures known to be unable to feed on lettuce containing the *Nr*-locus but have some resistance to pyrethroids or other insecticides. The remaining two avirulent biotypes (4850a and WT Kent, blue points in Figure 2) are loosely clustered together and are biotypes that are unable to feed on lettuce containing the *Nr*-locus, nor do they have any insecticide resistance.

To produce a robust and tractable dataset for analysis, the initial 689 DE genes ($p < 0.05$) were further refined by increasing the p -value cut-off to $p < 0.00001$. This revealed the 32 DE genes of highest significant difference between the virulent and avirulent biotypes. The RNA-seq analysis identified clusters of up- and down-regulated genes between the two conditions, highlighted using a heatmap (Figure 3).

Both resistance-breaking (Rb) biotypes were grouped together according to their origin (Rb1 and Rb2-Kent_CL; Rb3 and Rb4-UK631) and had similar gene expression levels despite feeding on lettuce with and without the *Nr*-locus. Out of the top 32 DE genes, one of the avirulent biotypes (SUS4-4850a) was down-regulating two of the DE gene clusters that were up-regulated in the other avirulent biotypes. Additionally, there was little differentiation in some of the DE genes down-regulated by the other avirulent biotypes.

The RNA-seq analysis revealed that 13 DE genes were up-regulated in the resistance-breaking biotypes, with the remaining 19 being down-regulated (Table 2). 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 2). The 19 down-regulated DE genes are also loosely clustered into four main groups.

A possible candidate gene responsible for the *Nr*-locus in resistance-breaking *N. ribisnigri*

Comparing the 689 DE genes ($p < 0.05$) between the WT and Rb in a 3-way Venn diagram, identified only one DE shared gene (gene 2698) between both Rb biotypes (Figure 4a). Due to both the log fold change and being the only gene shared between both virulent *N. ribisnigri* biotypes, this gene was further investigated as a potential candidate responsible for the *Nr*-locus. This gene was located on scaffold 2698; it was a sequence coding for 284 amino acids with a predicted protein size of 32.51 kDa and was identified as a hypothetical protein in *Aphis glycines* (A0A6G0SZ22-APHGL) using UniProt (Table 2). Using deep learning protein prediction software, gene 2698

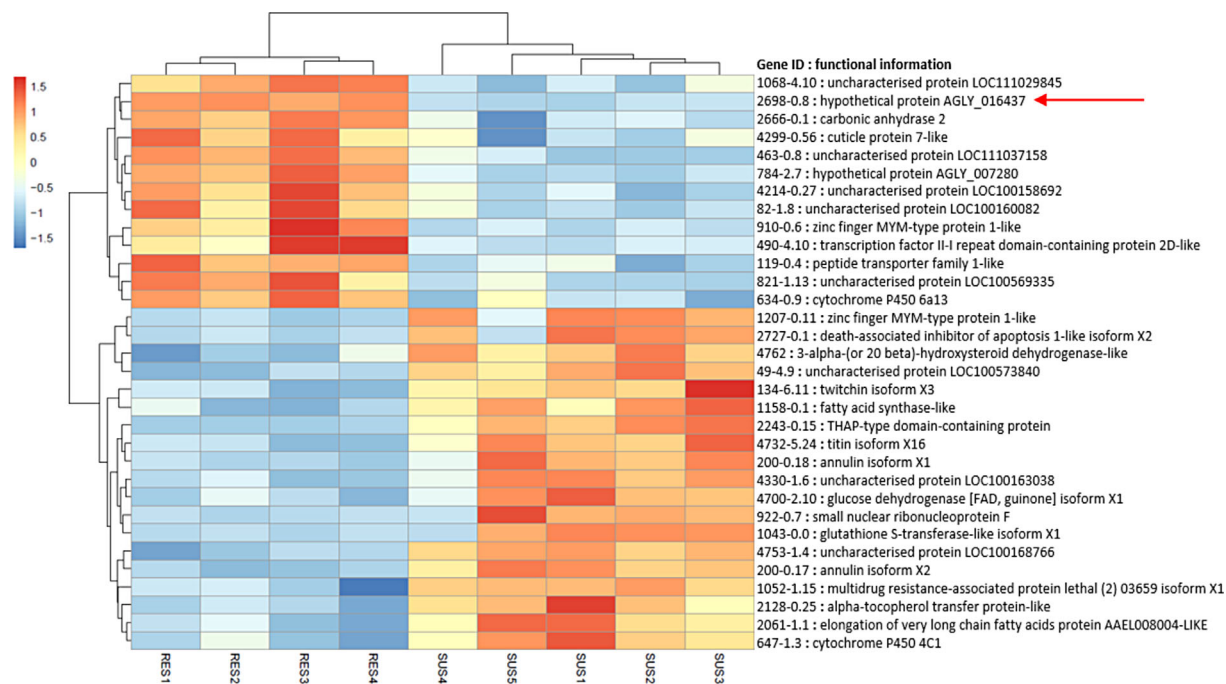


FIGURE 3 Heatmap of the top 32 differentially expressed (DE) genes ($p < 0.00001$) between *Nasonovia ribisnigri* resistance-breaking (virulent) and susceptible (avirulent) biotypes, showing clear division. RES = resistance-breaking (virulent); SUS = susceptible (avirulent). SUS1 = Nr4, SUS2 = Nr8, SUS3 = Nr29, SUS4 = 4850a, SUS5 = WT Kent, RES1 = Kent_CL (Eluarde), RES2 = Kent_CL (Pinokkio), RES3 = UK631 (Eluarde), RES4 = UK631 (Pinokkio). Colours of expression levels correspond to log fold change. Gene 2698 is indicated by a red arrow in the heat map.

has predicted localisations in both the cytoplasm and the nucleus (Figure S7). The mRNA encoding 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 identified for gene 2698 have a high level of associated transcripts for both resistance-breaking biotypes and a much lower number of transcripts for the avirulent biotypes, for all conditions and replicates (Figure 4b).

Two SNPs were identified in gene 2698 in nucleotide positions 9 and 50 at the start of the protein sequence. In the first single nucleotide polymorphism (SNP), from the WT_Kent, avirulent biotype, the allele was homozygous for a T residue (97% of 34 sequence reads) in the third-codon position of serine (position 9) residue AGT, whereas the Kent_CL (fed on resistant lettuce) virulent (resistance-breaking) biotype was homozygous for a C residue (100% of 248 sequence reads), which converted the serine to a threonine AGC. In the second SNP, from the WT_Kent avirulent biotype, the allele was homozygous for an A residue (96% of 43 sequence reads) in the first-codon position of methionine (position 50) residue ATG, whereas Kent_CL (fed on resistant lettuce) resistance-breaking biotype was homozygous for a T (100% of 290 sequence reads), which converted the methionine to a leucine TTG. These SNPs were not present in the low number of RNA-seq counts for the WT strains. Additionally, there were two SNPs identified on the WT strains (WT Kent) ~300 bp upstream of the promotor sequence for gene 2698. These SNPs are located 284 and 243 bp upstream of gene 2698, respectively. In the first SNP,

from the WT_Kent avirulent biotype, the allele was homozygous for a T residue (97% of 35 sequence reads) in the first- and second-codon position, which converts a valine (GTG) and cysteine (TGT) to an alanine (GCG) and arginine (CGT) in the resistance-breaking biotype, respectively. The second SNP, from the WT_Kent avirulent biotype, the allele was homozygous for an A residue (96% of 45 sequence reads) in the first-codon position which converts a methionine (ATG) to a leucine (TTG) in the resistance-breaking biotype. These had good RNA-seq coverage and were high quality with no strand bias, suggesting that they are not an artefact.

