
OPTIMISING CRISPR-CAS SYSTEMS FOR TARGETED GENE EDITING IN APHIDS AND WHITEFLIES

Reuben Philip James

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia

John Innes Centre

December 2024

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright

Abstract

The green peach aphid (*Myzus persicae*) and the silverleaf whitefly (*Bemisia tabaci*) are global, polyphagous hemipteran pests. Defined by piercing-sucking mouthparts, these insects feed on phloem sap, causing plant damage and transmitting viruses. They also secrete effector proteins to suppress plant defences. Gene editing technologies could be used to knock out effector, and viral transmission associated genes, enabling a deeper understanding of these damaging processes, with the potential of identifying an ‘Achilles heel’ to target for insect control.

This thesis investigated Receptor Mediated Ovary Transduction of Cargo (ReMOT Control) for gene editing in *M. persicae*. This involved injecting females with a Cas9 protein fused to a vitellogenin (Vg)-derived peptide, enabling the Cas9-sgRNA complex to reach developing preblastodermal embryos for germline editing. This technique has been successful in insects and one arachnid, indicating its potential in aphids. Often, CRISPR in insects occurs via egg injection; because aphids only lay eggs under autumn-winter conditions, ReMOT offers a more feasible alternative to direct egg injection.

To optimize ReMOT Control in *M. persicae*, Vgs were characterized, and an embryo-targeting peptide (MpRV) was identified. Its targeting efficiency was tested alongside the *Drosophila melanogaster* P2C peptide, which has been used for embryo-targeting in other arthropods. Cas9 fusions with MpRV and P2C were successfully expressed, purified, and shown to cleave the aphid *white* gene in vitro, a gene predicted to alter eye colour. Upon optimizing injection protocols, evidence of low-efficiency somatic gene editing in aphids was obtained.

For *B. tabaci*, a smaller CRISPR enzyme, CasΦ, was tested due to its compact size and AT-rich targeting preference. A CasΦ chimera fused with the whitefly Vg-derived peptide BtKV was developed, and ReMOT experiments showed preliminary evidence of *white* gene editing.

In summary, ReMOT Control technologies represent a promising advancement for achieving more efficient gene editing in aphids and whiteflies.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

“SCIENCE!”

“SCIENCE!”

– Sam Mugford

Table of Contents

Table of Contents

Abstract.....	1
“SCIENCE!”	2
Table of Contents.....	3
Acknowledgements	7
Funding Acknowledgement	10
List of Figures.....	11
List of Tables.....	13
List of abbreviations	16
1 General Introduction	18
1.1 Hemiptera	18
1.2 Aphids	19
1.2.1 Aphids secrete effectors to modulate plant defences	19
1.2.2 Aphid life cycle	20
1.2.3 Aphid embryogenesis and sex determination	24
1.3 Whiteflies.....	27
1.4 Gene editing in insects	29
1.4.1 RNA interference	29
1.4.2 CRISPR/Cas systems	30
1.4.3 CRISPR/Cas systems in insects	34
1.4.4 CRISPR/Cas systems in aphids.....	35
1.5 ReMOT Control	37
1.6 Vitellogenin	41
1.7 Optimising tools to study gene function in aphids and whiteflies.....	46
1.8 Aims	47
1.9 Contributions to this thesis	47
2 Myzus persicae vitellogenin harbours a region that may be used for REMOT control in aphids.	48
2.1 Introduction	48
2.1.1 Contributions to this chapter.....	55
2.2 Results	56
2.2.1 Aphids have lost the ‘canonical’ insect vitellogenin, but have retained vitellogenin-like ...	56
2.2.2 InterProScan reveals aphid Vg contains canonical Vg structural domains Vg-N, DUF1943 and VWD	58
2.2.3 MpVg contains RXXR cleavage site motifs, present in canonical Vgs	59
2.2.4 AlphaFold predictions reveal similarities between the structural domains of Vgs, and reveal that MpVg is more structurally similar to BtVg-like than BtVg.....	60
2.2.5 ‘MprV’ is an <i>M. persicae</i> alternative to ‘BtKV’ which aligns in sequence and structure	64

2.2.6	Many vitellogenesis signalling related genes have homologues in <i>M. persicae</i>	66
2.2.7	Morph-specific RNAseq analysis reveals MpVg is most expressed in males	68
2.3	Discussion	73
2.3.1	Aphid Vgs are Vg-like orthologues, which likely carry canonical Vg functions.....	73
2.3.2	Aphidomorpha and Coccoidae lost canonical Vg when they diverged from psyllids	76
2.3.3	The peptide MpRV is a promising candidate peptide for ReMOT control in aphids	79
2.4	Materials and Methods.....	82
2.4.1	Genome sequences used in the study	82
2.4.2	Creation of aphid protein database compiling protein sequences from 18 aphid genomes.....	82
2.4.3	BLAST searches for Vg homologues	83
2.4.4	Construction of a Vg, Vg-like and aphid Vg phylogeny	84
2.4.5	Vg structural domain prediction with InterPro.....	84
2.4.6	Construction of an aphid Vg alignment	85
2.4.7	Identification of an RXXR cleavage site motif in MpVg.....	85
2.4.8	Vg structure predictions in alphafold2	85
2.4.9	Comparison of MpVg, with silver lamprey lipovitellin and honeybee Vg structures	85
2.4.10	Foldseek analysis of MpVg vs BtVg-like and BtVg	86
2.4.11	MSA to find <i>M. persicae</i> alternative to 'BtKV'	86
2.4.12	Identification of Vg-signalling associated gene homologues in <i>M. persicae</i>	87
2.4.13	Morph specific RNAseq analysis of MpVg, MpVgR, and Vg-signalling associated genes	87
3	Generation and validation of embryo-localising fluorescent protein chimeras	88
3.1	Introduction	88
3.2	Contributions to this chapter.....	90
3.3	Results	90
3.3.1	mCherry is the best fluorophore for localisation experiments in <i>M. persicae</i> embryos	90
3.3.2	MpRV-mCherry is expressed and purified from <i>E. coli</i> on a large scale	92
3.3.3	Optimisation of <i>M. persicae</i> injections	93
3.3.4	MpRV-mCherry and P2C-mCherry localise to early-stage embryos 24 hours post injection of adult female mothers	95
3.4	Discussion	99
3.5	Materials and methods.....	104
3.5.1	Preparation of <i>M. persicae</i> ovary samples for confocal microscopy for autofluorescence analysis.....	104
3.5.2	Confocal microscopy and fluorescence intensity analysis for autofluorescence analysis	104
3.5.3	Generation of an MpRV fragment for golden gate assembly	105
3.5.4	Generation of an mCherry fragment for golden gate assembly.....	106
3.5.5	Golden gate assembly	108
3.5.6	Heat-shock transformation of DH5 α <i>E. coli</i> with the MpRV-mCherry construct and screening for successful clones	110
3.5.7	Expression and purification of MpRV-mCherry.....	110
3.5.8	Experimental design for assessment of uptake of chimeric embryo localising mCherry proteins by early-stage <i>M. persicae</i> embryos	112
3.5.9	<i>M. persicae</i> age-matched colony preparation.....	113
3.5.10	<i>M. persicae</i> adult female injection procedure.....	113
3.5.11	Sample preparation for MpRV- and P2C-mCherry embryo localisation analysis	116
3.5.12	Confocal microscopy for MpRV- and P2C-mCherry embryo localisation analysis	116
3.5.13	Fluorescence intensity analysis across early-stage embryos.....	117
4	Generation and validation of embryo-targeting Cas9 fusion RNPs.....	119

4.1	Introduction	119
4.2	Results	121
4.2.1	Optimisation of protocols for producing the MpRV-Cas9 chimera in <i>E. coli</i>	121
4.2.2	Producing the P2C-Cas9 chimera in <i>E. coli</i>	127
4.2.3	Design of sgRNAs for the <i>M. persicae white</i> gene	129
4.2.4	<i>In vitro</i> assays to assess if MpRV-Cas9-His and P2C-Cas9 chimeras cut <i>M. persicae white</i> gene fragments in the presence of sgRNAs	132
4.3	Discussion	135
4.4	Materials and Methods	139
4.4.1	Generation of a construct for expression of MpRV-Cas9	139
4.4.2	Small scale expression and purification of MpRV-Cas9	143
4.4.3	Large scale expression and purification of P2C-Cas9	145
4.4.4	Large scale expression and purification of MpRV-Cas9	146
4.4.5	Identification of an <i>M. persicae white</i> homologue as a target gene	146
4.4.6	Identification and generation of sgRNAs targeting <i>white</i>	147
4.4.7	Genomic DNA extraction from <i>M. persicae</i>	148
4.4.8	PCR of <i>white</i> target fragments	148
4.4.9	<i>In vitro</i> cutting assays	150
5	Gene editing of <i>Myzus persicae</i> by ReMOT control	151
5.1	Introduction	151
5.1.1	Contributions to this chapter	155
5.2	Results	155
5.2.1	Selection and optimization of aphid injection method	155
5.2.2	Optimization of rearing conditions to enhance likelihood of capturing edited individuals	156
5.2.3	Selection of genotyping procedure	157
5.2.4	Offspring of <i>M. persicae</i> females injected with MpRV-Cas9-sgRNA RNPs are unlikely to have edited target sites	157
5.2.5	Evidence of target site editing in offspring of <i>M. persicae</i> females injected with P2C-Cas9-sgRNA RNPs	160
5.2.6	Approach to detect large deletions in the region targeted by sgRNA1, 2 and 3	165
5.2.7	There is a natural SNP in the sgRNA3 target site	167
5.3	Discussion	168
5.4	Materials and Methods	176
5.4.1	Establishment of age matched <i>M. persicae</i> colonies	176
5.4.2	Micropipette preparation	176
5.4.3	Aphid injection procedure	176
5.4.4	Formation of MpRV- and P2C-Cas9-sgRNA RNPs	177
5.4.5	ReMOT control injection mixture preparation	178
5.4.6	ReMOT control experimental procedure	178
5.4.7	DNA extraction from pooled offspring of <i>M. persicae</i> injected with MpRV- or P2C-Cas9	180
5.4.8	Genotyping of offspring from injected <i>M. persicae</i>	181
5.4.9	Analysis of SNPs in sgRNA target sites	183
6	Investigating Casϕ as an alternative to Cas9 in ReMOT control in <i>Bemisia tabaci</i>	184
6.1	Introduction	184
6.1.1	Contributions to this chapter	186
6.2	Results	186
6.2.1	Construction of pET28a-BtKV-Cas ϕ -Cys	186

6.2.2	BtKV-Cas ϕ heterologous production in <i>E. coli</i>	187
6.2.3	Design and synthesis of Cas ϕ sgRNAs targeting <i>B. tabaci white</i>	188
6.2.4	BtKV-Cas ϕ cleaves <i>B. tabaci white</i> amplicons when complexed with sgRNAs	189
6.2.5	Low survival rate was observed in <i>B. tabaci</i> that had been injected with BtKV-Cas ϕ -sgRNA RNP	192
6.2.6	Evidence of editing of <i>white</i> was observed.....	193
6.3	Discussion	195
6.4	Materials and methods.....	198
6.4.1	Construction of pET28a-BtKV-Cas9-Cys	198
6.4.2	PCR to generate a Cas ϕ fragment for Gibson cloning	198
6.4.3	Gibson cloning to construct pET28a-BtKV-Cas ϕ -Cys	200
6.4.4	Small scale production of BtKV-Cas ϕ	201
6.4.5	Large scale production of BtKV-Cas ϕ	203
6.4.6	sgRNA design against <i>B. tabaci white</i>	204
6.4.7	sgRNA synthesis.....	204
6.4.8	<i>B. tabaci</i> genomic DNA extraction	206
6.4.9	PCR for fragments of <i>B. tabaci white</i>	207
6.4.10	In vitro cutting assays	208
6.4.11	Micropipette preparation	209
6.4.12	ReMOT Control experimental procedure	209
7	Discussion.....	211
7.1	Summary and discussion of research findings	211
7.2	Future directions	214
7.3	Conclusions	221
Appendix		222
	Full BUSCO scores of 19 Aphid genomes.....	223
	Sequences used in the vitellogenin and vitellogenin-like phylogeny.....	224
	General code for command line BLAST.....	230
	BLAST for Vg sequences in compiled proteomes of aphid genomes	231
	Alphafold script	233
	R script for generation of pLDDT plots	235
	Scripts for morph specific RNASeq analysis in R	238
	Tukey statistics for morph specific RNASeq analysis	244
	Raw data for fluorescence intensity across early-stage embryos.....	248
	Script to generate fluorescence intensity Box plot in R.....	250
	MpRV-Cas9 ReMOT control experiment tables.....	251
	P2C-Cas9 ReMOT control experiment tables	256
	Summary table of ICE analysis for MpRV-Cas9 experiments	263
	Summary table of ICE analysis for P2C-Cas9 experiments.....	270
References		277

Acknowledgements

Firstly, I would like to acknowledge my outstanding supervisory team. I have been blessed to have an extensive supervisory team with a range of expertise. Professor Saskia Hogenhout, thank you for taking a chance on me for this project, and for guiding me through my PhD superbly, taking care to help me become a forward thinking, confident scientist. The empathy shown when it comes to my health problems has been heavily appreciated. Professors Wendy Harwood and George Lomonossoff, thank you for your input and guidance as my secondary supervisors. Dr. Grant Hughes, thank you for all your input and for hosting me in your lab in Liverpool, which provided me valuable experience for experimental design. Thanks also for the ready supply of reagents. Dr. Marcus Guest, thank you for valuable input and for hosting me at Syngenta. I found the experience both valuable and very enjoyable. Finally, Dr. Sam Mugford, thank you for everything, whether it be technical help, wisdom from your seemingly endless repository of knowledge, advice on survival, or the many pints. Your unwavering positivity around science has kept me motivated throughout.

My colleagues in the Hogenhout lab (current and alumni) have provided invaluable support throughout my PhD. I thank everyone for their kindness and readiness to include me as a lab member. I have always felt part of the family here. Matteo, Weijie and Qun, thank you for invaluable support and advice in protein purification. My fellow PhD students, past and present – Roland, James, Josh and George – thank you for all of your support. It was a privilege to take this journey with your company. Nicholas, thank you for being a great undergrad project student. It was a privilege to be your supervisor. Thomas Mathers, thank you help in bioinformatic analysis, for your positivity, and for helping me plan my career every time we meet, even since you left the lab. Every time I see you, I have a fresh burst of motivation. Fede, thank you for insightful discussion, and for being a great friend. Mar, I'm not sure I can thank you enough. I don't know how I would have survived the many aphid injections and DNA extractions without your help. Your input on bioinformatic analysis has been invaluable and the feedback you have

provided me for this thesis has been greatly appreciated. I also thank you for your enthusiasm for lab morale, with birthday cards and lab photos.

I must give a specific thank you to my year in industry student and current lab pre-doc: Amber Hall. Amber, I could not have wished for a better student. You made my first supervisory experience feel easy, and you brought my little extra project to life, for which I will forever be grateful. Furthermore, you are a delight to be around, and you make the lab a happier place. You are a great friend, a great person, and if you want to, you will go on to be a truly fantastic scientist.

Outside of the lab, I have had the help of many others at the Norwich Research Park. Thank you Raf, Caroline and Adam for help with protein purification. Thank you Ioanna for your help in genotyping experimental design and for being a great friend. Thank you, Victor, Anna, Jess, Liz, Mark and Will, from the insectary. None of my aphid or whitefly experiments would have been possible without your help. Thank you, Eva and Sergio, for training me in confocal microscopy. Thank you, Ian and Mukund, for showing me your experiments in Liverpool, and for making me feel welcome. Thank you, Lucy, Carlo and Patricia, for welcoming me at Syngenta. It was a pleasure to work with you. James Milson, thanks for being a good mate, the Rec feels slightly emptier without you. Thomas Navarro, thank you for being a supportive and great friend. Your advice and support have been incredible. I loved every minute of working on the JSV with you and have enjoyed our friendship ever since. Long may it continue.

During my PhD, I was diagnosed with Type 1 diabetes. I must thank the staff of the Elsie Bartrum Diabetes centre for excellent treatment and care. I must also thank Brigid from the Elsie Bartrum Youth Carers team. Type 1 diabetes is a stressful disease and could have made completing my PhD near impossible. The care I have received has quickly helped me understand how to manage my condition with little disruption to my PhD.