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

Three previously used housekeeping genes (HKGs) (ribosomal protein S9 [RPS9], RPS18, RPL13) were identified in 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) (Figure 5). There was little variation between Ct values for both conditions (*N. ribisnigri* avirulent biotype feeding on host plants without the Nr-locus and *N. ribisnigri* feeding on host plants containing the Nr-locus) for all three HKGs. The Real-Time Quantitative Reverse Transcription PCR (qRT-PCR) analysis was highly optimised and had amplification efficiencies (E) for primers from 97.28% to 99.31%, and R^2 were >0.98 (Figure S8). Genom highlighted that all

TABLE 2 Differentially expressed (DE) genes in virulent (resistance-breaking) and avirulent biotypes of *Nasonovia ribisnigri* (Nr:1) feeding on both resistant (Nr-locus present) and susceptible (Nr-locus absent) lettuce host plants.

Gene	Log2 fold change	p-Value	Description
Up-regulated			
2698-0.8	11.47348	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	Lucosa dehydrogenase [FAD, quinone] isoform X1
49-4.9	-0.894127	2.51E-08	Uncharacterised protein LOC100573840

Note: 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 log2 fold change (LFC) estimates. Fold change values were considered significant if $p < 0.0001$.

three HKGs were stably expressed across both virulent and avirulent *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 R^2 of >0.98 (Figure S9). Only one primer pair was used for the gene expression analysis.

The gene 2698 was highly expressed in the resistance-breaking *N. ribisnigri* biotypes compared with the avirulent *N. ribisnigri* biotype (Figure 6).

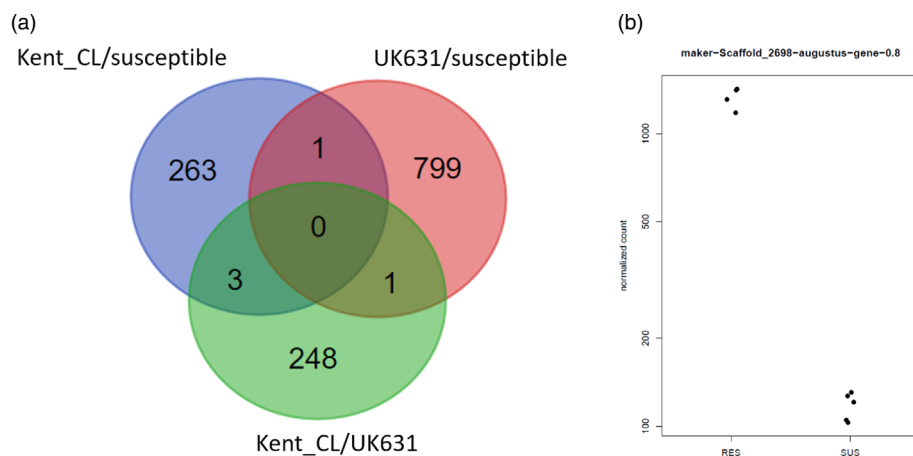


FIGURE 4 (a) Three-way Venn diagram of the differentially expressed genes in three contrasts (Kent_CL/susceptible; UK631/susceptible; Kent_CL/UK631) identified by DESeq2. Differentially expressed genes were considered significant and used in the diagram if $p < 0.05$. (b) RNA-sequencing counts for gene 2698 between resistance-breaking (Rb) and susceptible (SUS) biotypes. Susceptible comprise all avirulent biotypes and both Kent_CL and UK631 are virulent.

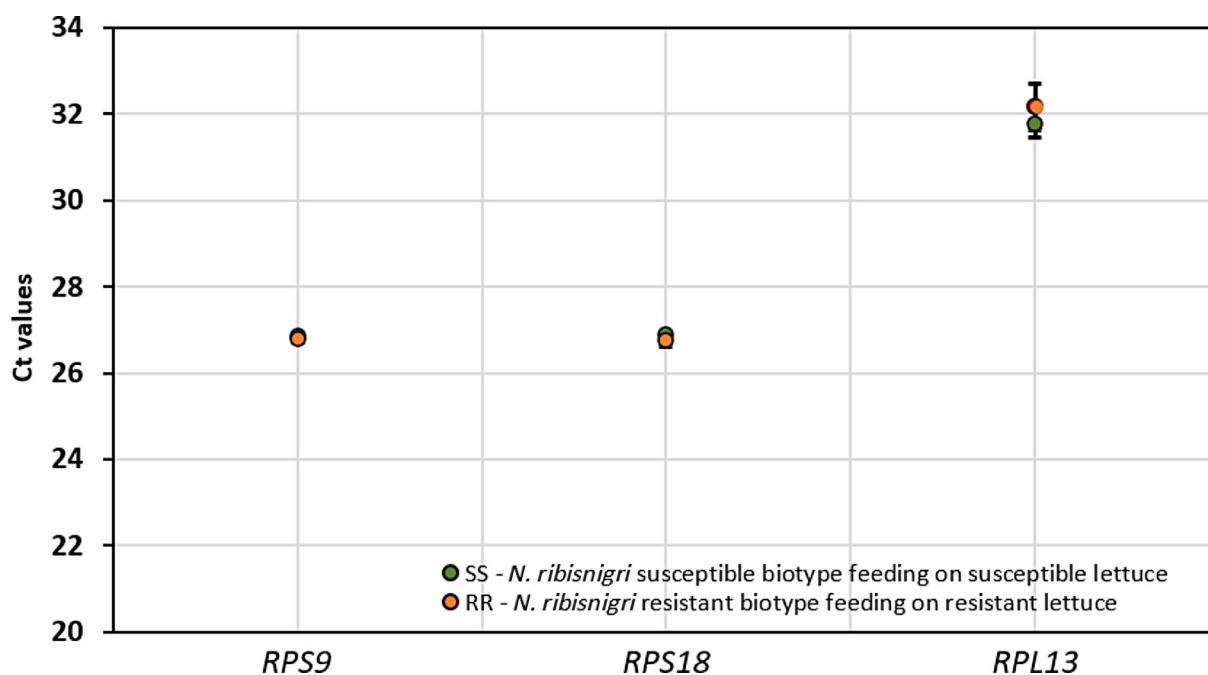


FIGURE 5 Ct values (\pm SE) obtained for the three different housekeeping genes (HKGs) under two experimental conditions. Green circles represent avirulent *Nasonovia ribisnigri* feeding on susceptible lettuce (without the Nr-locus), orange circles represent virulent *N. ribisnigri* feeding on resistant lettuce (with the Nr-locus). Three biological and three technical replicates were used to validate the HKGs. See Supporting Information for calculations.

Statistical analysis (t-test-MeV package) confirmed that WT_Kent (susceptible *N. ribisnigri* fed on susceptible lettuce - SS) was significantly different ($F = 96.54$, $p < 0.001$) from Kent_CL (resistance-break *N. ribisnigri* fed on resistant lettuce - RR), UK631 (RR) and UK631 (resistance-breaking *N. ribisnigri* fed on susceptible lettuce - RS). Resistance-breaking *N. ribisnigri* biotypes fed on lettuce containing the Nr-locus (cv. Eluarde) and without the gene (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-locus (Figure 6).

DISCUSSION

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

The differential expression analyses results indicated that one gene was significantly up-regulated in the resistance-breaking *N. ribisnigri* biotypes (gene 2698). Despite other significantly up- and down-regulated DE genes in the analysis, gene 2698 had the smallest log10

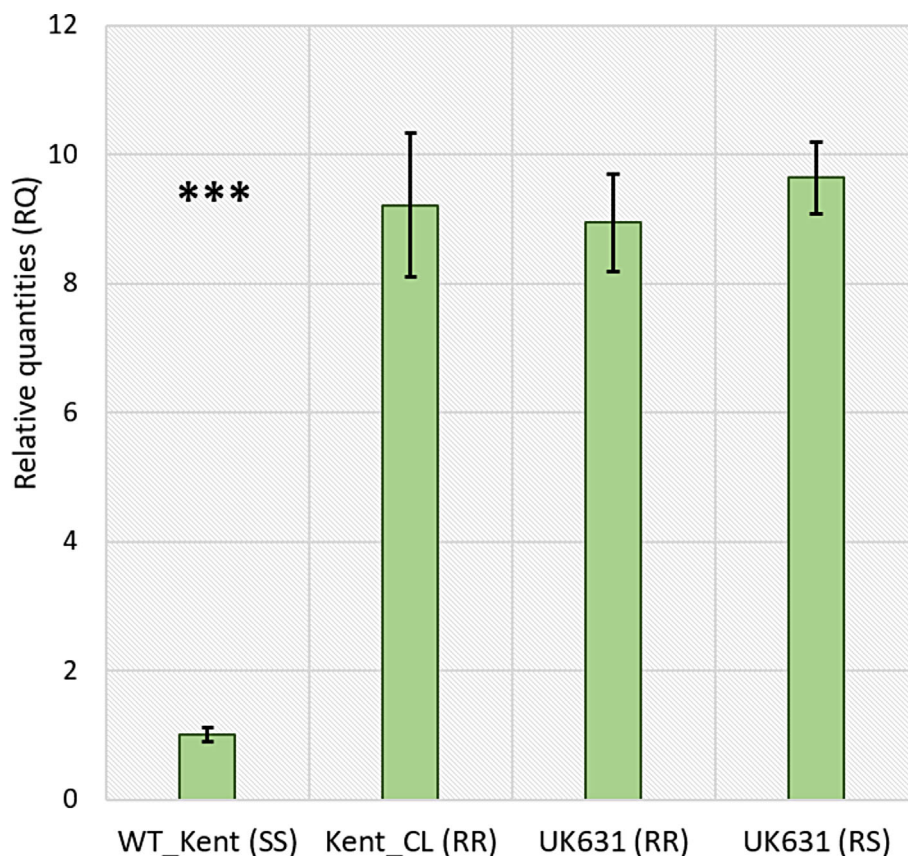


FIGURE 6 Comparison of gene-2698 expression levels using qRT-PCR between four conditions of *Nasonovia ribisnigri* cultures feeding on resistant (cv. Eluarde) and susceptible (cv. Pinokkio) lettuce cultivars. Both Kent_CL and UK631 are resistance-breaking biotypes able to feed on resistant lettuce cultivars. WT Kent is an avirulent biotype unable to feed on the resistant lettuce cultivars containing the Nr-locus. RR, resistance-breaking *N. ribisnigri* fed on resistant lettuce cultivar (cv. Eluarde); RS, resistance-breaking *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 ($F = 96.54$, $p < 0.001$). *** $p < 0.001$.

adjusted p -value (Figure S6). Further analysis confirmed that gene 2698 was the only DE gene that both resistance-breaking biotypes shared (Figure 4a) and was the most promising candidate gene conferring resistance to the Nr-locus in the lettuce host plant. The qRT-PCR confirmed that this gene is significantly up-regulated in the resistance-breaking biotypes (Figure 6). The monogenic Nr resistance 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). The results of the present study therefore gives further support to the hypothesis that a gene-for-gene resistance exists between the Nr-locus in the lettuce host plant and a corresponding gene (gene 2698) in the resistance-breaking *N. ribisnigri* biotype. Unfortunately, the function of the gene identified in this study is not known in any other species, and further work will be required to determine its biological function.

Interestingly, it was shown in the qRT-PCR results (Figure 6) that gene 2698 is up-regulated in virulent biotypes when they are feeding on lettuce not containing the Nr-locus, as well as lettuce containing the Nr-locus. This suggests that gene 2698 is constitutively expressed regardless of the presence or absence of the Nr-locus in the lettuce

host plant. Due to the unknown location of the mechanism/s of resistance within *N. ribisnigri*, whole organisms were used for the RNA extractions and subsequent RNA-seq analysis. It has been shown that using whole organisms for RNA-seq studies can produce false negatives and difficult to interpret positive signals, when the local expression pattern is variable (Johnson et al., 2013). Every effort was made to reduce the likelihood of false negative results, both with the experimental design and data analysis, but there are some limitations by using whole organisms as opposed to tissue-specific RNA extractions.

The overreliance and overuse of a single resistance gene in lettuce (Nr-locus) has resulted in the evolution of resistance-breaking *N. ribisnigri* biotypes (Thabuis et al., 2014). Combining multiple resistance genes or deleting susceptibility loci greatly reduces the selection pressure and can increase the longevity of resistant or tolerant cultivars (Delmas et al., 2016). For example, if a pest evolves and gains the ability to break one control gene, the plant is still protected by the remaining resistance genes. To provide more durable resistance in crop plants, it has been suggested to 'pyramid' resistance genes with mutations in susceptibility genes (Stuart, 2015). Pyramiding resistance genes in this way generally results in an increase durability of resistance in 'pyramid crops' compared to monogenetic resistance

(Mundt, 2018). Pyramiding resistance genes could be a more successful approach in controlling *N. ribisnigri* outbreaks in the future but will require the identification of new resistance loci. However, pyramiding resistance genes in one cultivar can still be overcome by pest insects if no other measures are applied to reduce the selection pressure and therefore is best used as a component of integrated pest management (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 avirulent *N. ribisnigri* biotypes were unable to feed on lettuce containing the Nr-locus (ten Broeke et al., 2013). The use of excised lettuce leaves in experiments resulted in a loss of resistance which suggested that the feeding mechanism is 'mobile' (Liu & 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 not to be toxic, as normal feeding behaviour and growth is resumed when avirulent aphids are transferred to lettuce cultivars not containing the Nr-locus (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-locus 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; Le Trionnaire et al., 2019), could be incorporated to mediate gene knockout of the SNPs identified in gene-2698 for the *N. ribisnigri* resistance-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-locus (loss-of-function strategy) (Housden et al., 2017). In addition, it could be possible to complement this approach with a gain-of-function, in which avirulent *N. ribisnigri* could be modified to over express gene-2698 and therefore determine if they are able to feed on lettuce containing the Nr-locus (Zimmer et al., 2018).

The two SNPs identified in gene 2698 in nucleotide positions 9 and 50 are unlikely the cause for the resistance-breaking biotype to overcome the Nr-locus in the lettuce host plant as these do not cause significant changes to the mRNA's expression. Interestingly, the SNP for the WT is in the 5' untranslated region (UTR) of the mRNA, located 243 bp upstream of the start codon for gene 2698 is homozygous for an A which converts a leucine to a methionine (start codon). This start codon located in the WT promotor could result in the formation of an upstream open reading frame, which could subsequently affect expression of the protein and prevent WT *N. ribisnigri* from feeding on lettuce containing the Nr-locus. Further research into gene 2698 is required to confirm these findings, but it is an interesting discovery.