Next, I must acknowledge my wonderful band members – past and present. The band has given me great stress relief during my PhD and performing with you all has been a pleasure. Gary, thank you for encouraging me to start the band, for being my first

drummer, and for all of the advice you have given me, whether it be PhD or music related. Charlotte, never change, you're fantastic. Victor and Ryan, you're both fantastic bassists, and it has been a pleasure to perform with you. Jorge, thank you for being a wonderful pianist, and for learning pretty much every song I threw at you. Jay, (another Hogenhout lab member), thank you for being a fantastic drummer, a great friend, and a great mentor. I have learnt so much from you and your journey to faculty. I'll miss you greatly when you're in Arizona. Jogi, thanks for being the best. Thanks for bringing my songs to life with your beautiful mind for lead guitar composition. Thanks for always picking up the phone. Thanks for supporting me through this process. Many great times have been had, and many yet lie ahead.

Outside of academia, I have had great friends who have supported me through this process. George, Ryan, Declan, Reece, Rory and Harvey, thanks for being great mates, whether it be from afar on the PS5, or in person when we get the chance. Rhys, thank you for always supporting me, for looking out for me, and for being an amazing friend.

To my partner Michelle, thank you for supporting me through this writing process. Putting up with a grumpy write-up bound PhD student is not easy, and I appreciate all you do to hold me up and keep me going.

Most of all I must thank my family. Mum and Dad, your support and guidance always keep me going. A better pair of parents could never be asked for, and I hope I grow up to be like you. Ellie, your strength is inspiring, motivating and seemingly unending. Thanks for being a great sister and a great friend. Granny and Grandpa, Naini and Taidi thank you for all your support over the years. Thank you all for believing in me.

Finally, I must thank my best friend Fin. Fin, growing up with you was the great joy of my life. It was always said that we spurred each other on with our competitiveness, and that as a result, we both had great grades. This was true, but more than that, your motivation and constant effort to do well, whether in life or in school, always inspired me. Over 7 years have passed since we lost you, and that inspiration you gave me remains every day, and helped push me to write this thesis. I know you are still cheering me on.

Funding Acknowledgement

This work was supported by the Norwich Research Park Biosciences Doctoral Training Partnership funded by UK Research and Innovation (UKRI) Biotechnology and Biological Sciences Research Council (BBSRC) as an iCASE studentship with Syngenta Ltd. Additional Support is provided by the BBSRC Institute Strategy Programmes (BBS/E/J/000PR9797 and BBS/E/J/230001B) awarded to the JIC. The JIC is grant-aided by the John Innes Foundation. We also acknowledge support from Liverpool School of Tropical Medicine.

List of Figures

Figure 1.1: A typical lifecycle of a host-alternating aphid.....	23
Figure 1.2: A diagram to show the stages of embryogenesis in the parthogenic aphid ovary.....	25
Figure 1.3: Images depicting an oviparous aphid ovary.....	26
Figure 1.4: The lifecycle of <i>Bemisia tabaci</i>	28
Figure 1.5: A schematic summary of the Class 1 and 2 CRISPR systems.	32
Figure 1.6: A comparison of ReMOT control and embryo injection methods of gene editing in mosquitos.....	40
Figure 1.7: Hormonal and nutritional control of Vg synthesis in the fat body.	45
Figure 2.1: A maximum likelihood phylogeny of hemipteran Vg and Vg-like proteins, including aphid Vgs..	58
Figure 2.2: A region of DUF1943 in aphid Vg sequences with contains low similarity but does not explain a lack of DUF1943 prediction by InterProScan in 5 species.	59
Figure 2.3: Predictions of the protein structures of MpVg, BtVg-like 1 and BtVg using Alphafold2.....	62
Figure 2.4: The predicted structure of MpVg, the solved crystal structure of <i>I. unicuspis</i> lipovitellin (Iu lipovitellin), and the modelled structure of <i>A. mellifera</i> Vg (AmVg) (136, 138)..	63
Figure 2.5: Structural alignments of MpVg against BtVg-like 1 and BtVg.....	63
Figure 2.6: A MSA showing how aphid Vg protein sequences align with BtKV, and the VgR binding peptide derived from MrVg (102).	64
Figure 2.7: The VgR binding peptides of BtVg and MpVg (named BtKV and MpRV respectively) reside in the same structural position.	65
Figure 2.8: Expression patterns of MpVg, MpVgR and JH-pathway associated genes (Met, SRC, FoxO [isoforms X1 and X2], LCMT1, PP2A, Cdc6, Cdc2, MCM3, 4 and 7, and Orc5).....	69
Figure 2.9: A heatmap with clustering to show expression levels of each JH pathway associated gene across all individuals of all 4 morphs.	70
Figure 2.10: Expression patterns of MpVg, MpVgR and ECR-pathway associated genes (EcR, Ultraspiracle, BC core protein-like, E75, E78C, and HR3).	71
Figure 2.11: A heatmap with clustering to show expression levels of each 20E pathway associated gene across all individuals of all 4 morphs.	72
Figure 3.1: An <i>M. persicae</i> embryo imaged under four excitation wavelengths, corresponding to the fluorophores CFP, GFP, RFP, and mPlum.	91
Figure 3.2: A plasmid map of pOPIN-F5-6his-MpRV-mCherry-6his-K, constructed using golden gate assembly.....	92
Figure 3.3: SDS-PAGE gel showing the sonicated sample, soluble fraction, and pure protein samples of MpRV-mCherry obtained by large scale purification.....	93
Figure 3.4: Example confocal microscopy images of early-stage embryos in ovaries originating from injected mothers.....	97
Figure 3.5: A Box plot showing mean fluorescence intensity of ‘early-stage’ embryos of ovaries from <i>M. persicae</i> adult females injected with P2C-mCherry (orange), MpRV-mCherry (blue), untagged mCherry (red) or IMAC buffer A4 (yellow).	98

Figure 4.1: SDS-Page gels showing different stages of optimisation of purification of MpRV-Cas9.	124
Figure 4.2: A plasmid map for the final version of the MpRV-Cas9 expression plasmid.	125
Figure 4.3: SDS-PAGE gels showing two IMAC stages (pre and post 3C cleavage) of MpRV-Cas9 purification.	126
Figure 4.4: A plasmid map of pET28a-P2C-g4sx3-Cas9, created and gifted by Ian Bennet and Grant Hughes at Liverpool School of Tropical Medicine.	128
Figure 4.5: SDS-PAGE gels showing elutions from the first and second IMAC and gel filtration for purification of P2C-Cas9.	128
Figure 4.6: A recreation of the alignment of B. tabaci Protein White against the top hit when subject to BLASTp against the M. persicae Clone O v2.1 (154) proteome, made in CLC main workbench (QIAGEN).	130
Figure 4.7: A map of the M. persicae white gene, coding sequence (CDS) and protein.	131
Figure 4.8: In vitro nuclease assays to assess P2C-Cas9 nuclease action with each of the white targeting sgRNAs.	133
Figure 4.9: A single electrophoresis gel showing the results of an in vitro nuclease assay testing MpRV-Cas9 RNPs.	134
Figure 5.1: An example of ICE analysis results for offspring of mother injected with MpRV-Cas9 RNPs.	159
Figure 5.2: Two M. persicae offspring displayed phenotypic evidence of successful genome editing.	162
Figure 5.3: ICE results for the reverse sequencing reaction of sgRNA1 target region of offspring of mothers injected with P2C-Cas9 RNPs.	163
Figure 5.4: Amplicon sequencing results from a negative (un-injected) control sample and two samples injected with P2C-Cas9.	164
Figure 5.5: Electrophoresis gels showing detection of products when investigating large deletions caused by genomic editing by ReMOT Control.	166
Figure 6.1: A plasmid map of pET28a-BtKV-Cas ϕ -Cys, created by swapping the Cas9-NLS-HA portion of pET28a-BtKV-Cas9-Cys with the Cas ϕ -2xSV40 NLS-2xFLAG portion of pPP441 by Gibson assembly.	187
Figure 6.2: SDS-PAGE shows a pure sample of BtKV-Cas ϕ was produced.	188
Figure 6.3: In vitro nuclease assays using BtKV-Cas ϕ with sgRNAs synthesised by Merck.	191
Figure 6.4: In vitro cutting assays with Cas ϕ sgRNAs synthesised by T7 transcription.	192
Figure 6.5: Two 4th instar nymphs showed phenotypic evidence of successful gene editing by BtKV-Cas ϕ -sgRNA RNPs.	193
Figure 6.6: An alignment with Sanger sequencing trace data of the sgRNA 3 target region of exon 3.	194
Figure 7.1: A schematic of ReMOT control in M. persicae.	214
Figure 7.2: A model of the hypothesised signalling pathways regulating vitellogenesis in aphids through day-length sensing and its impact on ILP expression.	217

List of Tables

Table 2.1: A table of homologues of Vg-signalling associated genes, the queries used to find them, the E values and Total scores. Finally, the <i>M. persicae</i> 'O' v2.0 gene ID and associated scaffold is shown, except for Kr-h1 and TOR which were only searched for by BLASTp (NCBI) and not subject to further analysis.....	67
Table 3.1: Mean fluorescence intensity across the <i>M. persicae</i> embryo under different excitation wavelengths.	91
Table 3.2: Forward and reverse oligonucleotides ordered for generation of a golden-gate suitable MpRV fragment.	106
Table 3.3: Primers used to amplify an mCherry fragment for use in a golden gate cloning reaction.	108
Table 4.1: Primers used for SDM of pET28a-MpRV-Cas9-Cys to remove a C-terminal stop codon. The codon change is lowercase and written into the forward primer.	140
Table 4.2: Primers used to generate an insert of Gb1-3C for Gibson assembly into the EcoRI cut site of pET28a-MpRV-Cas9-His. The parts of the primers overlapping with the vector sequences flanking the EcoRI cut site are lowercase. An additional base to maintain framing is underlined in the reverse primer.	142
Table 4.3: Primers used in colony PCR to analyse the product of Gibson assembly to insert Gb1-3C into pET28a-MpRV-Cas9-His.	143
Table 4.4: sgRNAs designed against <i>M. persicae</i> white coding sequence.	147
Table 4.5: Primers used to generate fragments containing sgRNA1, 2 and 3 targets for use in in vitro cutting assays.	149
Table 5.1: The number of aphids injected with P2C- or MpRV-Cas9 in each experiment, with the date of experiment and whether saponin was included in the injection mixture.	180
Table 5.2: Primers used for genotyping offspring of adult female <i>M. persicae</i> injected with MpRV- or P2C-Cas9. Primers were designed to amplify regions of white genomic DNA containing each sgRNA target site.	183
Table 6.1: Cas ϕ sgRNA target sequences, the white gene exons they target, and their 5' PAM sequences.	189
Table 6.2: Primers used to amplify a fragment suitable for Gibson assembly containing the Cas ϕ coding sequence from pPP441. 'Overlap' sections required for Gibson assembly are in lowercase. Extra base pairs, added to maintain framing, are in lowercase and underlined. Annealing sections of the primers are in uppercase.	200
Table 6.3: Oligos ordered for PCR DNA template for generation of sgRNA T7 transcription templates.	206
Table 6.4: Primers used to generate sgRNA T7 transcription templates by PCR.	206
Table 6.5: Primers used to amplify exon 2, 3 and 5 amplicons.	208
Table S1: BUSCO scores of the Aphid genomes used.....	223
Table S2: Vitellogenin and Vitellogenin-like sequences sourced from NCBI for use in a phylogeny (Figure 3.1).	224
Table S2: Aphid vitellogenin sequences used in a phylogeny (Figure 3.1) and the genome annotations to which they belong. All were downloaded internally on the Norwich	

Bioscience Institutes (NBI) High Performance Computing (HPC) cluster. Most are published in Mathers et al., 2022 (151). <i>M. persicae</i> genome clone O v2.1 is published in Liu et al., 2024 (154). <i>A. pisum</i> v1.0 is published in Mathers et al., 2021 (140). <i>M. cerasi</i> genome assembly is published in Thorpe et al., 2019 (152) and the annotation was published in Mathers et al., 2022 (151).	228
Table S 3: Tukey results for JH Pathway associated genes.	244
Table S 4: Tukey results for 20E pathway associated genes.	246
Table S4: Raw fluorescence intensity data from Image J analysis of early stage embryos from MpRV-, P2C- and untagged mCherry, and buffer A4 injected aphids.	248
Table S6: Injected adult counts for the ReMOT control injections with MpRV-Cas9 performed on 14/02/2024.	251
Table S7: Progeny counts and eye colour inspection for the ReMOT control injections with MpRV-Cas9 performed on 14/02/2024. 'Eyes' refers to number of aphids with an eye colour change.	251
Table S8: Injected adult and progeny counts, and eye colour inspection for ReMOT control injections with MpRV-Cas9 on 21/02/2024. Only data for day 10 was recorded in this case.	252
Table S9: Injected adult counts for the ReMOT control injections with MpRV-Cas9 performed on 12/04/2024.	253
Table S10: Progeny counts and eye colour inspection for the ReMOT control injections with MpRV-Cas9 performed on 12/04/2024. 'Eyes' refers to number of aphids with an eye colour change.	253
Table S11: Injected adult counts for the ReMOT control injections with MpRV-Cas9 performed on 01/05/2024. In this experiment, lack of survival was likely due to improper watering of the plant. **Any escapees were collected here.	254
Table S12: Progeny counts and eye colour inspection for the ReMOT control injections with MpRV-Cas9 performed on 01/05/2024. 'Eyes' refers to number of aphids with an eye colour change. **Any escapees were collected here.	254
Table S13: Injected adult counts for the ReMOT control injections with MpRV-Cas9 performed on 13/06/2024.	255
Table S14: Progeny counts and eye colour inspection for the ReMOT control injections with MpRV-Cas9 performed on 13/06/2024. 'Eyes' refers to number of aphids with an eye colour change.	255
Table S15: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 15/12/2023.	256
Table S16: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 15/12/2023. 'Eyes' refers to number of aphids with an eye colour change.	256
Table S17: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 15/02/2024.	257
Table S18: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 15/02/2024. 'Eyes' refers to number of aphids with an eye colour change.	257
Table S19: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 23/02/2024.	258

Table S20: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 23/02/2024. 'Eyes' refers to number of aphids with an eye colour change.....	258
Table S21: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 08/03/2024.	259
Table S22: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 08/03/2024. 'Eyes' refers to number of aphids with an eye colour change.....	259
Table S23: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 15/03/2024. On Day 3 adults were removed from the original clip cage to a separate clip cage demarked by the 'a' suffix.	260
Table S24: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 15/03/2024. 'Eyes' refers to number of aphids with an eye colour change. On Day 3 adults were removed from the original clip cage to a separate clip cage demarked by the 'a' suffix. Remaining progeny from both clip cages were counted and screened.	261
Table S25: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 03/05/2023.	262
Table S26: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 03/05/2024. 'Eyes' refers to number of aphids with an eye colour change.....	262
Table S27: Summary table of all ICE analysis on amplicons from pools of progeny from MpRV-Cas9 injected aphids.....	263
Table S28: Summary table of all ICE analysis on amplicons from pools of progeny from P2C-Cas9 injected aphids.	270

List of abbreviations

ReMOT control: Receptor Mediated Ovary Transduction of Cargo

CRISPR: Clustered Regularly Interspaced Palindromic Repeats

RNPs: Ribonucleoproteins

sgRNA: single guide RNA

Vg: Vitellogenin

VgR: Vitellogenin receptor

Vn: vitellin

MpVg: *Myzus persicae* vitellogenin

MpVg-N: *Myzus persicae* vitellogenin lipoprotein N domain

MpVgR: *Myzus persicae* vitellogenin receptor

MpRV: *Myzus persicae* VgR binding peptide

BtKV: *Bemisia tabaci* VgR binding peptide

BtVg: *Bemisia tabaci* vitellogenin

BtVgR: *Bemisia tabaci* vitellogenin receptor

PBS: phosphate buffer saline

CFP: Cyan fluorescent protein

GFP: Green fluorescent protein

LDLR: low-density lipoprotein receptor

ILP(s): Insulin-like peptide(s)

InR: Insulin receptor

NSCI: Neurosecretory cells group I

pi: *pars intercerebralis*

PRR: pattern recognition receptor

PAMP: pathogen associated molecular patterns

LPS: lipopolysaccharide

PGN: peptidoglycan

LTA: Lipoteichoic acid

BmOTP: *Bombyx mori* Ovary Targeting Peptide

TRIS: tris(hydroxymethyl)aminomethane

EDTA: Ethylenediaminetetraacetic acid

List of abbreviations

TAE: TRIS acetate EDTA

PCR: polymerase chain reaction

DMSO: dimethyl sulfoxide

LB: Luria broth

BSA: Bovine serum albumin

SOC: Super Optimal broth with Catabolite repression

IMAC: Immobilised-metal affinity chromatography

DAPI: 4',6-diamidino-2-phenylindole

ROI: Region of interest

IPTG: Isopropyl β -D-1-thiogalactopyranoside

CPPs: Cell-penetrating peptides

DIPA-CRISPR: Direct Parental CRISPR

HDR: Homology directed repair

AAVs: Adeno-associated viruses

VLP: Virus-like particle

1 General Introduction

1.1 Hemiptera

The insect order Hemiptera diverged from other hemimetabolous insects 320 to 370 million years ago and is an order of insects characterised by the presence of stylets (specialised mouthparts) for piercing and sucking, and a resulting feeding behaviour (1-4). Further, all hemipteran insects have holocentric chromosomes (5). Within Hemiptera are two suborders: Sternorrhyncha, which includes aphids and whiteflies; and Auchenorrhyncha that includes planthoppers (Fulgoromorpha), leafhoppers and spittlebugs (Cicadomorpha), and true bugs (Heteroptera) such as *Lygus* species (1). Most hemipteran insects rely on phloem sap as their dominant food source (4). This is true for most hemipteran insects, with notable heteropteran exceptions including kissing bugs (*Rhodnius spp.*) and bed bugs (*Cimex lectularis*), which are exclusively hematophagous (6, 7), some xylem feeders including spittlebugs and sharpshooters (8), and mesophyll feeders such as the green mirid bug (*Apolygus lucorum*) (9). Hemiptera contains some of the most important agricultural pest insects, responsible for massive economical losses mostly via the plant viruses they transmit; whiteflies and aphids are both major agricultural pest insects which feed on plant nutrients from the phloem sap, which are essential for plant growth and development (10-12).