Construction of the first reference genome for *N. ribisnigri*

This study presents the first genome assembly and annotation for *N. ribisnigri*, both for a virulent and virulent (resistance-breaking)

biotype. The avirulent (Nr:0) genome had the highest completeness score (94.2% complete and single copy) and the virulent biotype (Nr:1) had the most contiguous assembly with a completeness of 92.9% complete and single copy (Table 1). These scores represent a high standard of completeness, compared with previously published aphid genomes (Figure 1) (Biello et al., 2020; Li et al., 2019; Mathers et al., 2020; Shahid et al., 2021). The removal of both the aphid symbiont (*Buchnera aphidicola*) and host plant (*L. sativa*), overall, further improved both the assembly statistics and completeness of the genomes by ~1%.

From the assembled genomes, there was a discrepancy between genome size of the Nr:0 and Nr:1 biotypes (Table 1). The genome assemblies for UK631 (Nr:0) and Ely (Nr:1) biotypes were similar in genome size (~366 Mb), N50, number of scaffolds and completeness (BUSCO), but this is likely to be a result of lower assembly qualities. The higher quality genome assemblies of Nr8 (Nr:0), WT Kent (Nr:0) and Kent CL (Nr:1) in fact differ in genome size, 423, 423 and 384 Mb, respectively. The differences observed in genome size is likely a result of the variation in assembly quality as Nr8 had fewer duplications and missing genes than Kent CL. The higher N50 and fewer scaffolds in the Nr8 genome would have also attributed to the higher estimated genome size. Additionally, the K-mer coverage for the Nr8 was higher with fewer errors than Kent CL and supports this conclusion. Therefore, an estimated genome size of 423 Mb for *N. ribisnigri* is more accurate.

The genome sizes of both the avirulent and virulent *N. ribisnigri* biotypes were shown to be comparable to other aphid species, although the estimation for the resistance-breaking biotype is likely low (Figure 1). The pea aphid (*A. pisum*) genome is estimated to be a size of ~446.6 Mb (The International Aphid Genome Consortium 2010), a draft assembly of the soybean aphid (*A. glycines*) (Matsumura) has been estimated at 317.1 Mb (Wenger et al., 2020), the corn leaf aphid (*Rhopalosiphum maidis*) (Fitch) near-complete (95.8%) genome is estimated at 321 Mb (Chen et al., 2019), Russian wheat aphid (*Diuraphis noxia*) (Kurdjumov) is 421 Mb (Nicholson et al., 2015), the peach-potato aphid (*Myzus persicae*) (Sulzer) is estimated to be 347 Mb (Mathers et al., 2017) and an improved bird-cherry oat aphid (*Rhopalosiphum padi*) (Linnaeus) genome is estimated to be of 321 Mb (Morales-Hojas et al., 2020). The number of genes identified from IPS in both the avirulent and virulent were ~30,000 and ~28,800, respectively. The number of genes identified by IPS in existing aphids (BIPAA, 2021) range between 19,182 (*A. glycines*) and 31,604 (*M. persicae*), and it is likely that the estimation produced by Augustus is reasonable. The transcript prediction of ~77,000 is likely an over estimation and an artefact of the assembly process.

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 virulent (resistance breaking) and avirulent *N. ribisnigri* biotypes provides researchers with a tool to study *N. ribisnigri* at much greater resolution. Additionally, elucidating a potential candidate gene of resistance in the resistance-breaking biotype will enable targeted approaches to thoroughly investigate this elusive mechanism of resistance.

EXPERIMENTAL PROCEDURES

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 *L. sativa* cv. 'Auvona RZ', a variety that does not contain the *Nr-locus* 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 16:8 h (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-locus*) at G's Fresh (Cambridgeshire, UK) in 2018. All clones were determined and validated to be virulent or avirulent by subsequent culturing (same conditions as above) on a resistant (*Nr*) variety (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 (Figure S10). All *N. ribisnigri* lines derive 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. A total of ~20 individuals from each isogenic line were used to obtain enough high molecular weight DNA for sequencing.

DNA extraction

N. ribisnigri high molecular weight (HMW) DNA was extracted using a QIAGEN® Genomic-tip (20/G) extraction kit. 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 s and homogenised using a sterile pestle. Subsequently, 1 mL 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 min at 37°C, followed by the addition of Proteinase K (0.8 mg/mL) (removing any protein contaminants) and incubation for 2 h at 50°C. This solution was centrifuged for 20 min at 21,130 g to pellet the debris. The clarified lysate was transferred to a Genomic-tip and the solution was left to flow-through by gravity. The Genomic-tips were washed four times with 1 mL wash buffer and HMW DNA was finally eluted into 2 mL Eppendorf tubes 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 min at 21,130 g, washed with 70% ethanol, and centrifuged again at 10 min

at 50002347g). Finally, the DNA pellet was left to dry for ~1 min, resuspended in an appropriate volume of molecular grade water, and placed into an orbital shaker for 2 h, 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 downstream processes.

Sequencing

DNA that passed in-house quality checks (Qubit BR Assay Kit [Thermo Fisher 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. Previous Illumina sequence data were provided by Warwick Crop Centre from a resistance-breaking biotype (Kent CL), and this was incorporated into the assembly of the Kent CL genome. To complement the Illumina sequence data, long-read sequence data were generated using a MinION (Oxford NANOPORE Technologies) at Rothamsted Research. Two SpotON flow cells (R9.4.1) were used to produce both virulent and avirulent long-read sequence data. *N. ribisnigri* HMW DNA was extracted following the method above, and the Genomic DNA by Ligation SQK-LSK109 protocol was used to construct the library (Oxford NANOPORE Technologies). This protocol includes DNA repair and end-prep, adapter ligation and clean-up followed by priming and loading the SpotON flow cell. For each run, approximately 1 µg 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 resistance-breaking sequence data and Nr8 was used to produce the avirulent (wild type) sequence data. All reads have been submitted to the National Center for Biotechnology Information (NCBI) (BioProject ID: PRJNA857679) (Figures S11 and S12).

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). MaSuRCA (Maryland Super Read Cabog Assembler) was used to generate a hybrid assembly for each biotype with the short Illumina and long ONT reads. 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

virulent and avirulent *N. ribisnigri* genomes (<http://qb.cshl.edu/genomescope/genomescope2.0/>) (Figure S1). All genomes generated and presented in the manuscript have been submitted to the NCBI (BioProject ID: PRJNA857679) (Figure S11).

Removal of contaminants

To identify and remove sequence contamination by 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 were identified to be the host plant lettuce (*L. 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 identified using the NCBI Web BLAST (nucleotide) database. 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.

Assessment of *N. ribisnigri* genome completeness using BUSCO

Once all contaminants had been removed, 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).

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 (Hoff & Stanke, 2013), using the pea aphid (*A. pisum*) and the peach-potato aphid (*M. persicae*) as a training set. *M. 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 against the newly assembled *N. ribisnigri* genomes (resistance-breaking 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., 2021; Huerta-Cepas et al., 2019).