1.2 Aphids

Aphids are hemipteran insects belonging to the family Aphidoidea. There are more than 4700 species of aphids distributed globally, around 450 of which are among the most important crop pests (2, 13). Most aphids are specialised to a limited host range. Examples are the cabbage aphid (*Brevicoryne brassicae*) and the pea aphid (*Acyrtosiphon pisum*) (13). However, some aphids, such as the green peach aphid (*Myzus persicae*) are generalists and can establish colonies on a wide range of hosts around the world (13, 14). To supplement their nutrient poor diet, all aphids carry the obligate symbiont *Buchnera aphidicola*, which is transmitted from mother to daughter in bacteriocytes during reproduction (15, 16). Aphids are also a significant vector of plant viruses including 110 virus species of the genus *Potyvirus* (2, 10).

1.2.1 Aphids secrete effectors to modulate plant defences

Aphids can establish massive colonies extremely quickly. One reason for this is that they secrete effector proteins which modulate plant defences (1, 17). Upon landing on a plant host, an aphid probes the leaf, which involves stylet penetration of the epidermis cell wall, injection of saliva containing effector proteins, and ingestion of the cell components. If the plant is compatible, the stylets continue to the phloem to establish a long-term feeding site (1, 17). During this process, oral secretions are injected which contain effectors that modulate plant defences (1, 17). Examples from *M. persicae* include Mp10, which binds to AMSH deubiquitinase enzymes in plant cells to suppress immunity (18, 19), and cathepsin B (CathB) proteins, which localise to processing (p)-

bodies in plant cells and then recruit key immune regulators suppressing plant defence (20, 21). By these means, effector mediated plant defence modulation allows aphid populations to thrive on their plant hosts. Research into the molecular mechanisms of aphid effector proteins to understand their mechanisms is limited as there is not currently any simple and reliable method of gene knock-out in aphids.

1.2.2 Aphid life cycle

Most aphids switch between sexual and parthenogenic reproduction depending on seasonal photoperiodic changes (22). One holocycle includes many rounds of parthenogenic reproduction followed by one round of sexual reproduction. The latter involves the laying of eggs, from which new parthenogenic females hatch (14, 22). There is much variation in aphid life cycles among different species (22). However, a general life cycle can be described (14, 22-25). In the summer, viviparous (i.e., giving birth to live young) female aphids reproduce parthenogenetically, and give birth to viviparous females, which will continue to reproduce until the autumn. Parthenogenic aphids that give birth to parthenogenic aphids are called virginoparae. In the autumn, photoperiodic changes (daylight shortening) cause the viviparous females to produce sexual female and male aphid morphs, which can be winged or wingless. The parthenogenic females which produced sexual morphs are called 'sexuparae'. The sexual morphs mate, and the females lay cold-resistant overwintering eggs. In the early spring, viviparous parthenogenic females (called a foundress or fundatrix) hatch from the eggs. This represents one holocycle (Figure 1.1) (14, 22-25). It should be noted that the conditions required for overwintering eggs to hatch successfully are unclear and may vary

depending on geographical location; in a laboratory environment, eggs of the pea aphid have successfully hatched after being kept at 4°C for 85 days (26).

Many aphids, such as *A. pisum* colonise a single host for their entire lifecycle. However, about 10% of aphid species, including *M. persicae*, have host-alternating life cycles (Heteroecy) which involve mating and egg laying on one plant (primary host) during the autumn, winter and early spring, and parthenogenic reproduction on another plant (secondary host) during the late spring and summer (22). One notable difference between life cycles of non-host alternating and heteroecious species is that heteroecious sexuparae give birth to winged migrant morphs which fly to the primary host, at which point the female migrant morphs give birth to sexual wingless females, which mate with the winged or wingless males (Figure 1.1) (27). The primary host is often a woody plant, which are more nutritious in the spring; the secondary host is often a herbaceous plant, providing more nutrition in summer (22). Life cycles that involve year-round parthenogenesis (anholocycly) are derived from the holocycle, and can occur for multiple reasons, such as primary host availability or climate (14, 21, 23, 25). For example, holocyclic and anholocyclic *M. persicae* lineages exist depending on their global location and the resulting effect on climate and availability of their primary host plants, which are peach and (possibly) related *Prunus* species (14, 21, 23, 25). In the UK, for example, the peach is rare, and therefore *M. persicae* are thought to overwinter predominantly parthenogenetically in protected sites, such as greenhouses, and outside on crops and weeds in milder winters. It is unclear how holocyclic lineages survive, but it is possible that *M. persicae* can sexually reproduce on multiple *Prunus* species. It is

also unclear how overwintering occurs in particularly cold (sub-zero temperature) conditions (14, 21, 23, 25).

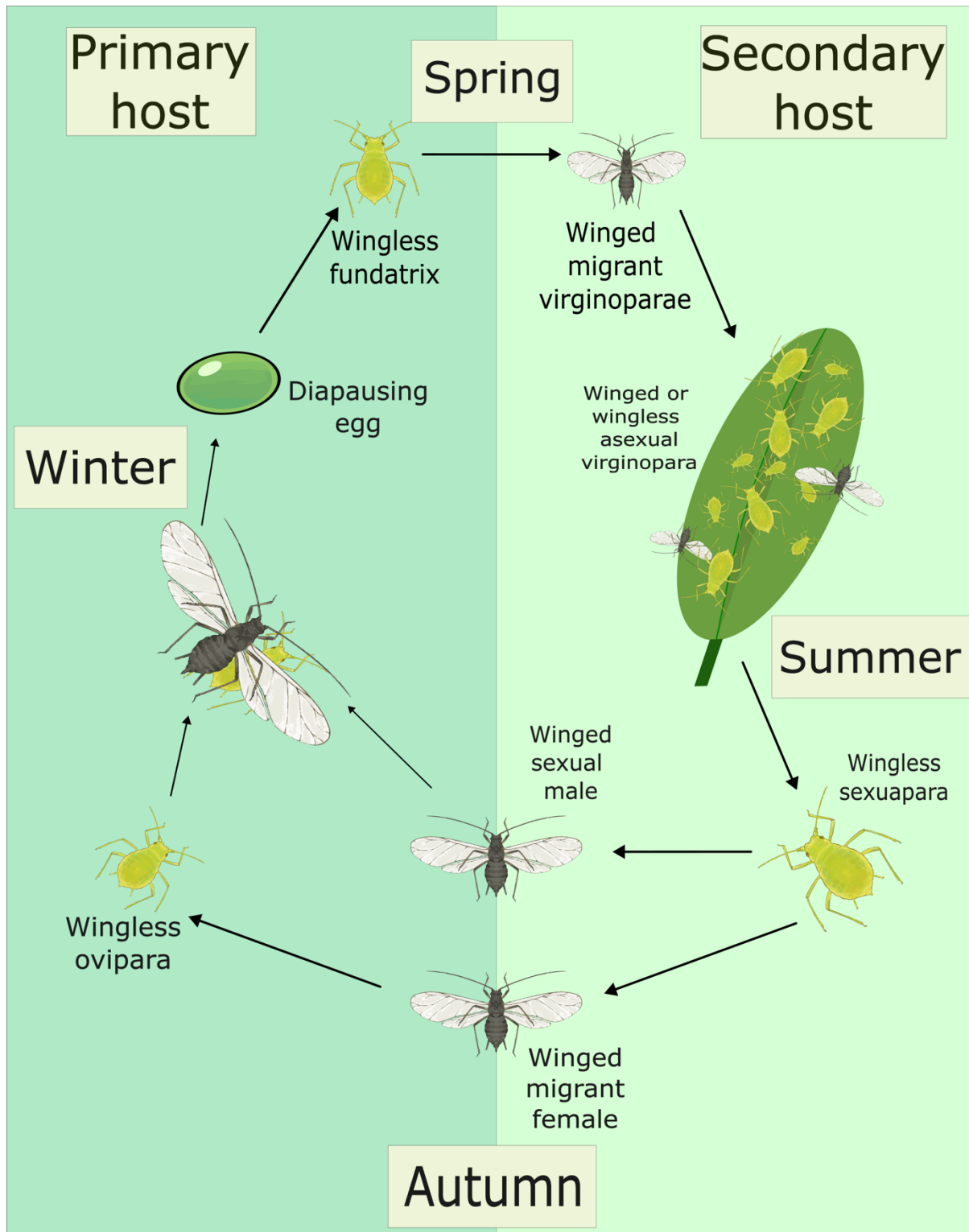


Figure 1.1: A typical lifecycle of a host-alternating aphid. In host-alternating species, upon detection of longer nights in autumn, parthenogenic females switch to produce winged migrant morphs which migrate to the primary host plant. The female migrants give birth to wingless sexual females, which mate with the winged males. The females then lay eggs which undergo an obligate diapause before the fundatrix can hatch in spring. The fundatrix females produce winged migrant virginoparae, which migrate to the secondary host, where parthenogenic reproduction of winged or wingless parthenogenic viviparous females occurs over summer. This figure was produced by the author with information from Hardy et al., 2017, and Yan et al., 2020 (22, 27).

1.2.3 Aphid embryogenesis and sex determination

Viviparous parthenogenic aphids, like all insects, have ovaries split into multiple ovarioles. At the anterior tip of each ovariole lie the nurse cells, posterior to which are several germ cells which are held until they mature into oocytes. The nurse and germ cells (32 cells in total) are in fact derived from a single embryonic germ cell which has undergone 5 rounds of division. After this, up to 6 embryos are found with increasing volume and maturity. Embryos mature over 20 stages: stage 0-4 are preblastodermal, then in stage 4-6 the blastoderm forms. Then, at stage 7 the *Buchnera aphidicola*-containing bacteriocyte is taken up. The embryos continue to grow and develop over the remaining stages, with stage 20 containing fully developed ovaries of its own with young embryos (28-31) (Figure 1.2). This phenomenon – of aphids being born pregnant with embryos that contain younger embryos – is called the telescoping of generations, and is one of the reasons that aphids are able to reproduce so quickly (28). The ovaries of oviparous females develop similarly with some marked differences. The oocyte grows much larger in sexual aphid embryos. Further, after choriogenesis and vitellogenesis, a follicular epithelium remains intact. All ovarioles are capable of producing eggs, which contain single nuclei (29) (Figure 1.3). Aphids are diploid for autosomes and have an XX/X0 sex determination system: female aphids have 2 X chromosomes, while male aphids only carry one X. Unlike some other hemipteran insects, models predict that the X chromosome in aphids carries male-beneficial traits, while autosomes carry traits which are beneficial to females, likely due to their female-dominated life cycles (32-34).

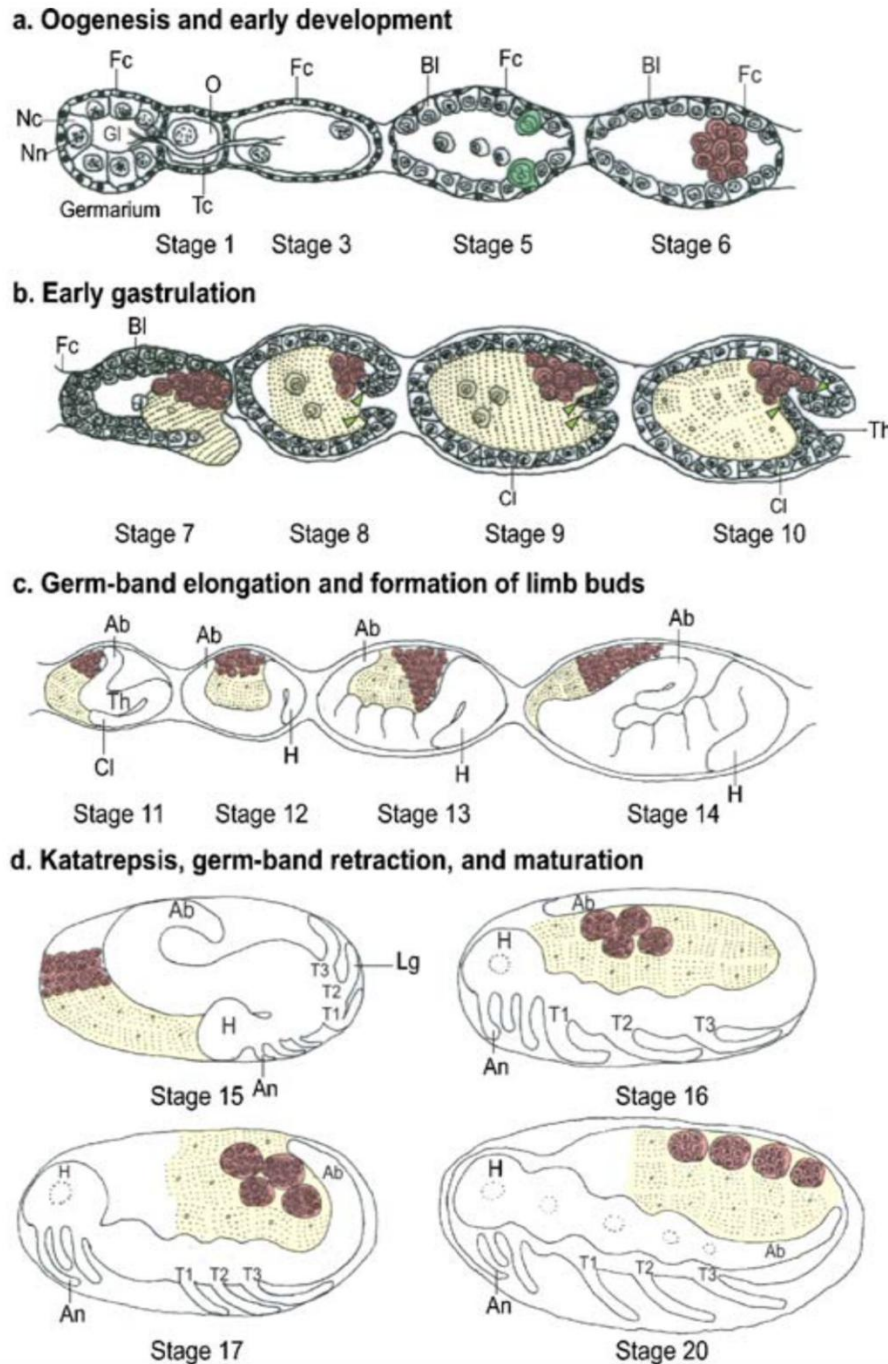


Figure 1.2: A diagram to show the stages of embryogenesis in the parthenogenic aphid ovary. **(a) Oogenesis and early development.** The germarium contains nurse cells surrounded by follicle cells (Fc). At stage 1, a single oocyte is surrounded Fc, with a single nucleus. A tropic cord (Tc) connects the germarium to stages 1-3. At stage 5 the blastoderm (Bl) forms, and the initial germ cells for the following generation (green) are present. At stage 6, multiple germ cells are present (red). **(b) Early gastrulation.** At stage 7, the bacteriocyte (yellow), housing *Buchnera aphidicola*, is introduced. Germline (red) migration is introduced via the invaginating germ band at stage 9 (green arrowheads). The cephalic lobe (cl) and initial thorax (Th) begin to develop at stages 9 and 10 respectively. **(c) Germ-band elongation and formation of limb buds.** Embryogenesis continues with development of the abdomen (Ab) and head (H). Limb buds also begin to form. This occurs across stages 11-14. **(d) Katatrepsis, germ-band retraction, and maturation.** The germline (red) is pushed to the anterior during katatrepsis (stage 15-16), then guided to the embryonic cavity within the abdomen (Ab). Thoracic segments (T1-3), and the antenna (An) form. Stage 20 represents a mature embryo ready for birth. This figure is adapted from Chang et al., 2007 with permission from the publisher (31).