RNA-seq methods and differential gene expression analysis

RNA extraction and RNA-seq library preparation

RNA-seq data for *N. ribisnigri* were obtained from Warwick Crop Centre (Wellesbourne Campus, University of Warwick). Aphid cultures were maintained long-term on a lettuce cultivar Pinokkio, which does not contain the Nr-locus. Prior to raising aphids for sequence analysis, their expected ability to survive or not on the Nr-locus-containing cultivar (Eluarde) was verified from culture experiments. For RNA-seq analysis, two Rb aphid strains (Kent CL and UK631) were reared on both lettuce cultivars containing the Nr-locus (cv. Eluarde) and without (cv. Pinokkio). The WT aphids were only reared on Pinokkio (Table 3). For each biotype, a single apterous female was placed onto the various conditions listed in Figure S10. Once the aphid colonies had established, approximately 50 aphids (mixed apterae and alates) of each treatment were pooled for RNA extraction. Whole organisms were used due to the uncertainty of where the mechanism/s of resistance were located (e.g., stylet/salivary, metabolic or within the midgut). 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). Then, 100 bp paired-end sequencing was performed on an Illumina HiSeq platform at Wellcome Trust Centre for Human Genetics, Oxford, with the resulting read counts ranging from 12,634,191 to 37,227,165 per sample.

Quality control of the RNA-seq reads was conducted on all datasets with FastQC prior to further downstream analysis (Andrews, 2010; Wingett & Andrews, 2018). The newly assembled reference genome of *N. ribisnigri* avirulent strain (Nr8_123) was used for mapping the reads using HISAT2 software with default options.

TABLE 3 *Nasonovia ribisnigri* cultures that were used for the RNA-seq data extraction and sequencing.

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

Note: Nr:0 = *Nr*-locus avirulent; Nr:1 = *Nr*-locus avirulent; Pyrethroid R = pyrethroid resistant; Insecticide R = insecticide resistant; (S) = susceptible (does not contain *Nr*-locus); (R) = resistant (contains *Nr*-locus). For the differentially expressed (DE) gene analysis using DESeq2, RNA-seq data were grouped into three groups: blue (susceptible), orange (resistant_Kent) and green (resistant_UK). These data were used for the transcriptome assembly and RNA-seq analysis for differentially expressed (DE) gene analysis. Blue consists of the five avirulent lineages, orange consists of the virulent Kent CL lineage and, green consists of the virulent UK631 lineage used in the DESeq2 analysis and Venn diagram comparison.

Transcriptome assembly

After all reads were quality checked using FastQC (Wingett & Andrews, 2018), reads were aligned to the draft *N. ribisnigri* reference genome from the present study (avirulent, Nr8_123) using Hisat2 to create short read alignments (Table 1). 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 (avirulent, Nr8_123) and assembled using Trinity v2.5.1, using default settings (Figure S3A).

Differential gene expression analysis

The same RNA-seq data were used for the differential gene expression analysis (see above). RNA-seq data were aligned to the *N. ribisnigri* genome annotation (Nr8) using *featureCounts* (Liao et al., 2014). HISAT2 was used to align all RNA-seq data from all conditions (Table 3) to each gene in the *N. ribisnigri* genome annotation. *featureCounts* subsequently used the alignment file to create a gene count table for use in further downstream analysis. The DESeq2 package (Anders & Huber, 2010) was used to test for differential gene expression between three conditions which consisted of an avirulent complex of three lineages (susceptible to the *Nr*-locus in the host plant) and two lineages virulent to the *Nr*-locus able to overcome the *Nr*-locus (Table 3). These three groups were used in the DESeq2 analysis which used the RNA-seq data count tables in *featurecounts* (Figure 2). Since this study was focused on elucidating the potential mechanisms of resistance to the *Nr*-locus in lettuce only, *N. ribisnigri* cultures Nr4, Nr8 and Nr29 (which have a level of resistance to either pyrethroids or other insecticides) were grouped together with 4850a and WT Kent as the 'avirulent complex group', as they are all unable to feed on lettuce containing the *Nr*-locus. Data rows with no or

single counts were removed, and count data were log transformed (log2) prior to analysis. Gene expression log fold changes were considered significant if $p < 0.05$ (alpha). The apeglm package was utilised to generate more accurate LFC estimates (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 virulent (Nr:1) and avirulent (Nr:0) *N. ribisnigri* biotypes. The null hypothesis for each gene was that there is no differential expression between the avirulent or the two virulent groups (i.e., LFC = 0). A power analysis was performed on the count data produced from the RNA-seq to determine whether the number of replicates are sufficiently powered for the discovery of differentially expressed genes (Wu et al., 2015; Schurch et al., 2016). The power analysis was conducted in RStudio (4.2.3) using the PROspective Power Evaluation for RNAseq (PROPER), Dispersion Shrinkage for Sequencing data (DSS), DESeq2 and edgeR packages (R Core Team, 2020), and the RStudio code is provided in the Supporting Information (S16).

All DESeq2 analysis was conducted in RStudio (4.0.5) (R Core Team, 2020) (Supporting Information S17). The ggplot package was used to create a PCA plot of the experiment to visualise the overall similarities of the samples in the three groups (Figure 2). To visualise the significantly DE genes between the avirulent and virulent biotypes, a volcano plot package (EnhancedVolcano) was used (Figure S6). For the volcano plot only, both DE genes of the resistant (virulent) groups were pooled and plotted against the DE genes of the avirulent 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, followed by a more stringent p -value of 0.00001, to reduce the number of identified DE genes to a tractable and highly significant dataset. A Venn diagram was created using webtools to visually compare the differentially expressed genes between the avirulent group and the two virulent groups (<https://bioinformatics.psb.ugent.be/webtools/Venn/>).

Validation of RNA-seq data by qRT-PCR

Aphid and plant cultures

qRT-PCR was incorporated to confirm the results of the RNA-seq data. Two avirulent cultures (Nr:0) and two virulent 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, and each culture was triplicated to create three biological replicates per treatment (12 conditions) (Figure S10). Each culture was maintained in previous laboratory conditions (described above) until ~50 individuals were produced. Each isogenic line was collected in 1.5 mL Eppendorf tubes containing RNeasy Protect (Qiagen) and frozen at -80°C prior to RNA extraction.

Reference gene discovery

The reference genes, *RPS9*, ribosomal protein S18 (*RPS18*) and ribosomal protein L13a (*RPL13*), were used as internal controls for the qRT-PCR. The *RPS9* gene was chosen based on previous work which demonstrated that it had a stable expression among soybean aphids (*Aphis glycines* (Matsumura)) on avirulent and virulent host plants (Bansal et al., 2013). The ribosomal proteins *RPS18* and *RPL13* have been shown to be good reference genes for 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 & Hoover, 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>) (Figure S12).

Identification and validation differentially expressed genes in *N. ribisnigri*

One significantly differentially expressed (DE) gene of interest (gene-2698) 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). 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>) (Figure S13). Methodology for qRT-PCR analysis on

N. ribisnigri to validate RNA-seq data can be found in the supporting information (S14). A t-test (Modelling of Extreme Values (MeV) package) was used to test for differences between each condition, with a Bonferroni corrected *p*-value of 0.008 or less was considered statistically significant.

Reference genes stability analysis

The software algorithm GeNorm was used to determine the stability of the three candidate reference genes (*RPS9*, *RPS18* and *RPL13a*) (Andersen et al., 2004; Vandesompele et al., 2002), using three biological replicates (Figure S10) and three technical replicates to validate the genes. The inputs for GeNorm were the raw expression values for each gene (using the equation, $2^{(-\Delta\text{Ct})}$). 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 is not a suitable reference gene (Figure S15). Fold change and relative expression level were determined using the comparative Ct method ($2^{-\Delta\Delta\text{Ct}}$) (Schmittgen & Livak, 2008). Statistical analysis was performed using a t-test (MeV package).