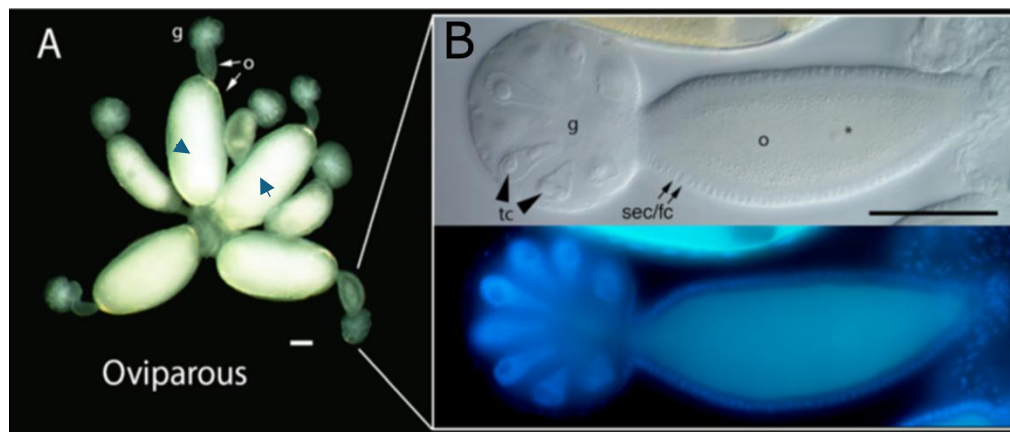


Figure 1.3: Images depicting an oviparous aphid ovary. (A) Bright field image of an oviparous aphid ovary. The germarium (g) sits posterior to 1 to 2 oocytes. More developed oocytes are yolk filled (examples marked by blue arrow heads). (B, upper panel) Magnified bright field image of the germarium and previtellogenic oocyte. Trophocytes (tc) are marked by black arrow heads. Somatic epithelial or follicle cells (sec/fc) make up the epithelium around the oocyte. (B, lower panel) DAPI stained view of upper panel, showing nuclei. The oocyte nucleus is barely visible over autofluorescence. This figure is adapted from Bickel et al., 2013 with permission from the publisher (35).

1.3 Whiteflies

Alongside aphids in the Sternorrhyncha suborder are another major group of sap-sucking crop pests, the whiteflies (Aleyrodidae). The silverleaf whitefly, *Bemisia tabaci* (Hemiptera), is a cryptic species complex consisting of at least 34 genetic groups, and 392 haplotypes (36). The Middle East-Asia Minor 1 (MEAM1) species complex is considered one of the most significant biotypes (37). They are a global polyphagous pest, responsible for crop losses due to feeding damage and viral infection. Moreover, *B. tabaci* infest over 1000 plant species and transmit more than 300 plant pathogenic viruses such as begomoviruses (38-44). Like aphids and other phloem-feeding Hemiptera, *B. tabaci* release oral secretions containing effectors upon feeding, which modulate plant defences (45). The *B. tabaci* life cycle consists of 6 stages: egg, 4 immature instar stages, and adult (46, 47) (Figure 1.4). Unlike aphids they are obligate sexual, oviparous reproducers. Moreover, they have a haplodiploid sex determination system, where males have only one copy of each chromosome (48). An advantage of haplodiploidy for researchers is that proof of principle gene editing is simpler to screen for, as a single edited allele can give rise to an observable phenotype in the haploid males (49). Recently, gene editing in *B. tabaci* has been achieved via Receptor Mediated Ovary Transduction of Cargo (ReMOT Control) (49). This method, involving injection of pre-made RNPs into the mother, which localise to the germline to perform gene editing, was necessary due to the small size of *B. tabaci* embryos and high mortality of injected eggs (49).

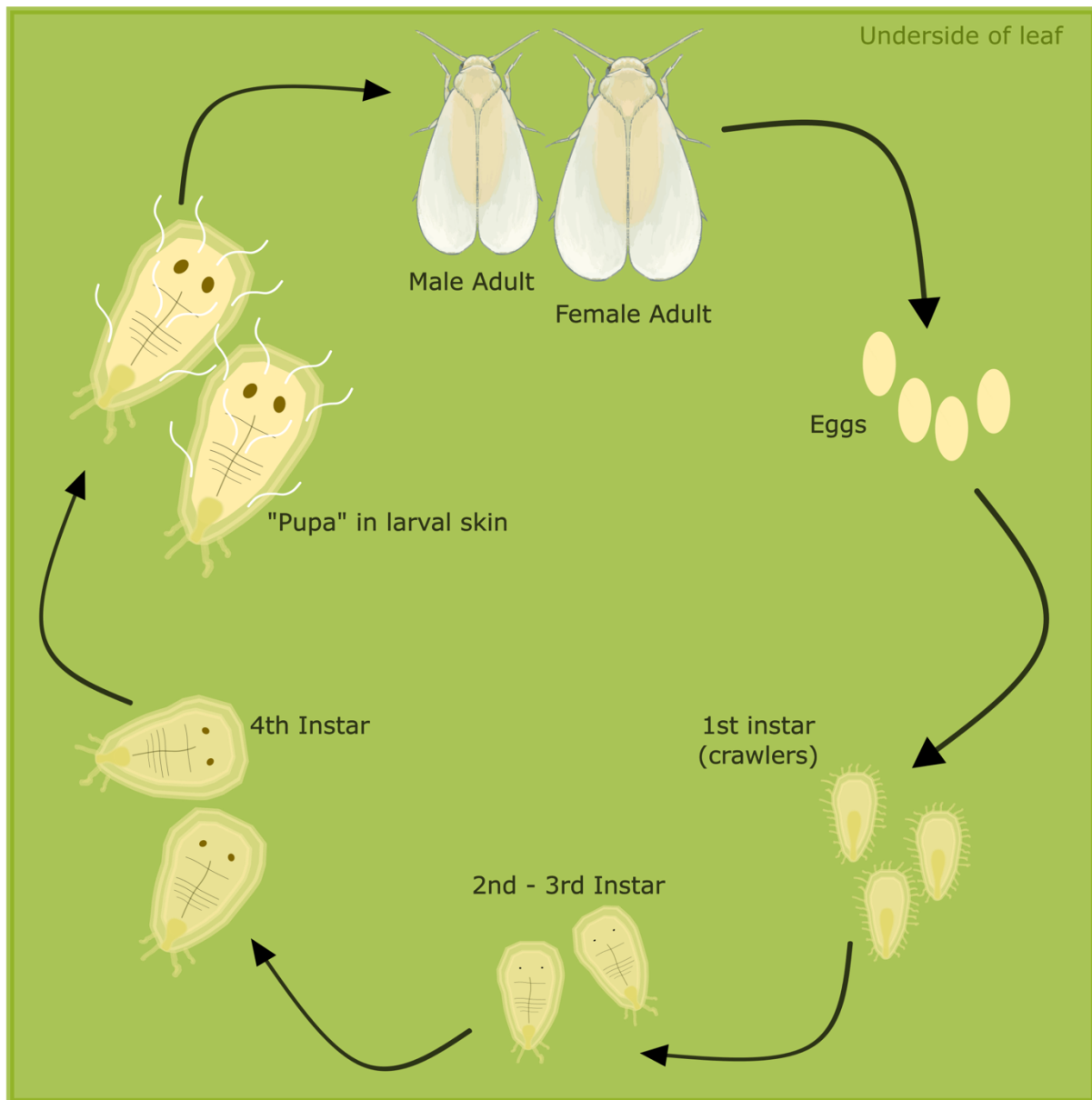


Figure 1.4: The lifecycle of *Bemisia tabaci*. Female adults lay eggs which attach to the underside of the leaf via the pedicel. From these eggs hatch 1st instar 'crawlers' which crawl to a spot on the underside of the leaf suitable for feeding where they anchor themselves for the remainder of larval development. During the 1st – 4th instar stages, development of the whitefly occurs inside a translucent cuticle or 'larval skin'. During 2nd-3rd instar phases, development continues, and eyes become visible as two small round spots. The 4th instar and "Pupa" stages are both in fact the same stage. The eyes enlarge into red brown, fully developed bipartite compound eyes. Meanwhile, wings develop. By the time of eclosion, the cuticle is almost transparent, with the pharate adult visible underneath. Adult male and female whitefly then emerge. This figure was produced by the author with information from Walker et al., 2010 (50).

1.4 Gene editing in insects

1.4.1 RNA interference

One method to study gene function is RNA interference (RNAi). RNAi induces gene knockdowns or silencing. Double stranded (ds)RNA is introduced – often by injection or ingestion – and is processed into silencing (si)RNAs by the enzyme Dicer. These 21-23 nucleotide RNAs become part of the RNA-induced silencing complex (RISC) which, guided by the siRNA, cleaves mRNA at the target site (51). In cleaving mRNA, RNAi can ‘silence’ or ‘knockdown’ a gene. This method has been employed for the study of aphid genes, often with the target of using RNAi as a pest control method (52-57). For aphids, plant-mediated RNAi is often used (55-57). The approach often involves generating transgenic plant lines expressing dsRNA of a target. The aphids then feed continuously on the transgenic plant, thereby receiving the dsRNA, initiating RNAi. Through use of RNAi, multiple genes, including *CCHa1R* (54), *Imf2-like* (55), *Mphb* (56) and *Ya* genes (57), have been knocked down, causing reduced fecundity and/or survivability in aphids. However, whilst expression of genes can be reduced by more than 50% using this method (55, 57), a complete, CRISPR-mediated gene knockout leads to a more pronounced phenotype, and can be better for the study of gene function. Further, RNAi can be inconsistent, with variation in gene expression making reproduction of results difficult. Therefore, whilst RNAi may be a useful pest control mechanism in future, CRISPR is a better tool for the study of gene function, and for finding an “Achilleas’ heel” target for pest control in aphids.

1.4.2 CRISPR/Cas systems

CRISPR (clustered regularly interspaced short palindromic repeats)/Cas (CRISPR-associated gene) systems evolved as a form of adaptive immunity in bacteria and archaea (58-62). The hallmark of this system is the CRISPR locus, which consists of short palindromic repeats separated by protospacers. Protospacers are short fragments derived from foreign DNA from invading pathogens, which are integrated into the CRISPR locus via recognition of protospacer adjacent motifs (PAMs) by Cas1, an integrase protein, and Cas2, an endoribonuclease protein (58, 60-64). The CRISPR locus is transcribed and then processed into mature CRISPR RNAs (crRNAs) from pre-crRNAs by endonucleases, either as part of Cascade (CRISPR-associated complex of antiviral defence), a protein complex made up of CasA, B, C, D and E, or as a single enzyme such as Cas6 (60, 61). These mature crRNAs form a complex with a Cas protein where they act as a guide to the complementary sequence of the invading pathogen. Upon arrival to this site, the Cas protein cuts the foreign DNA, deactivating the pathogen (58, 59). As of the most recent update, there are two classes containing six types and thirty-three subtypes (65). The most used genome editors belong to class 2. Class 2 CRISPR/Cas systems have a single large, multidomain protein, such as Cas9, which performs the same functions as the complexes made in class 1 systems; the effector protein is often involved in pre-crRNA processing, crRNA binding and target cleavage (65, 66). There are 3 types of CRISPR/Cas system in Class 2: type II, V and VI (Figure 1.5) (65).

Type II is the most used type of CRISPR/Cas system in programmable gene editing and is hallmarked by the protein Cas9 (58, 61, 65, 66). Cas9 was characterised as derived from

Streptococcus pyogenes and was shown to be involved in pre-crRNA processing (67) as well as target binding and cleavage (58). In this system, pre-crRNA is processed by the double-stranded (ds) RNA specific RNase III in the presence of Cas9; the process of crRNA maturation in this system is guided by a trans-activating crRNA (tracrRNA) which has complementarity to the repeat regions in the CRISPR locus (58, 67). It was also discovered that the base-paired structure left between the tracrRNA and the mature crRNA after this process was essential in activating Cas9's ability to cleave the target DNA (58). The target DNA is recognised due to complementarity to the crRNA, and then cleaved by the HNH and RuvC endonuclease domains of Cas9 (58).

Type V systems differ from type II systems in that their effector enzyme, Cas12, contains one RuvC-like domain that cleave both strands of the target DNA, rather than one RuvC and one HNH domain (65, 68-70). Moreover, Cas12a has a non-specific single-stranded deoxyribonuclease activity which is triggered by a conformational change caused by crRNA-target hybridization (69). Cas12a has become a popular choice for use in mammalian genome editing, showing signs of greater specificity than Cas9 (65, 68). However, it has not often been used in entomological studies. Recently, new Cas12 family members have been discovered that are significantly smaller than Cas9 and Cas12a. For example, Cas ϕ (Cas12j) discovered in biggiephages (71), and Cas12f discovered in *Acidibacillus sulfuroxidans* (72).

Type VI systems are hallmarked by Cas13, which contains two HEPN domains, which performs pre-crRNA into mature RNA as well as cleavage of the target DNA (65, 73). Further, Cas13 displays non-specific *trans* acting RNase activity upon activation, which

stops growth of the phage-infected host cell, preventing infection to other bacterial cells (65, 73). Cas13 has been shown to be effective at editing the *Drosophila* genome (74), and it's RNase activity has meant it has been useful in viral detection, including for SARS-CoV-2 (75, 76).

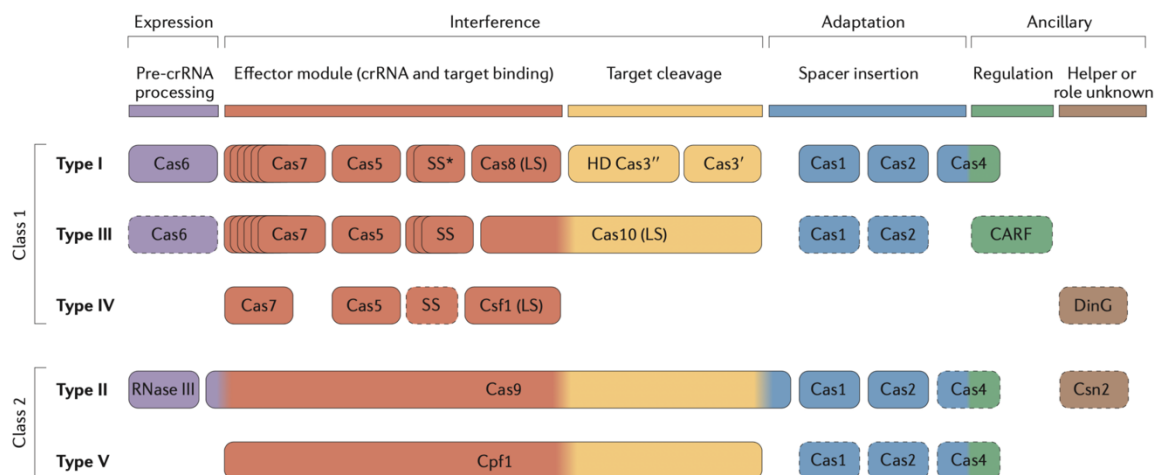


Figure 1.5: A schematic summary of the Class 1 and 2 CRISPR systems. Class 1 includes Type I, III and IV systems, while Class 2 includes Type II, V, and VI (not shown) systems. The main defining differences between the classes are that Class 1 systems include a multi-subunit effector and target cleavage module, while Class 2 systems achieve this with a single protein. Within Class 2, Type 2 systems use a single protein such as Cas9, with pre-crRNA processing from RNase III, while Type V systems use a single protein such as Cas12a (Cpf1) for both functions. This figure is reproduced with the publisher's permission from Makarova et al., 2015 (66).

In 2012, Nobel laureates Jennifer Doudna and Emmanuelle Charpentier published their ground-breaking study on the type II CRISPR/Cas system in *S. pyogenes*, wherein they demonstrated its programmable nature and its potential for use in gene editing (58). As previously described, Cas9 proteins hallmark type II systems, and are responsible for the cleavage of the target DNA (58). The protein is activated by the complementary base pairing of tracrRNA to mature crRNA in the ribonucleoprotein (RNP) complex. It is this RNA interaction which is key to the programmable nature of this system (58). Jinek et al., demonstrated that by synthesising a chimeric RNA comprised of the tracrRNA element and the crRNA element joined by a linker loop, termed a single guide RNA (sgRNA), one could design the crRNA element to guide an active Cas9 protein to a target of ones choosing and cleave it (58). After this, Cong et al., demonstrated that knock-ins were also possible using this system. By mutating the RuvC domain of Cas9, the protein is converted to a nickase, causing only a single stranded (ss) break rather than a ds break. The ss break is repaired efficiently by homology directed repair (HDR) (77). By providing a template for the HDR, one could introduce insertions and deletions into the genome (77). Over the following year, this system was shown to be efficient at genome editing through knockouts via cleavage, as well as knock-ins via homology directed repair (HDR) with a template, in various organisms, including bacteria, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, Zebrafish, *Drosophila melanogaster* and cultured human and mouse cells (77-85).