AUTHOR CONTRIBUTIONS

Dion Garrett: Conceptualization; investigation; writing – original draft; methodology; validation; visualization; writing – review and editing; software; formal analysis; data curation; resources. **Graham Teakle:** Methodology; validation; visualization; writing – review and editing; software; resources; supervision; project administration. **Rosemary Collier:** Visualization; writing – review and editing; supervision; project administration. **James R. Bell:** Visualization; writing – review and editing; project administration; supervision. **Sergio Cerezo-Medina:** Software; validation; visualization; methodology; resources; formal analysis. **Ramiro Morales-Hojas:** Investigation; methodology; visualization; writing – review and editing; software; formal analysis; project administration; supervision.

ACKNOWLEDGEMENTS

The authors are grateful to collaborators and staff who have contributed from the Rothamsted Insect Survey. We would also like to thank the Alison Jackson and Richard Stark (University of Warwick) for providing the RNA-seq data used for the assembly of the *N. ribisnigri* transcriptome and differential gene expression analysis (DESeq2) and Marian Elliott for the *N. ribisnigri* cultures used in this project. No ethics approval was required for this research and the authors declare no competing interests. All sequence reads and genome assemblies have been submitted to the National Centre for Biotechnology Information (NCBI) (BioProject ID: PRJNA857679) (Tables S3 and S4).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All sequence reads and genome assemblies have been submitted to the National Centre for Biotechnology Information (NCBI) (BioProject ID: PRJNA857679) (Tables S8 and S9).

ORCID

Dion Garrett  <https://orcid.org/0000-0001-7350-2763>

Graham Teakle  <https://orcid.org/0000-0003-1410-0806>

James R. Bell  <https://orcid.org/0000-0003-1640-6120>

Sergio Cerezo-Medina  <https://orcid.org/0000-0001-9234-8669>

Ramiro Morales-Hojas  <https://orcid.org/0000-0002-4120-5964>

REFERENCES

- Aarts, R., Schut, J.W., Driessen, R. & Reinink, K. (1999) Integrated control of aphids on lettuce varieties resistant to *Nasonovia ribisnigri*. *Mededelingen-Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen*, 64(3; PART A), 11–16.
- Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Čech, M. et al. (2018) The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, 46(W1), W537–W544.
- Anders, S. & Huber, W. (2010) Differential expression analysis for sequence count data. *Nature Precedings*, 11, R106. Available from: <https://doi.org/10.1093/nar/gky379.46.W537.W544>
- Andersen, C.L., Jensen, J.L. & 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*, 64(15), 5245–5250.
- Andrews, S. (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Andrews, S. (2019) FastQC A Quality Control Tool for High Throughput Sequence Data. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Bansal, R., Mian, M.R., Mittapalli, O. & 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), 286–295. Available from: <https://doi.org/10.1017/S0007485312000697>
- Barzman, M., Bärberi, P., Birch, A.N.E., Boonekamp, P., Dachbrodt-Saaydeh, S., Graf, B. et al. (2015) Eight principles of integrated pest management. *Agronomy for Sustainable Development*, 35, 1199–1215. Available from: <https://doi.org/10.1007/s13593-015-0327-9>
- Biello, R., Singh, A., Godfrey, C.J., Fernández, F.F., Mugford, S.T., Powell, G. et al. (2020) A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Molecular Ecology Resources*, 21(1), 316–326. Available from: <https://doi.org/10.1111/1755-0998.13258>
- BIPAA. (2021) Bioinformatics Platform for Agroecosystem Arthropods. Available at: <https://bipaa.genouest.org/is/> [Accessed: 23th June 2020].
- Blackman, R.L. & Eastop V.F. (2017) *Taxonomic issues, in Aphids as Crop Pests*, ed. by van Emden HF and Harrington R. CABI, Wallingford, 1–29.
- Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A. et al. (2021) The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, 49(D1), D344–D354.
- Campbell, M.S., Holt, C., Moore, B. & Yandell, M. (2014) Genome annotation and curation using MAKER and MAKER-P. *Current Protocols in Bioinformatics*, 48(1), 4–11. Available from: <https://doi.org/10.1002/0471250953.bi0411s48>
- Chen, W., Shakir, S., Bigham, M., Richter, A., Fei, Z. & Jander, G. (2019) Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). *GigaScience*, 8, 33. Available from: <https://doi.org/10.1093/gigascience/giz033>
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N. et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121), 819–823.
- Delmas, C.E., Fabre, F., Jolivet, J., Mazet, I.D., Richart Cervera, S., Deliere, L. and Delmotte, F. (2016) Adaptation of a plant pathogen to partial host resistance: selection for greater aggressiveness in grapevine downy mildew. *Evolutionary Applications*, 9(5), 709–725. Available at: <https://doi.org/10.1111/eva.12368>
- Diaz, B.M., Barrios, L. & Fereres, A. (2012) Interplant movement and spatial distribution of alate and apterous morphs of *Nasonovia ribisnigri* (Homoptera: Aphididae) on lettuce. *Bulletin of Entomological Research*, 102(4), 406–414. Available from: <https://doi.org/10.1017/S0007485311000745>
- Eenink, A.H., Groenwold, R. & 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*, 31, 291–299. Available from: <https://doi.org/10.1007/BF00021643>
- Field, L.M., Bass, C., Davies, T.G.E., Williamson, M.S. & Zhou, J.J. (2017) Aphid genomics and its contribution to understanding aphids as crop pests. In: *Aphids as crop pests*. Wallingford UK: CABI, pp. 37–49. Available from: <https://doi.org/10.1079/9781780647098.0037>
- Flor, H.H. (1971) Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, 9, 275–296.
- Francis, W.R., Christianson, L.M., Kiko, R., Powers, M.L., Shaner, N.C. & Haddock, S.H. (2013) A comparison across non-model animals suggests an optimal sequencing depth for de novo transcriptome assembly. *BMC Genomics*, 14(1), 1–12. Available from: <https://doi.org/10.1186/1471-2164-14-167>
- Gao, L.L., Klingler, J.P., Anderson, J.P., Edwards, O.R. & Singh, K.B. (2008) Characterization of pea aphid resistance in *Medicago truncatula*. *Plant Physiology*, 146(3), 996–1009. Available from: <https://doi.org/10.1104/pp.107.111971>
- Hatchett, J.H. & 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*, 63(5), 1400–1407. Available from: <https://doi.org/10.1093/aesa/63.5.1400>
- Himabindu, K., Suneetha, K., Sama, V.S.A.K. & 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*, 174, 179–187. Available from: <https://doi.org/10.1007/s10681-009-0106-2>
- Hoff, K.J. & Stanke, M. (2013) WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research*, 41(W1), W123–W128. Available from: <https://doi.org/10.1093/nar/gkt418>
- Hough, G.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. et al. (2017) Loss-of-function genetic tools for animal models: cross-species and cross-platform differences. *Nature Reviews Genetics*, 18(1), 24–40. Available from: <https://doi.org/10.1038/nrg.2016.118>
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H. et al. (2019) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1), D309–D314. Available from: <https://doi.org/10.1093/nar/gky1085>
- Huson, D.H., Auch, A.F., Qi, J. & Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Research*, 17(3), 377–386. Available from: <https://doi.org/10.1101/gr.5969107>
- International Aphid Genomics Consortium. (2010) Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PloS Biology*, 8, e100313. Available from: <https://doi.org/10.1371/journal.pbio.1000313>
- Jiang, X., Zhang, Q., Qin, Y., Yin, H., Zhang, S., Li, Q. et al. (2019) A chromosome-level draft genome of the grain aphid *Sitobion miscanthi*. *GigaScience*, 8(8), giz101. Available from: <https://doi.org/10.1093/gigascience/giz101>