1.4.3 CRISPR/Cas systems in insects

The CRISPR/Cas9 system was first used in insects by Gratz et al., 2013, who demonstrated that the system works when deployed in *Drosophila melanogaster* and that one could use homology directed repair (HDR) to introduce genetic elements (80). The authors achieved this by co-injecting preblastodermal embryos with plasmids encoding Cas9 (phsp70-Cas9) and a sgRNA targeting the first exon of the *yellow* gene (pU6-BbsI-chiRNA) (80). The deletion of the *yellow* gene's function causes a characteristic yellowing of the adult cuticle and larval mouthparts and therefore successful knock-outs of this gene are easy to detect (80). Further, to assess the possibility of using HDR to introduce genetic elements, a donor template containing an attP ϕ C31 phage recombination site flanked by 60nt homology arms to the *yellow* gene was injected with the plasmids. This insertion can then be detected by PCR (80). Since then, the CRISPR/Cas9 system has been deployed in a variety of insects, including the malaria vector *Anopheles* mosquitoes (86-89), the disease vector mosquito *Aedes aegypti* (90) and, more recently, *Bemisia tabaci* (49).

Developing CRISPR/Cas systems as a tool in insects can provide both knowledge of gene function, as well as opportunity for gene-drives (i.e., the spread of a knock-out gene through a population) and pest insect management. Importantly, the efficiency of the CRISPR/Cas9 system in disease vectors opens the door for opportunities to reduce the spread of such diseases. For example, Hammond et al., (2016) demonstrated that by using microinjection of CRISPR/Cas9 constructs targeting the *Anopheles gambiae* *AGAP007280* loci, one can induce a gene drive with transmission rates to progeny of

91.4-99.6%; population modelling and cage experiments suggest that targeting this locus meets the minimum requirements for an effective gene drive (87).

Until 2018, the method of choice for transformation of insect lines by CRISPR/Cas was microinjection of preblastodermal embryos (26, 80, 81, 86, 87, 91, 92). Unfortunately, this method has notable drawbacks. Firstly, microinjection is challenging, and the equipment needed is expensive, and therefore not every lab has the facilities to perform this part of the methodology. Furthermore, the injection of preblastodermal embryos is often difficult or impossible, because in many insects, microinjection damages the eggs too much and many insects do not lay eggs synchronously or at all (90). Or, in some cases eggs are not laid until the post blastoderm stage (93).

1.4.4 CRISPR/Cas systems in aphids

One CRISPR/Cas9 transformation protocol exists for aphids and was developed for the pea aphid (*Acyrtosiphon pisum*). This method which relies on the sexual reproductive cycle of aphids, resulting in a 7-month long procedure (26). The first step is to force aphid females to switch their developing embryos from those that will undergo viviparous parthenogenesis to males and females which will partake in sexual reproduction. This is triggered by switching from a long-day (LD) photoperiod (18h light, 8h dark), simulating spring and summer, to a short-day (SD) photoperiod (12h light, 12h dark), simulating autumn. This must be done in such a way that males and females are perfectly synchronised so that mating is at its maximum resulting in enough fertilised eggs for microinjection. This step is followed by the mating and egg laying step, where males and

females were placed on a plant and allowed to mate, and sexual females began laying eggs around 3 days later, peaking after 5 days. Then eggs were microinjected with Cas9 and gRNAs, before being placed in a Petri dish with wet filter paper and moved to a climactic chamber at 15°C. Aphid eggs acquire a black serosal cuticle, necessary for cold resistance, via melanisation. Therefore, injected eggs are monitored for melanisation, with complete melanisation after 3-4 days indicating a non-disruptive injection. Eggs which did not complete melanisation were discarded. Eggs that were not damaged were either moved directly to plants or were treated to avoid infection before moving to plants. At this point, the plants were moved to a climactic chamber at 4°C for 85 days, simulating the 3-month obligate diapause period that aphid eggs typically undergo before hatching. After this, plants were moved back to 18°C LD conditions and monitored for hatching aphids. These aphids were individualised to single plants until they began viviparous parthenogenesis, at which point the clonal lineage was maintained by isolating 2 individuals per generation to a new plant (26). Each clonal lineage was screened for mutations via PCR and sequencing of the target region (*Stylin-01*)(26). After melanisation mutation rates were 70-80%, but only 1-11% of the injected eggs hatched. A total of 17 lineages were obtained, 6 of which had been mutated (2 homozygous and 4 heterozygous) and the final germline transmission of the mutations was ~35% (26). This method was therefore shown to be relatively efficient at generating transgenic lines but has many drawbacks. The foremost of these is the fact that it takes 7 months of careful experimentation to achieve the clonal lineages. It is true that once edited, the resulting clonal aphid lineage can be maintained for long periods of time. However, if one then decided to mutate a second site (a related gene up or downstream for example) it would take another 7 months to do so. This method's drawbacks are

directly related to the fact that the current methodology for CRISPR/Cas9 mediated transformation in insects relies on microinjection of eggs, which is challenging and often causes damage to the eggs.

1.5 ReMOT Control

In 2018, a novel method was described for delivery of Cas9 and sgRNAs to embryonic insect cells, which negates the need for challenging egg microinjection and opens the potential for many more insect species to be studied using gene editing tools (90). The method, termed Receptor Mediated Transduction of Cargo (ReMOT) control, involves injecting adult females with a chimeric protein made up of an ovary-targeting peptide fused to Cas9 to facilitate uptake of the RNP into the developing embryo inside the adult to increase the chance of germline editing (90). The first example of a yolk-targeting peptide used was P2C, a 41 aa peptide derived from the *Drosophila melanogaster* yolk protein (DmYP1). To find this peptide DmYP1 was fragmented into fragments P2-P6 which were individually fused to EGFP in a pAc5-STABLE1-neo backbone, which was injected into the haemolymph of vitellogenic *Anopheles gambiae* females. The ovaries were then dissected after 16h and screened for EGFP fluorescence; the N-terminal P2 fragment showed the highest transduction into the ovaries. P2 was then further fragmented into P2A, P2B and P2C and subjected to a similar study; the fragments and EGFP were cloned as fusions into pET28a-Cas9-Cys. These were expressed in *E. coli* where, and the protein was purified. Upon injection, it was found that P2C showed the greatest transduction to the ovaries (90). It was then demonstrated that P2C would not affect the action of Cas9 when expressed as a fusion protein by standard microinjection

of preblastodermal embryos with an sgRNA targeting the kynurenine monooxygenase (*kmo*) gene. When homozygous, the recessive white *kmo^w* alleles produce white-eyed female mosquitos, making screening for successful edits easy to do immediately upon egg hatching (90). This was done by targeting two separate sites in the *kmo* gene one at a time. Mosaic individuals resulting from the first target indicate only one allele was edited. These were subject to editing with the second target to confirm this (90). Ultimately, the idea that P2C could chaperone Cas9 with an sgRNA to the embryos and cause stable gene edits was tested. The first problem faced was that without an endosomal release reagent (EER) no editing was observed. As a result, reagents used in drug delivery studies were tested and it was found that chloroquine at a concentration of 0.5-2 mM was sufficient to alleviate this problem (90). The P2C-Cas9 fusion was then mixed with sgRNA constructs along with the EER and injected intrathoracically, whereupon the P2C domain delivered the RNP to the embryos where gene editing takes place. Offspring were screened for edits by observation of the eyes as previously described (90). Heritability of the induced *kmo* mutation was checked by crossing G₀ individuals that had been edited successfully, with Wh-Iso8-*kmo*⁴⁶⁰ individuals and screening them for white eyes. Sequencing demonstrated that 100% of the progeny carried both the 460 and 519 site mutations (90).

This method was first demonstrated in *Aedes aegypti* and *Anopheles gambiae* (90), but has since been shown to be effective in many other species from diverse orders including other Diptera such as *Culex pipiens* and *Anopheles stephensi* (94, 95), Hymenoptera (*Nasonia vitripennis*) (96), Coleoptera (*Tribolium castaneum*) (97), Hemiptera (*Rhodnius prolixus*, *Bemisia tabaci* and *Diaphorina citri*) (49, 98, 99), Lepidoptera (*Bombyx mori*)

(100), and even the black legged tick, *Ixodes scapularis* (Arachnida) (101) demonstrating its potential for use in multiple insect species that would previously have been more difficult to study. This is largely due to the lack of reliance of microinjection of insect eggs, which are much smaller than the abdomen, and are easily damaged. A comparison of ReMOT control with classical embryo injection-based gene editing in mosquitos is presented in Figure 1.6 (90).

In some cases, including in *B. tabaci* and *R. prolixus*, P2C did not perform robustly as a chaperone for Cas9, and therefore part of the binding sequence of *B. tabaci* vitellogenin (BtVg) was used (49, 98). This was found by multiple sequence alignment against other insect Vgs and a known vitellogenin receptor (VgR) binding domain discovered by peptide array in the giant freshwater shrimp *Macrobrachium rosenbergii* (49, 102). This peptide was found to be effective in both *B. tabaci* and *R. prolixus* when targeting the gene *white* (49, 98).

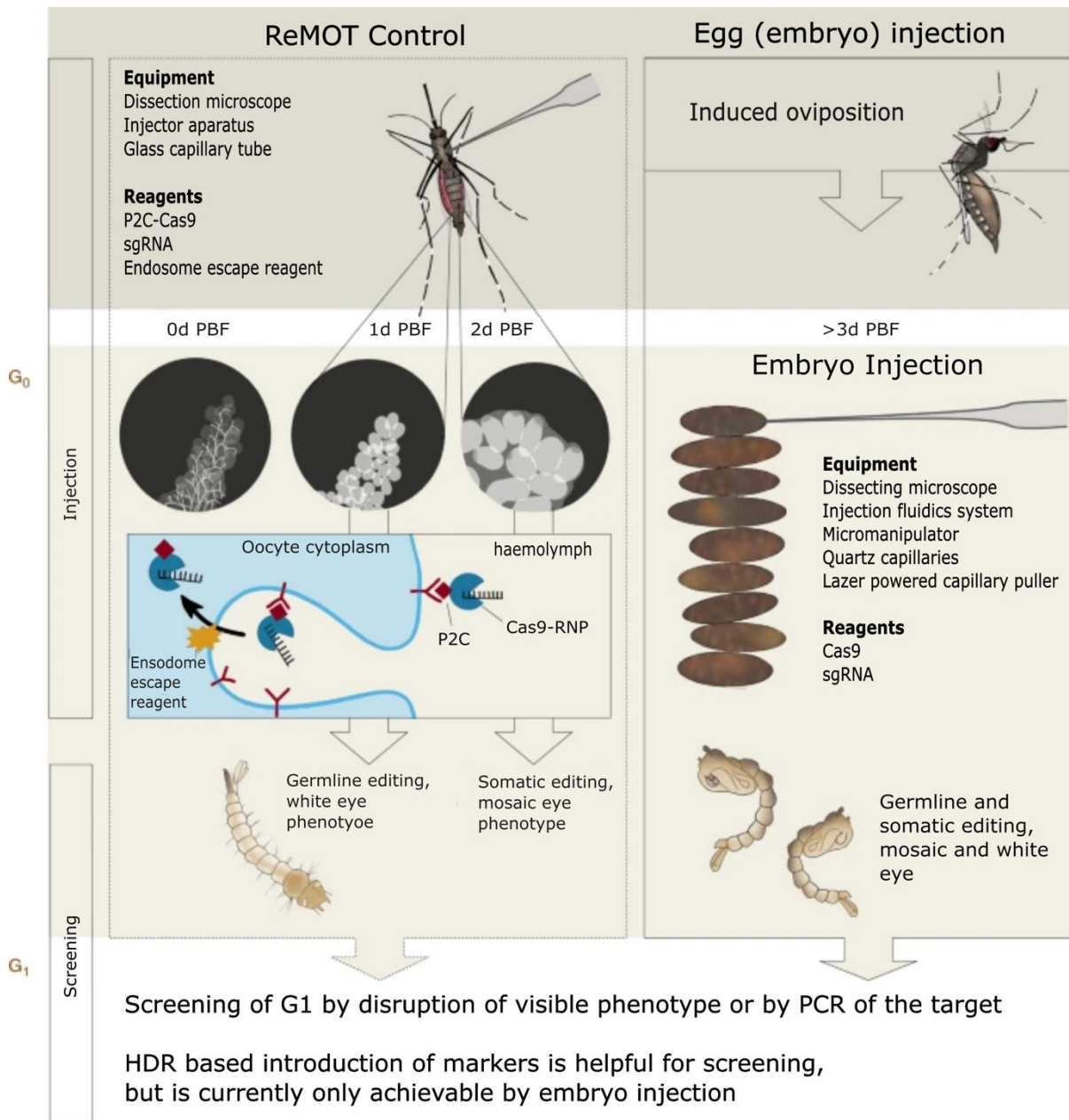


Figure 1.6: A comparison of ReMOT control and embryo injection methods of gene editing in mosquitos. For embryo injection, oviposition must be induced, occurring >3 days (d) post blood feed (PFB). Then, highly specialised quartz needles must be used to inject the eggs one by one under a high magnification microscope. Emergent offspring can be germline or somatically edited. Using the white gene as an example target here, this would result in both full and mosaic white-eye phenotypes. Screening for edits from this method can be done by disruption of a marker and PCR, as well as insertion of a marker by homology directed repair (HDR). For ReMOT control, adult females are injected with engineered P2C-Cas9 ribonucleoprotein (RNP) complexed with the white-targeting sgRNA, and an endosome escape reagent (EER) 1-2 days PBF. The P2C enables receptor mediated endocytoses into the oocytes, and the EER enables endosomal escape into the cytoplasm for access to the nucleus. Offspring of mothers injected 1d PBF are germline edited, and have a full white eye phenotype, while those from mothers injected 2d PBF are somatically edited and have a mosaic phenotype. This phenotype is used for screening, along with PCR. However, with ReMOT control, insertion of a marker via HDR is not possible. This figure was adapted from Chaverra-Rodriguez et al., 2018 (90) under creative commons liscence <http://creativecommons.org/licenses/by/4.0/>.

1.6 Vitellogenin

Vgs are phospholipoglycoproteins which are, as a class of yolk protein, in part responsible for maintaining the nutrition requirements of the developing embryo (103, 104). In insects, these proteins are synthesised in the fat body, are proteolytically cleaved and heavily co- and post-translationally modified to facilitate transport of nutrients (carbohydrates and lipids for example) before release towards the ovaries (103, 104). After this cleavage the large (140-190kDa) and small (40-60kda) subunits assemble into oligomeric proteins (usually dimers) for transport to the oocytes (104). Upon arrival to the oocytes Vg is taken up via receptor mediated endocytosis by Vitellogenin Receptors (VgRs), which are members of the low-density lipoprotein receptor (LDLR) family (103). After incorporation into the oocytes, Vgs are stored in a crystalline form as vitellins (Vns) which act as a food store for the embryo (104). Vitellogenins belong to the broader family of LDLs, that also includes Vitellogenin-like proteins and apo-lipophorins. Insect Vgs usually have four distinct domains: (i) an N-terminal signal peptide, (ii) a lipoprotein N-terminal domain, (iii) a domain of unknown function (DUF1934) and (iv) a Von Willebrand factor domain (VWD) (105). Notably, some aphids, including *Myzus persicae* and *Diuraphis noxia* have been reported to lack DUF1934 and have generally smaller N-terminal lipoprotein domains (105). Furthermore, In a phylogenetic analysis of Vitellogenins, aphid Vgs show discordance with the species phylogeny grouping together as an outgroup to all other insect Vgs and not in the same clade as other hemipteran Vgs (105) suggesting something interesting has happened in the evolutionary history of aphid Vgs. Insect VgRs contain multiple domains: LDLR domain class A; epidermal growth

factor (EGF) and growth factor-like domains; LDLR YWTD domains; calcium binding EGF-like domains; a transmembrane region and some low complexity regions. Unlike Aphid Vgs, Aphid VgRs appear to group with other hemipteran VgRs when subject to phylogenetic analysis (105). In the VgR, LDLR domain class A repeat clusters are responsible for ligand binding. These repeats are largely composed of negatively charged residues, which are thought to interact with positively charged residues on the ligand (103). Further, each repeat has 6 cysteine residues which form disulphide bonds, the abolition of which has been shown to abolish Vg-VgR interactions (102, 106). The binding domain of Vg itself is localised to the vitellogenin-N domain, specifically in the β -sheet domain (102). Upon binding, Vg-VgR complexes localise to clathrin coated pits which pinch off intracellularly to form vesicles. These vesicles travel to an endosome, where acidification results in the dissociation of the Vg-VgR complex. The receptor is then recycled to the cells surface while the Vg is stored as vitellin (Vn) in mature yolk bodies for future use (104).