- Johnson, B.R., Atallah, J. & Plachetzki, D.C. (2013) The importance of tissue specificity for RNA-seq: highlighting the errors of composite structure extractions. *BMC Genomics*, 14(1), 1–8. Available from: <https://doi.org/10.1186/1471-2164-14-586>
- Katis, N., Tsitsipis, J.A., Stevens, M. & Powell, G. (2007) In: van Emden, H.F. & Harrington, R. (Eds.) *Transmission of plant viruses. Aphids as crop pests*. Wallingford: CAB, pp. 353–390. Available from: <https://doi.org/10.1079/9780851998190.0353>
- Kerr, A. (1987) The impact of molecular genetics on plant pathology. *Annual Review of Phytopathology*, 25(1), 87–110. Available from: <https://doi.org/10.1146/annurev.py.25.090187.000511>
- Koramutla, M.K., Amini, R. & Bhattacharya, R. (2016) Comprehensive evaluation of candidate reference genes for qRT-PCR studies of gene expression in mustard aphid, *Lipaphis erysimi* (Kalt). *Scientific Reports*, 6, 1–10. Available from: <https://doi.org/10.1038/srep25883>
- Le Trionnaire, G., Tanguy, S., Hudaverdian, S., Gléonnet, F., Richard, G., Cayrol, B. et al. (2019) An integrated protocol for targeted mutagenesis with CRISPR-Cas9 system in the pea aphid. *Insect Biochemistry and Molecular Biology*, 110, 34–44. Available from: <https://doi.org/10.1016/j.ibmb.2019.04.016>
- Li, B., Bickel, R.D., Parker, B.J., Saleh Ziabari, O., Liu, F., Vellichirammal, N.N. et al. (2020) A large genomic insertion containing a duplicated follistatin gene is linked to the pea aphid male wing dimorphism. *eLife*, 9, e50608. Available from: <https://doi.org/10.7554/eLife.50608>
- Li, Y., Park, H., Smith, T.E. & Moran, N.A. (2019) Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Molecular Biology and Evolution*, 36, 2143–2156. Available from: <https://doi.org/10.1093/molbev/msz138>
- Liao, Y., Smyth, G.K. & Shi, W. (2014) FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30, 923–930. Available from: <https://doi.org/10.1093/bioinformatics/btt656>
- Liu, Y.B. & McCreight, J.D. (2006) Responses of *Nasonovia ribisnigri* (Homoptera: Aphididae) to susceptible and resistant lettuce. *Journal of Economic Entomology*, 99(3), 972–978. Available from: <https://doi.org/10.1093/jee/99.3.972>
- Luethy, R. & Hoover, C. (2004) Hardware and software systems for accelerating common bioinformatics sequence analysis algorithms. *Drug Discovery Today: BIOSILICO*, 2(1), 12–17. Available from: [https://doi.org/10.1016/S1741-8364\(04\)02383-2](https://doi.org/10.1016/S1741-8364(04)02383-2)
- Lv, W., Du, B., Shangguan, X., Zhao, Y., Pan, Y., Zhu, L. et al. (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*, 15(1), 1–16. Available from: <https://doi.org/10.1186/1471-2164-15-674>
- Mathers, T.C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S.T., Baa-Puyoulet, P. et al. (2017) Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biology*, 18(1), 1–20. Available from: <https://doi.org/10.1186/s13059-017-1202-6>
- Mathers, T.C., Mugford, S.T., Hogenhout, S.A. & Tripathi, L. (2020) Genome sequence of the banana aphid, *Pentalonia nigronervosa* Coquerel (Hemiptera: Aphididae) and its symbionts. *G3: genes, genomes*, 10(12), 4315–4321. Available from: <https://doi.org/10.1534/g3.120.401358>
- Morales-Hojas, R., Gonzalez-Urriarte, A., Alvira Iraizoz, F., Jenkins, T., Alderson, L., Kruger, T. et al. (2020) Population genetic structure and predominance of cyclical parthenogenesis in the bird cherry-oat aphid *Rhopalosiphum padi* in England. *Evolutionary Applications*, 13(5), 1009–1025. Available from: <https://doi.org/10.1111/eva.12917>
- Mundt, C.C. (2018) Pyramiding for resistance durability: theory and practice. *Phytopathology*, 108(7), 792–802. Available from: <https://doi.org/10.1094/PHYTO-12-17-0426-RVW>
- Nault, L.R. (1997) Arthropod transmission of plant viruses: a new synthesis. *Annals of the Entomological Society of America*, 90(5), 521–541. Available from: <https://doi.org/10.1093/aesa/90.5.521>
- Nicholson, S.J., Nickerson, M.L., Dean, M., Song, Y., Hoyt, P.R., Rhee, H. et al. (2015) The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC Genomics*, 16(1), 1–16. Available from: <https://doi.org/10.1186/s12864-015-1525-1>
- Niemann, J.U., Menssen, M. & Poehling, H.M. (2022) Reducing initial aphid infestation by use of coloured mulch foils and newly developed biodegradable spray-films. *Journal für Kulturpflanzen*, 74(3–4), 49–62.
- OmicsBox—Bioinformatics made easy. (2019) BioBam Bioinformatics. Version 1.4.11. Available at: www.biobam.com/omicsbox. 3rd March 2009.
- Parra, G., Bradnam, K. & Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, 23(9), 1061–1067. Available from: <https://doi.org/10.1093/bioinformatics/btm071>
- R CORE TEAM. (2020) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rawat, N., Neeraja, C.N., Nair, S. & 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*, 5, 1–15. Available from: <https://doi.org/10.1186/1939-8433-5-8>
- Reinink, K. & Dieleman, F.L. (1989) Comparison of sources of resistance to leaf aphids in lettuce (*Lactuca sativa* L.). *Euphytica*, 40, 21–29. Available from: <https://doi.org/10.1007/BF00023293>
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. et al. (2011) Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24–26. Available from: <https://doi.org/10.1038/nbt.1754>
- Schmittgen, T.D. & Livak, K.J. (2008) Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, 3(6), 1101–1108. Available from: <https://doi.org/10.1038/nprot.2008.73>
- Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V. et al. (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, 22(6), 839–851. Available from: <https://doi.org/10.1261/rna.053959.115>
- Shahid, S., Shi, Y., Yang, C., Li, J., Ali, M.Y., Smaghe, G. et al. (2021) CCHamide2-receptor regulates feeding behavior in the pea aphid, *Acyrtosiphon pisum*. *Peptides*, 143, 170596. Available from: <https://doi.org/10.1016/j.peptides.2021.170596>
- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. & Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212. Available from: <https://doi.org/10.1093/bioinformatics/btv351>
- Singh, K.S., Cordeiro, E.M., Troczka, B.J., Pym, A., Mackisack, J., Mathers, T.C. et al. (2021) Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest *Myzus persicae*. *Communications Biology*, 4(1), 847. Available from: <https://doi.org/10.5061/dryad.vhhmgqnt7>
- Smith, A., Hubley, R. & Green, P. (2013) RepeatMasker Open-4.0. *RepeatMasker Open-4.0*.
- Stuart, J. (2015) Insect effectors and gene-for-gene interactions with host plants. *Current Opinion in Insect Science*, 9, 56–61. Available from: <https://doi.org/10.1016/j.cois.2015.02.010>
- Tagu, D., Klingler, J.P., Moya, A. & Simon, J.C. (2008) Early progress in aphid genomics and consequences for plant-aphid interactions studies. *Molecular Plant-Microbe Interactions*, 21(6), 701–708. Available from: <https://doi.org/10.1094/MPMI-21-6-0701>
- Tatchell, G.M., Collier, R.H. & Hough, G.L. (2017) IPM case studies: leafy salad crops. In: van Emden, H.F. & Harrington, R. (Eds.) *Aphids as Crop*