Regulation of vitellogenesis by signalling pathways is well understood (but not in aphids). The most prominent signalling pathways are controlled by juvenile hormone (JH), amino acid/target-of-rapamycin (AA/TOR), and ecdysteroid 20-hydroxyecdysone (20E) (107). In some cases, such as mosquitos, vitellogenesis is activated by feeding, which in turn activates the AA/TOR and insulin pathways (107-112). These pathways then upregulate both the JH and 20E pathways (107). This is relevant to ReMOT control-based gene editing as the method relies on active vitellogenesis (90). Therefore, understanding the regulation of vitellogenesis could lead to optimisations of injection timing in ReMOT

control (90). For example, mosquitos are injected 24h post blood meal to coincide with the height of vitellogenesis (90).

JH primarily interacts with methoprene-tolerant (Met), a member of the basic helix-loop-helix (bHLH)-Per-Arnt-Sim (PAS) family of transcription factors. Met dimerises with another bHLH-PAS protein, Taiman (Tai) (or its orthologs FISC or *p160*/steroid receptor coactivator [SRC]), to regulate JH responsive genes (107, 113-115). One JH responsive gene is *Krüppel-homolog 1* (*Kr-h1*) which codes for a zinc finger transcription factor. *Kr-h1* has been shown to have different effects on vitellogenesis in different insects: in *Aedes aegypti* *Kr-h1* both activates and represses different JH responsive genes to regulate vitellogenesis and egg maturation (116), however in *Cimex lectularius* *Kr-h1* knockdown by RNAi appeared to have no significant effect on vitellogenesis (117). The JH-Met/Tai complex is known to induce massive polyploidy in cells of adult female *Locusta migratoria* to enable massive Vg synthesis (118). The genes involved in this process are *chromosome maintenance genes 3, 5, 7* (*Mcm3*, *Mcm4*, *Mcm7*), *cell-division-cycle 6* (*Cdc6*), *cyclin-dependant kinase 6* (*Cdk6*) and *adenovirus e2 factor-1* (*E3f1*) (118-120). To activate their transcription the JH-Met/Tai complex binds E-box or E-box-like motifs in the promoters of these genes. Knockdown of *Mcm4 and 7*, *Cdc6*, *Cdk6* and *E2f1* dramatically decreased Vg expression (118-120). Another way that JH induces polyploidy in *L. migratoria* is through dephosphorylation of FoxO via leucine carboxyl methyltransferase 1 (LCMT1)-mediated protein phosphatase 2A (PP2A) activation. This FoxO dephosphorylation allows it to travel to the nucleus where it activates transcription of *cell-division-cycle 2* (*Cdc2*) and *origin-recognition complex subunit 5* (*Orc5*). Knockdown of any of these genes resulted in reduced ploidy in the fat body and

consequently reduced Vg expression (121). In *A. aegypti*, to further upregulate Vg expression, JH acts through Met to induce *regulator of ribosome synthesis 1 (RRS1)* and *ribosomal protein L32 (RpL32)* expression, which increases ribosomal biogenesis and therefore allows massive Vg synthesis (122). The repressive functions of JH are also accounted for; JH induces *Hairy* expression which dimerises with *Groucho* to form a transcriptional repressor complex (123).

20E acts via a dimer of the ecdysone receptor (EcR) and ultraspiracle (Usp) (109, 124-126). The EcR/Usp dimer exists at low 20E levels as a dimer which acts as a repressor. However, when 20E levels increase EcR/Usp becomes ligated and consequently undergoes conformational change which allows it to bind DNA with greater affinity and recruit coactivators, allowing it to activate gene expression (109, 125, 126). The DNA binding sites involved in this are called ecdysone responsive elements (EcREs) and consist of two nucleotide hexamers (AGGTCA) which are separated by a varied number of nucleotides and form inverted or direct palindromes (109). In *A. aegypti* EcREs can be found in the 5' regulatory region of Vg, suggesting that EcR/Usp is directly responsible for upregulating its synthesis (109, 111). However, EcR/Usp also upregulates E74B, E75A, *Broad complex* (BrC)-Z2 and HR3, which can all bind sites in the *AaVg* promotor, suggesting that these transcription factors act in combination with EcR/Usp to upregulate Vg expression (109, 111).

Vitellogenesis is not well characterised in aphids. One study in the brown citrus aphid *Aphis citricidus* suggests that aphid Vgs follow a canonical expression pattern, with high expression in the fat body of asexual adult females (105). Coupled with high VgR

expression in nymphs and ovaries, this implies that Vg plays a role in embryo development of aphids. However, in asexual aphids the need for Vg-associated reproductive and developmental functions is likely decreased, or even abolished, as the embryos develop inside the mother in filamentous ovaries comprised of ovarioles containing multiple embryos at different developmental stages (28). Arising from their sexual reproductive cycle, eggs of multiple aphid species have been described as being filled with yolk upon deposition (28), indicating vitellogenesis plays a role in egg development. Understanding Vg and vitellogenesis in aphids would inform whether a Vg-derived peptide is likely to be suitable for ReMOT control in aphids.

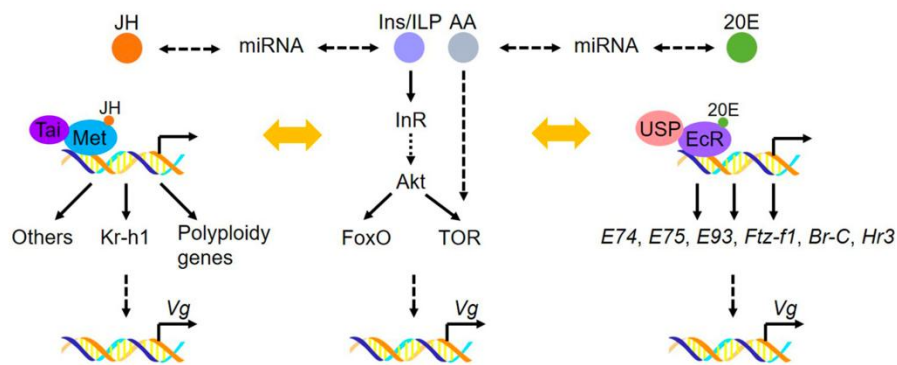


Figure 1.7: Hormonal and nutritional control of Vg synthesis in the fat body. Juvenile hormone (JH) interacts with Met/Tai (or SRC) to upregulate Kr-h1 and polyploidy genes, increasing Vg transcription. Ecdysone (20E) interacts with EcR/USP, which mediates upregulation of 20E responsive genes, which lead to increased Vg transcription. Insulin/Insulin-like peptides (ILPs) and nutrition in the form of amino acids (AA) mediate Vg transcription via interplay with both the JH and 20E pathways, through FoxO and TOR. This figure is adapted from Wu et al., 2021 (107) under creative commons licence <http://creativecommons.org/licenses/by/4.0/>.

1.7 Optimising tools to study gene function in aphids and whiteflies

This thesis aims to develop and optimise ReMOT control systems for use in aphids and whiteflies. No ReMOT control system has been developed for use in aphids; this thesis will present and discuss work completed to design and optimise a ReMOT control method in the green peach aphid, *M. persicae*. In pursuit of this goal, *M. persicae* Vg (MpVg) is characterised to find a suitable ovary targeting peptide with which to fuse Cas9. Chimeric mCherry fusion proteins are produced to analyse the embryo targeting properties of Vg derived peptides. Chimeric Vg-peptide-Cas9 fusions are produced and are injected into adult female aphids. Offspring are screened for successful genome editing. ReMOT control has been used for successful gene editing in *B. tabaci* (49). Using this system as a basis, the use of Cas ϕ as an alternative editor to Cas9 in ReMOT control is studied. The design and optimisation of these methods will enable new fundamental studies on *M. persicae* and *B. tabaci* concerning, for example, effectors secreted by these insects which contribute to their success as agricultural pests. This may enable the discovery of an ‘Achilleas heal’ to exploit for control and management of these pests.

1.8 Aims

- Characterise *M. persicae* vitellogenin evolution and possible functions
- Mine a *M. persicae* Vg receptor binding ligand for use in ReMOT control, and analyse its embryo-targeting ability
- Produce chimeric VgR-ligand-Cas9 ligands which enable germline gene editing in *Myzus persicae* via adult injection
- Explore potential improvements to ReMOT control in *B. tabaci* using an alternate BtKV-Cas ϕ protein chimera

1.9 Contributions to this thesis

- In Chapter 2, scripts used to generate expression heatmaps were provided by George Seddon-Roberts, with minor tweaking by myself.
- Injections described in Chapter 5 were primarily carried out by me, and the experimental design and technical set up was designed by me. However, some help was given by Dr. Sam Mugford and Dr. Mar Marzo in carrying out injections. Further, some DNA extractions were carried out by Dr. Mar Marzo.
- In chapter 6, concerning ReMOT in whitefly using BtKV-CasPhi2, experimental work was carried out by year in industry student Amber Hall. The project and experiments were designed and supervised by me. Data interpretation, figures and all writing was done by me.

2 *Myzus persicae* vitellogenin harbours a region that may be used for REMOT control in aphids.

2.1 Introduction

Gene editing by ReMOT control requires the targeting of Cas9-sgRNA ribonucleoprotein complexes (RNPs) to the insect germ line. This is via the use of chimeric Cas9 proteins engineered to include a germ line-targeting peptide. These chimeric Cas9 proteins are then complexed with sgRNAs in vitro, before injection into mother insects. The included peptide then guides the complex from the injection site to the germ line (90). The embryo-localising peptides in these chimeric proteins are derived from yolk precursor proteins (YPPs) such as DmYP1 or vitellogenin (Vg) (49, 89, 90, 94-101). By far the most often used peptide is P2C, derived from DmYP1 by deletion assay (90). P2C has been effective in ReMOT control studies on a variety of insects including multiple mosquitos (Diptera), *Nasonia vitripennis* (Hymenoptera), *Tribolium castaneum* (Coleoptera), the Asian citrus psyllid *Diaphorina citri* (Hemiptera) and even one arachnid – the black legged tick *Ixodes scapularis* (89, 90, 94-97, 99, 101). In the close aphid relative, *Bemisia tabaci* (Hemiptera), P2C was ineffective, whereas ‘BtKV’, derived from the receptor binding domain of *B. tabaci* Vg (BtVg) was successful in enabling gene editing by ReMOT control (49). BtKV was also used for ReMOT control in another hemipteran: *Rhodnius prolixus* (98). BtKV was found by multiple sequence alignment (MSA) against the Vg receptor (VgR) binding region of *Macrobrachium rosenbergii* Vg, found by peptide array (49, 102). However, the most closely related species to *M. persicae* to have a successful ReMOT

control system is the Asian citrus psyllid, *Diaphorina citri*, for which P2C was used successfully (99).

In insects, Vgs are synthesised in the fat body, where they are post-translationally modified and released into the haemolymph, whereupon they travel to oocytes and are taken up via receptor-mediated endocytosis upon interaction with the VgR (104). Vgs make up the bulk of YPPs in most insect oocytes (103). As such, many studies on insect Vgs have been carried out, including hemipteran species such as *B. tabaci* and *N. lugens* (127, 128).

Canonical vitellogenins are made up of an N-terminal signal peptide and 3 distinct domains. These are the Vg-N, DUF1943, and VWD. The Vg-N domain is further split into two subdomains: the α -helical domain and the β -barrel domain (103, 104). The Vg-N domain has been associated with VgR binding (102). In the VgR, LDLR domain class A repeat clusters are responsible for ligand binding. These repeats are largely composed of negatively charged residues, which are thought to interact with positively charged residues on the ligand (103). Further, each repeat has 6 cysteine residues which form disulphide bonds, the abolition of which has been shown to abolish Vg-VgR interactions (102, 106). Upon binding, Vg is taken up by receptor-mediated endocytosis. Vg-VgR complexes localise to clathrin coated pits which pinch off intracellularly to form vesicles. These vesicles travel to an endosome, where acidification results in the dissociation of the Vg-VgR complex. The receptor is then recycled to the cells surface while the Vg is stored as Vn in mature yolk bodies for future nutrition (104). The functions of the DUF1943 and VWD domains are less well understood. Recently, studies have shown

how these domains function. Perhaps unexpectedly, given its function as a yolk protein, Vg has been described as a pattern recognition receptor (PRR) which can trigger innate immune responses (129-131). In the fat greenling fish, *Hexagrammos otakii*, by binding to pathogen associated molecular patterns (PAMPs) such as lipopolysaccharide (LPS) and peptidoglycan (PGN), Vg can act as opsonin, promoting phagocytosis by macrophages (129). Further, a study in the crab *Eriocheir sinensis* found that Vg specifically binds bacteria and exerts immune functions via both the VWD, and that binding of EsplgR via the DUF1943 promotes phagocytosis (130). Moreover, purified recombinant scallop *Patinopecten yessoensis* Vg domains DUF1943 and VWD were able to bind LPS and LTA on bacterial cell wall. Also, purified native PyVg displayed antibacterial activity against both gram positive and negative bacteria (131). Similar findings have been shown in insects. In *Bombyx mori*; BmVg inhibited growth of gram negative *E. coli* and gram positive *Bacillus subtilis* (132). Further, in the honey bee, *Apis mellifera*, it has been shown that Vg can bind immune elicitors, and transport them to the egg, priming the developing embryo for immune activation (133). Taken together, these findings suggest that Vgs have a conserved immunological function, including in insects.

More recently, vitellogenin has been shown to be an effector protein, suppressing H₂O₂ burst when secreted into plants via the salivary glands upon feeding by the planthoppers *N. lugens*, *Laodelphax striatellus* and *Sogatella furcifera* (134, 135).

Further, in *Nephotettix cincticeps*, Vg has been shown to aid transmission of viruses to plant phloem via feeding (135).

The Vgs of many insects are cleaved once in the fat body at an RXXR motif, usually found in the Vg-N domain of the protein (103, 104). In the honey bee, *Apis mellifera*, the cleavage site is found in the polyserine linker between the two subdomains of Vg-N (136). It remains unclear whether MpVg contains an RXXR motif or if it undergoes cleavage.

Signalling pathways that give rise to vitellogenesis are well understood in some insects. The most prominent signalling pathways are controlled by juvenile hormone (JH), ecdysteroid 20-hydroxyecdysone (20E) and nutrition (in the amino acid/target of rapamycin [AA/TOR] pathway) (107). In some cases, such as mosquitoes, vitellogenesis is activated by feeding, which in turn activates the AA/TOR and insulin pathways (107-112). These pathways then upregulate both the JH and 20E pathways (107). While these pathways are understood in some insects, including *R. prolixus*, *L. migratoria*, and *A. aegypti*, (107, 113-115) no characterisation of these pathways/genes has been done thus far for aphids.

JH primarily interacts with methoprene-tolerant (Met), which dimerises with Taiman (Tai) (or its orthologs FISC or *p160*/steroid receptor coactivator [SRC]), to regulate JH responsive genes (107, 113-115). One JH responsive gene is *Krüppel-homolog 1* (*Kr-h1*) which has been shown to have different effects on vitellogenesis in different insects: in *Aedes aegypti* *Kr-h1* both activates and represses different JH responsive genes to regulate vitellogenesis and egg maturation (116), however in *Cimex lectularius* *Kr-h1* knockdown by RNAi appeared to have no significant effect on vitellogenesis (117). The JH-Met/Tai complex is known to induce massive polyploidy in cells of adult female *Locusta migratoria* to enable massive Vg synthesis (118). The genes involved in this

process are *chromosome maintenance genes 3, 5, 7* (*Mcm3*, *Mcm4*, *Mcm7*), *cell-division-cycle 6* (*Cdc6*), *cyclin-dependant kinase 6* (*Cdk6*) and *adenovirus e2 factor-1* (*E3f1*) (118-120). To activate their transcription the JH-Met/Tai complex binds E-box or E-box-like motifs in the promoters of these genes. Knockdown of *Mcm4* and *7*, *Cdc6*, *Cdk6* and *E2f1* dramatically decreased Vg expression (118-120). Another way that JH induces polyploidy in *L. migratoria* is through dephosphorylation of FoxO via leucine carboxyl methyltransferase 1 (LCMT1)-mediated protein phosphatase 2A (PP2A) activation. This FoxO dephosphorylation allows it to travel to the nucleus where it activates transcription of *cell-division-cycle 2* (*Cdc2*) and *origin-recognition complex subunit 5* (*Orc5*). Knockdown of any of these genes resulted in reduced ploidy in the fat body and consequently reduced Vg expression (121). In *A. aegypti*, to further upregulate Vg expression, JH acts through Met to induce *regulator of ribosome synthesis 1* (*RRS1*) and *ribosomal protein L32* (*RpL32*) expression, which increases ribosomal biogenesis and therefore allows massive Vg synthesis (122). Finally, JH induces *Hairy* expression which dimerises with *Groucho* to form a transcriptional repressor complex, which mediates repressive JH function (123).