- Pests, 2nd edition, pp. 569–577. Available from: <https://doi.org/10.1079/9781780647098.0569>
- ten Broeke, C.J., Dicke, M. & 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*, 148(2), 130–141. Wallingford UK: CABI. Available from: <https://doi.org/10.1111/eea.12084>
- Thabuis, A.P.P., Teekens, K.C., 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. [Accessed: 5th May 2021]. Available at: <https://patentimages.storage.googleapis.com/e9/72/17/1a444388508a4d/US8829280.pdf>
- 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: 13th September 2021].
- Tjallingii, W.F. & Esch, T.H. (1993) Fine structure of aphid stylet routes in plant tissues in correlation with EPG signals. *Physiological Entomology*, 18(3), 317–328. Available from: <https://doi.org/10.1111/j.1365-3032.1993.tb00604.x>
- Trocza, B.J., Singh, K.S., Zimmer, C.T., Vontas, J., Nauen, R., Hayward, A. et al. (2021) Molecular innovations underlying resistance to nicotine and neonicotinoids in the aphid *Myzus persicae*. *Pest Management Science*, 77(12), 5311–5320. Available from: <https://doi.org/10.1002/ps.6558>
- van der Arend, A.J. (2003) The possibility of *Nasonovia ribisnigri* resistance breaking biotype development due to plant host resistance: a literature study. In: van Hintum, T.J.L., Lebeda, A., Pink, D., Schut, J.W. *Eucarpia leafy vegetables conference*, The Netherlands, Centre for Genetic Resources. pp.75–81.
- Van Helden, M., Tjallingii, W.F. & Dieleman, F.L. (1993) The resistance of lettuce (*Lactuca sativa* L.) to *Nasonovia ribisnigri*: bionomics of *N. ribisnigri* on near isogenic lettuce lines. *Entomologia Experimentalis et Applicata*, 66(1), 53–58. Available from: <https://doi.org/10.1111/j.1570-7458.1993.tb00692.x>
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology*, 3(7), 1–12. Available from: <https://doi.org/10.1186/gb-2002-3-7-research0034>
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S. et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, 9(11), e112963. Available from: <https://doi.org/10.1371/journal.pone.0112963>
- Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G. et al. (2018) BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, 35(3), 543–548. Available from: <https://doi.org/10.1093/molbev/msx319>
- Wei, Z., Hu, W., Lin, Q., Cheng, X., Tong, M., Zhu, L. et al. (2009) Understanding rice plant resistance to the Brown planthopper (*Nilaparvata lugens*): a proteomic approach. *Proteomics*, 9(10), 2798–2808. Available from: <https://doi.org/10.1002/pmic.200800840>
- Wenger, J.A., Cassone, B.J., Legeai, F., Johnston, J.S., Bansal, R., Yates, A.D. et al. (2020) Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect Biochemistry and Molecular Biology*, 123, 102917. Available from: <https://doi.org/10.1016/j.ibmb.2017.01.005>
- Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V. et al. (2007) Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 35-(suppl_1), D5–D12. Available from: <https://doi.org/10.1093/nar/gkl1031>
- Wingett, S.W. & Andrews, S. (2018) FastQ screen: a tool for multi-genome mapping and quality control. *F1000Research*, 7, 1338. Available from: <https://doi.org/10.12688/f1000research.15931.2>
- Wu, H., Wang, C. & Wu, Z. (2015) PROPER: comprehensive power evaluation for differential expression using RNA-seq. *Bioinformatics*, 31(2), 233–241.
- Zhao, C., Escalante, L.N., Chen, H., Benatti, T.R., Qu, J., Chellapilla, S. et al. (2015) A massive expansion of effector genes underlies gall formation in the wheat pest *Mayetiola destructor*. *Current Biology*, 25(5), 613–620. Available from: <https://doi.org/10.1016/j.cub.2014.12.057>
- 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*, 35(12), 2084–2092. Available from: <https://doi.org/10.1093/bioinformatics/bty895>
- Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L. & Yorke, J.A. (2013) The MaSuRCA genome assembler. *Bioinformatics*, 29(21), 2669–2677. Available from: <https://doi.org/10.1093/bioinformatics/btt476>
- Zimmer, C.T., Garrood, W.T., Singh, K.S., Randall, E., Lueke, B., Gutbrod, O. et al. (2018) Neofunctionalization of duplicated P450 genes drives the evolution of insecticide resistance in the brown planthopper. *Current Biology*, 28(2), 268–274. Available from: <https://doi.org/10.1016/j.cub.2017.11.060>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Data S1. k-mer graphs for the *N. ribisnigri* genomes. 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. (A) Wild-type *N. ribisnigri* (Nr8_123) k-mer plots (Illumina data only). (B) Resistant *N. ribisnigri* k-mer plots (Illumina data only).

S2. (A) Species distribution for the wild-type genome (Nr_8) highlighting the number of shared genes with other organisms from BLAST using OmicsBox (v. 3.1.8). (B) Top-hit species distribution for the wild-type genome (Nr_8) highlighting the number of shared top-hit genes with other organisms from a BLAST search using OmicsBox (v.3.1.8).

S3. (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 (*A. 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.

S4. HISAT2 results of number of reads which aligned to the Nr_8 WT genome annotation using feature counts and mapping statistics.

S5. 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 resistance-breaking and susceptible biotype.

S6. Volcano plot of the 18,872 genes in both resistance-breaking (Nr:1) and susceptible (Nr:0) *N. ribisnigri* biotypes. Six hundred eighty-nine 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.

S7. Protein prediction outputs. Both DeepLoc 2.0 (top) and BUSCA (below) predict localisation in both the cytoplasm and nucleus.

S8. Primer efficiencies of *RPS9* and *RPS18*.

S9. Primer efficiencies of *PP1* and *PP2*.

S10. Conditions used for qRT-PCR experiment for *N. ribisnigri*. Nr:0 = susceptible biotype (unable to feed on resistant lettuce containing the Nr-gene).

S11. Accession numbers for all genomes are available on the NCBI (BioProject ID: PRJNA857679).

S12. Accession numbers for all sequence reads which are available on the NCBI (BioProject ID: PRJNA857679).

S13. Primer sequences and amplicon characteristics of resistance genes and house-keeping genes for qRT-PCR validation experiment for *N. ribisnigri*.

S14. qRT-PCR validation methodology.

S15. Housekeeping (HKGs) stability (GeNorm) (A) 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 HKGs are highlighted in green and red, respectively. (B) Average stability of HK1, HK2 and HK3. Least stable on the left (HK3) and most stable on the right (HK1 and HK2).

S16. Power analysis code.

How to cite this article: Garrett, D., Teakle, G., Collier, R., Bell, J.R., Cerezo-Medina, S. & Morales-Hojas, R. (2024) Genome assembly and transcriptomic analysis to elucidate the ability of *Nasonovia ribisnigri* to break host plant resistance. *Insect Molecular Biology*, 1–18. Available from: <https://doi.org/10.1111/imb.12894>