20E acts via a dimer of the ecdysone receptor (EcR) and ultraspiracle (Usp) (109, 124-126). The EcR/Usp dimer exists at low 20E levels, which acts as a repressor. However, when 20E levels increase EcR/Usp becomes ligated allowing it to bind DNA and activate gene expression (109, 125, 126). The DNA binding sites involved in this are called ecdysone responsive elements (EcREs) (109). In *A. aegypti* EcREs can be found in the 5' regulatory region of Vg, suggesting that EcR/Usp is directly responsible for upregulating its synthesis (109, 111). However, EcR/Usp also upregulates E74B, E75A, *Broad complex*

(BrC)-Z2 and HR3, which can all bind sites in the *AaVg* promotor, suggesting that these transcription factors act in combination with EcR/Usp to upregulate *Vg* expression (109, 111).

All the arthropods where ReMOT control has been successful have been sexual reproducers which lay yolk-filled eggs. Most of the time, aphids reproduce asexually through parthenogenic viviparity, with adult females giving birth to multiple clones without fertilisation. These aphid morphs are called ‘*virginoparae*’ (14, 22-25). In asexual reproducing aphids, the embryos develop inside the mother in filamentous ovaries comprised of ovarioles containing multiple embryos at different developmental stages (28). Stage 1-4 represent oocytes that are pre-blastoderm-formation and are connected to the germaria (stage 0) by a trophic chord. At stage 5 the blastoderm is formed, and at stage 6 the maternal bacteriocyte containing *Buchnera aphidicola* is incorporated (28). It has been shown that aphids express *VgR*, which is conserved from other hemipteran species (105). Therefore, it is possible that the first 4 stages are fed nutrition via *Vg* uptake by *VgR* and transport along the trophic cord. Indeed, in the closely related *Bemisia tabaci*, *Vg* uptake occurs both via the trophic cord, and via receptor mediated endocytosis (137). Presumably, from stage 5 onwards the embryos can obtain nutrients from other sources, possibly by further *Vg* uptake, uptake of other resources, or via *B. aphidicola*. It has been shown that the trophic cord in sexual ovaries is much larger than in asexual ovaries (29), perhaps to accommodate the greater need for *Vg* uptake and transport to sexual oocytes, destined to become overwintering eggs. If all nutritive yolk is delivered by transport along the trophic cord, it is possible that *Vg* is only taken up in germaria, and oocytes receive it solely from the trophic cord.

To this point, the only aphid Vg that has been studied is from the brown citrus aphid *A. citricidus* (105). In this study, AcVg and AcVgR show morph- and tissue-specific expression patterns, consistent with the canonical reproductive roles of Vg. AcVg was predominantly expressed in the fat body of adults, while AcVgR was primarily expressed in the oocytes. However, there were some points which raised further questions. A phylogenetic analysis of insect Vgs revealed that aphid Vgs are distantly related to those of other Hemiptera. Furthermore, aphid Vgs lack the DUF1943 domain, making them significantly smaller than most insect Vgs that possess this domain. The DUF1943 domain may explain why aphid Vgs group distinctly from other hemipteran Vgs (12). The reason for the loss of the DUF1943 domain is described as 'unclear' (105). Other hemipterans possess multiple Vg-like genes in addition to the canonical Vg (127, 128). In a phylogeny that includes Vg and Vg-like genes in *N. lugens*, the Vgs of *D. noxia*, *A. pisum*, and *A. medicaginis* cluster with Vg-like proteins rather than canonical Vg proteins (128), suggesting that aphid Vgs may be ancestral to Vg-like proteins in other insects and that the canonical Vg has been lost in aphids. In the successful ReMOT control gene editing of *B. tabaci*, the peptide 'BtKV' is derived from canonical BtVg, rather than BtVg-like (49). The study with *N. lugens* also constructed a phylogeny of insect VgRs, revealing that aphid VgRs are well conserved, contain the expected domains, and cluster with other hemipteran VgRs (105).

Here, we identified a Vg-like protein in *M. persicae* (MpVg), but not a canonical Vg. The structural domains were also analysed. Further, homologues of signalling pathway genes were identified, and expression was analysed. By analysing the structural and

sequence similarities of MpVg compared to other Vgs, the region MpRV was identified as a potential VgR binding region, which could be used for ReMOT control in *M. persicae*.

2.1.1 Contributions to this chapter

Dr. Rea Antoniou-Kourounioti provided the script to perform Alphafold 2 analysis in the command line. Dr. Sam Mugford provided the script to produce pLDDT plots in R, and I amended it for my use.

2.2 Results

2.2.1 Aphids have lost the ‘canonical’ insect vitellogenin, but have retained vitellogenin-like

To begin searching for an alternative to ‘BtKV’ in *M. persicae* that could be used for ReMOT control, I first searched for a Vg homologue in *M. persicae* and other aphids. Upon searching 19 aphid genomes, including *M. persicae*, (all these genomes had BUSCO scores of >C:95% except *M. persicae* 2.1 which has C:91.4%. Further information on BUSCO scores is in the appendix) for Vg and Vg-like homologues in aphids, no canonical Vg protein sequences were found. However, Vg-like protein sequences were identified in all aphid proteomes searched. Further, in some species, two transcripts were discovered which differed by the addition of a single serine residue at position 588. This suggests that aphids do not have the canonical Vg gene. To further investigate this, a phylogenetic tree of hemipteran Vg and Vg-like protein sequences, including aphid Vgs, was produced, which shows that all aphid sequences included group with Vg-like proteins rather than canonical Vg proteins (Figure 2.1). In searching for hemipteran Vg and Vg-like homologues, many sequences were found. For many species, homologues for both canonical Vg and Vg-like were identified; other than aphids, the only other species that contained a Vg-like, but not a canonical Vg homologue was the grape phylloxera (*D. vitifoliae*). All sequences used in the phylogeny are presented in the appendix (Supplementary Table S1 and S2). From this analysis, one observes that aphid Vgs sit within the Vg-like clade, and have most recently diverged from *D. vitifoliae*, which

also do not have a canonical Vg homologue. This suggests that aphids have lost canonical Vg and have retained only Vg-like.

Upon noticing that *D. vitifoliae* had no canonical Vg sequence, I also checked for Vg and Vg-like sequences in scale insects of the closely related Coccoidae family. Only a Vg-like homologue from the European fruit lecanium, *Parthenoclecanium corni*, was found. Therefore, the Phylloxeroidea, Aphidoidea and Coccoidae families appear to have lost canonical Vg, suggesting that it was lost when these families diverged from psyllids.

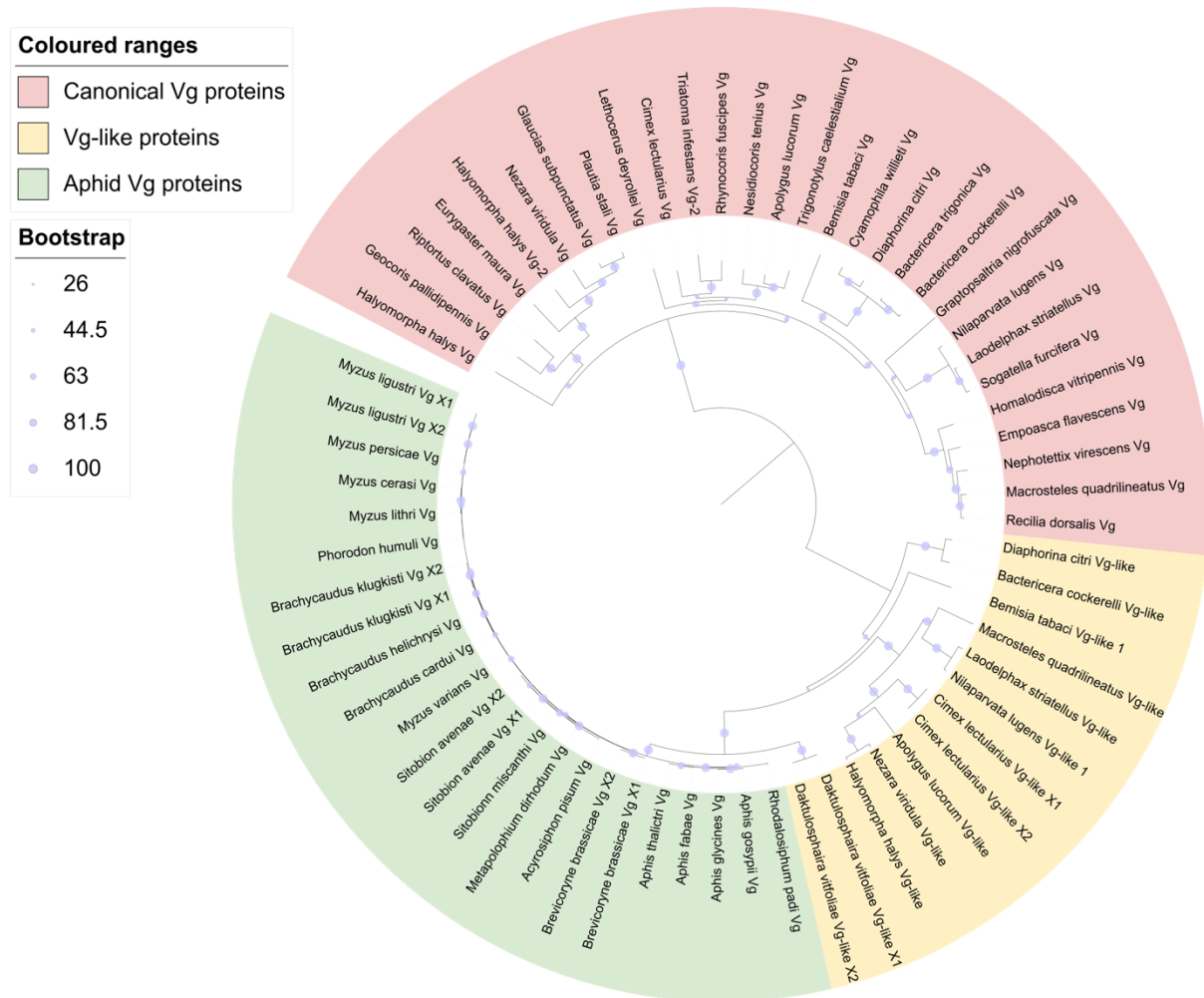


Figure 2.1: A maximum likelihood phylogeny of hemipteran Vg and Vg-like proteins, including aphid Vgs. Canonical Vg sequences are highlighted in red and are a distinct clade from the Vg-like sequences highlighted in yellow. Aphid Vg-like sequences are highlighted in green and sit within the Vg-like clade. 100 bootstraps were included. Bootstrap values are shown by blue circles ranging in size.

2.2.2 InterProScan reveals aphid Vg contains canonical Vg structural domains Vg-N, DUF1943 and VWD

Each aphid Vg sequence used in the phylogenetic tree (Figure 2.1) was subject to InterProScan to identify conserved domains. In most of these sequences, all the Vg-N, DUF1943, and VWD domains were predicted. However, in *M. persicae*, *P. humuli*, *M. varians*, *B. klugkisti*, and *A. thalictri*, the DUF1943 was not identified. To investigate this, an alignment containing the Vg sequences from the 19 aphid genomes was constructed. A region of variation was found in the DUF1943 from position 661-708, but no obvious

differences from sequences that InterPro identified as DUF1943-containing were found (Figure 2.2). In the 5 sequences affected, the DUF1943 was predicted manually by using the alignment to annotate the DUF1943 region of the other aphids. This finding conflicts with the conclusions found by Shang et al., (2018) who suggested that aphids lacked DUF1943, and that this may be what makes them distinct from other Vgs (105). These data suggest aphid Vgs do in fact contain DUF1943, but some do not have it predicted by InterPro, the reason for which remains unclear.

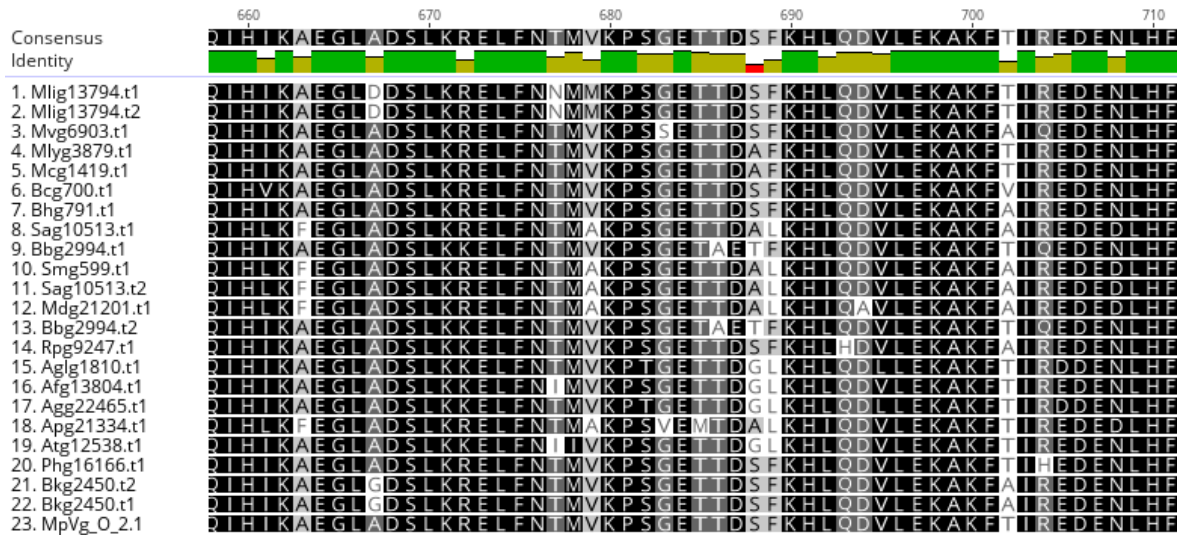


Figure 2.2: A region of DUF1943 in aphid Vg sequences with contains low similarity but does not explain a lack of DUF1943 prediction by InterProScan in 5 species. The Vg sequences of *M. persicae*, *P. humuli*, *M. varians*, *B. klugkisti*, and *A. thalictra* were not predicted to contain DUF1943 by InterProScan, but all others were.

2.2.3 MpVg contains RXXR cleavage site motifs, present in canonical Vgs

Another canonical feature of insect Vgs is cleavage in the fat body, and downstream oligomerisation of the cleavage products. This cleavage occurs at RXXR cleavage sites, most often found in the Vg-N domain (103, 104). The MpVg protein sequence was manually searched for an RXXR cleavage site motif. Three were found: an RHIR motif was

found in the Vg-N domain at positions 494-497; an RFQR motif was found at positions 760-763 in the DUF1943; an RLLR motif was found in the region between the DUF1943 and VWD domains at positions 1039-1042. Whether MpVg is cleaved at any of these sites remains unclear.

2.2.4 Alphafold predictions reveal similarities between the structural domains of Vgs, and reveal that MpVg is more structurally similar to BtVg-like than BtVg

To assess whether the structure was conserved between aphid Vgs and other insect Vgs and Vg-likes, Alphafold2 was used to generate structural predictions (Figure 2.3). UCSF Chimera X was then used to visualise the structures, label domains, and perform structural alignments (Figures 2.3 and 2.4).

When comparing the Alphafold predictions of MpVg, BtVg-like, and BtVg, one observes that domain structures appear to be conserved. However, BtVg (canonical Vg) contains large flexible regions which slit up the β -barrel and α -helical subdomains of the Vg-N domain, which are absent from MpVg and MpVg-like.

Our predictions of MpVg, BtVg-like and BtVg show high confidence in pLDDT scores in the key structural domains Vg-N, DUF1943 and VWD. The models show lower pLDDT scores in the regions linking these domains. This is especially pronounced in the flexible linker regions of BtVg, linking the two halves of Vg-N (Figure 2.3).

Further confidence in the predicted MpVg structure is gained by comparing it to the solved *I. unicuspis* lipovitellin structure, and the modelled *A. mellifera* Vg structure (Figure 2.4) (136, 138). Visual inspection reveals that the domains sit in similar positions in both MpVg and *A. mellifera* Vg. The *I. unicuspis* lipovitellin structure was solved as a dimer of two chains originating from the Vg precursor, hence the VWD is not present, and the coordination of the DUF1943 is different from other models (139). However, the Vg-N domain, associated with receptor binding, sits in a similar position and orientation to the Vg-N domain in both *A. mellifera* Vg and MpVg. These similarities suggest that my MpVg Alphafold prediction is likely to be accurate for the monomeric state of MpVg.

When MpVg is structurally aligned to BtVg-like 1 and BtVg, MpVg is more similar to BtVg-like 1. While the conserved canonical Vg structural domains align well in both cases, there are regions of BtVg which are not conserved in MpVg and therefore do not align at all (Figure 2.5). To further evidence this, foldseek was run on the structure predictions used. The comparison between MpVg and BtVg-like 1 had a query and target TM score 9.452E-01, and 9.438E-01 respectively, while MpVg compared with BtVg had scores of 7.937E-01, and 5.023E-01. The higher scores of MpVg vs BtVg-like 1 show that MpVg is more similar in structure to BtVg-like 1 than to BtVg.

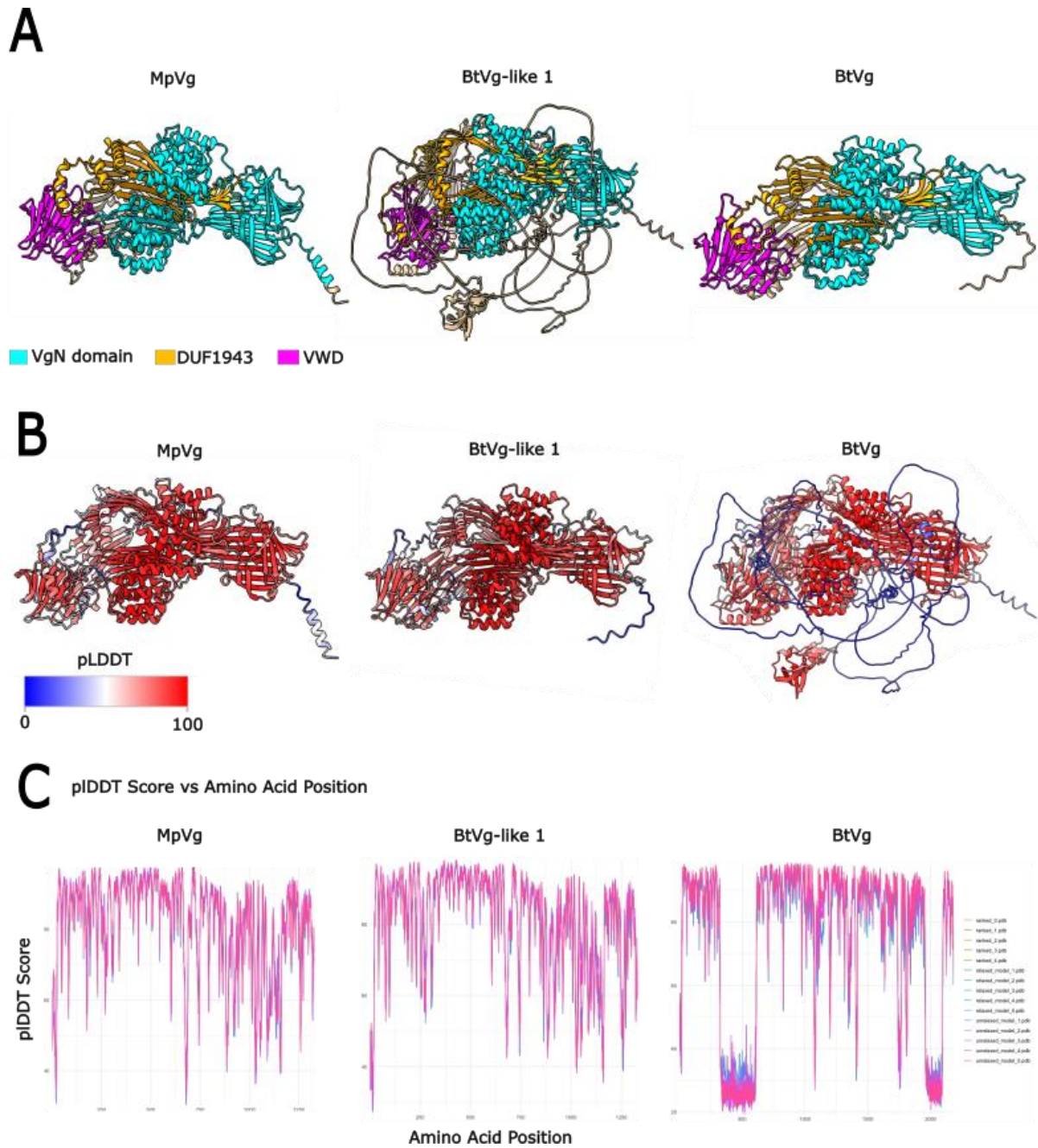


Figure 2.3: Predictions of the protein structures of MpVg, BtVg-like 1 and BtVg using AlphaFold2. (A) The predicted structures of MpVg, BtVg-like 1 and BtVg are presented, showing their structural domains as predicted by InterProScan. The Vg-N domain is coloured in red; the DUF1943 is coloured in magenta; the VWD is coloured in cyan. (B) The predicted structures of MpVg, BtVg-like 1 and BtVg, coloured according to the pLDDT score at each amino acid position. For each of these models 'unrelaxed_model_1' was used. The pLDDT is represented by a blue-red gradient, with blue to red indicating low to high pLDDT scores. (C) Plots of pLDDT score against amino acid position for each predicted model of MpVg, BtVg-like 1 and BtVg-like.

Myzus persicae vitellogenin harbours a region that may be used for REMOT control in aphids.

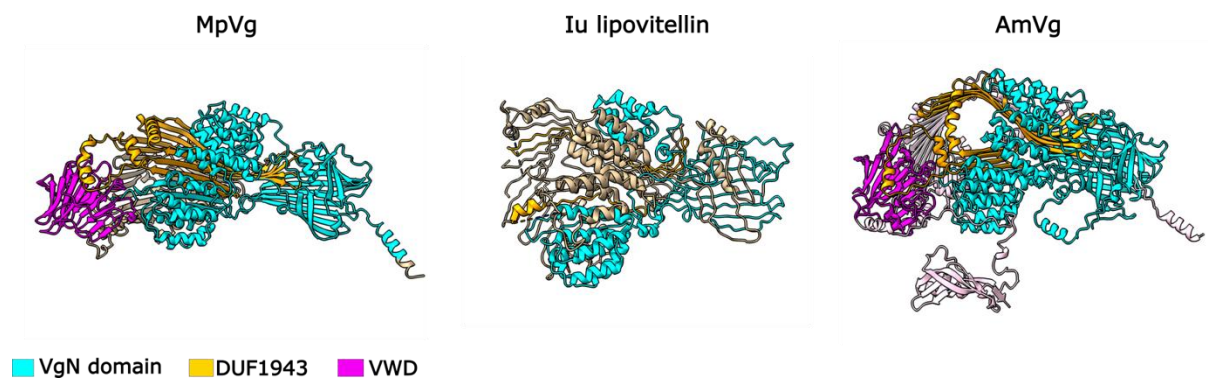


Figure 2.4: The predicted structure of MpVg, the solved crystal structure of *I. unicuspis* lipovitellin (*Iu* lipovitellin), and the modelled structure of *A. mellifera* Vg (*AmVg*) (136, 138). *Iu* lipovitellin is a crystal structure solved as a dimer of two chains originating from the Vg precursor, hence only Vg-N and DUF1943 are mapped here. One observes the similarities between the labelled domains of the three structures. All three contain the Vg-N and DUF1943 domains in similar positions, with *Iu* lipovitellin showing differences likely due to its dimeric state. Both *AmVg* and *MpVg* also show the VWD in similar positions.

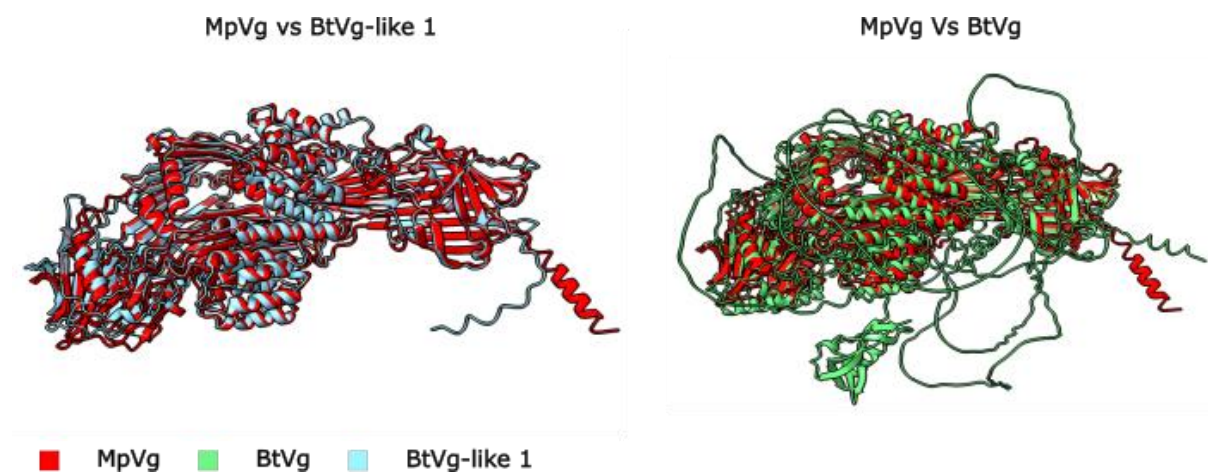


Figure 2.5: Structural alignments of MpVg against BtVg-like 1 and BtVg. MpVg is coloured red; BtVg-like 1 is coloured cyan; BtVg is coloured pink. The entirety of MpVg and BtVg-like 1 align well. MpVg and BtVg align well in well-predicted structural domains, but BtVg has large extra regions that MpVg lacks, causing a more distant alignment.

2.2.5 ‘MpRV’ is an *M. persicae* alternative to ‘BtKV’ which aligns in sequence and structure

To find a potential VgR binding peptide in MpVg, I performed a MSA of MpVg against two known VgR binding peptides, BtVg, *A. aegypti* Vg, *A. gambiae* Vg, *D. noxia* Vg and *A. citricidus* Vg. The two known VgR binding regions were BtKV and the *M. rosenbergii* VgR binding region, upon which BtKV was based (49, 102). This alignment was based on the alignment used to derive BtKV from BtVg (49). While aphid Vgs differ in sequence from the canonical Vgs in this alignment, conserved residues are aligned, allowing the peptide sequence RPSFAAQETGV to stand out as a potential VgR binding peptide of MpVg (Figure 2.6). This peptide was named MpRV.

Because the sequences of BtVg and MpVg are dissimilar, I decided to map MpRV AND BtKV onto their respective Alphafold models to see whether the two peptides structurally aligned (Figure 2.7). This revealed that MpRV and BtKV both reside in the β -barrel of the Vg-N domain of their respective protein models. This lends further evidence to the hypothesis that MpRV is a VgR binding peptide of MpVg, which could enable embryonic uptake of cargo.

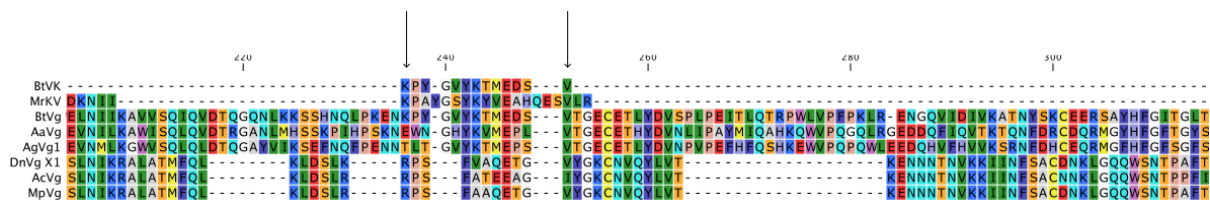


Figure 2.6: A MSA showing how aphid Vg protein sequences align with BtKV, and the VgR binding peptide derived from MrVg (102). The VgR binding peptide of *M. rosenbergii* and *B. tabaci* (BtKV) were aligned to vitellogenin sequences from *B. tabaci*, *A. aegypti*, *A. gambiae*, *D. noxia*, *A. glycines* and *M. persicae*. Arrows represent the proposed VgR binding peptide region. This reveals a potential VgR binding peptide in *M. persicae* named MpRV with the sequence RPSFAAQETGV.

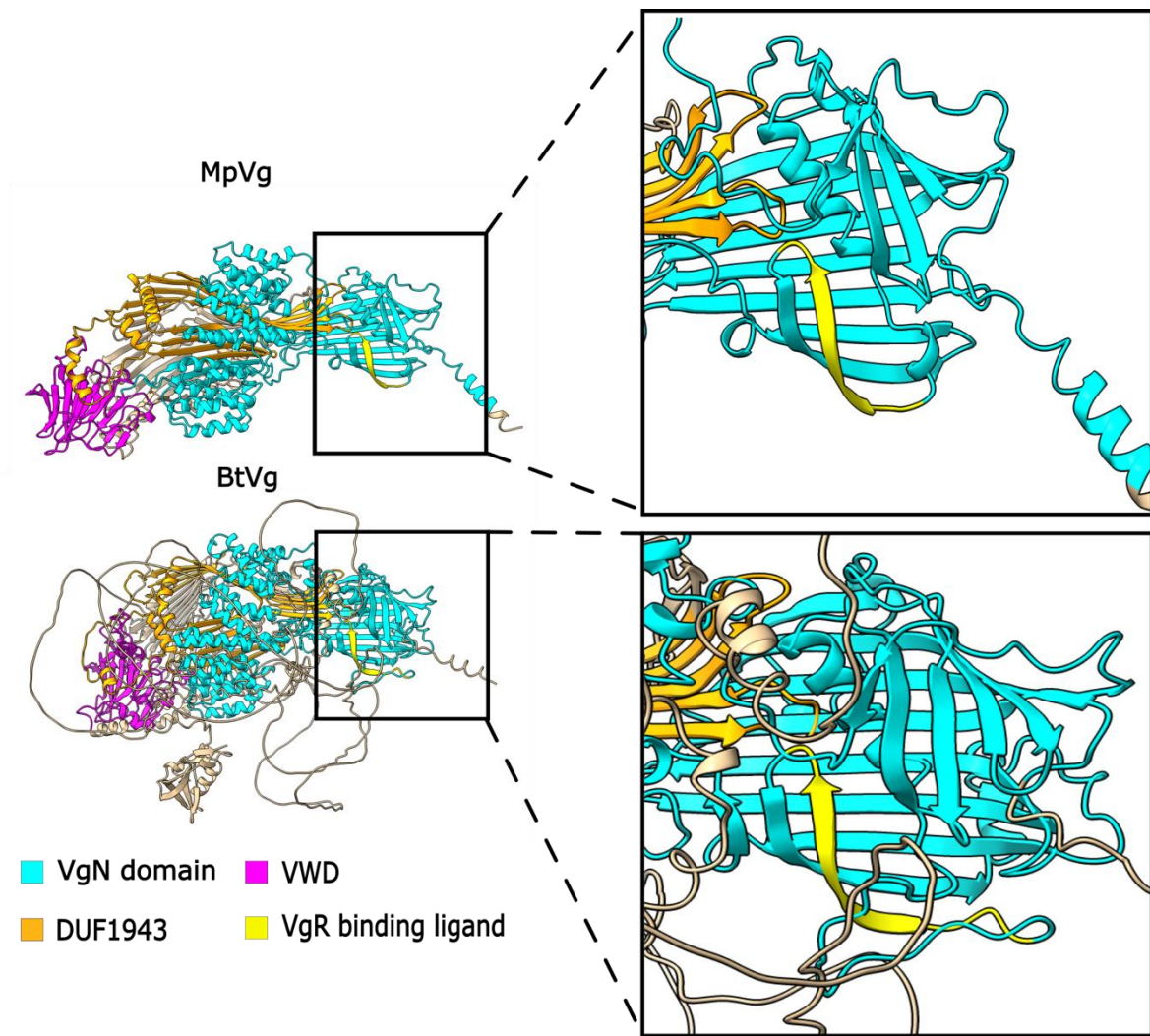


Figure 2.7: The VgR binding peptides of BtVg and MpVg (named BtKV and MpRV respectively) reside in the same structural position. The sequences of BtKV and MpRV were mapped onto BtVg and MpVg structural predictions respectively, coloured here in yellow. This revealed that both peptides reside the same structural location in the Vg-N domain of their respective proteins.

2.2.6 Many vitellogenesis signalling related genes have homologues in *M. persicae*

To gain insight into vitellogenesis in *M. persicae*, BLASTp (NCBI) was used to identify *M. persicae* homologues of Vg-signalling genes. Then, these sequences were used to identify the corresponding gene IDs in the *M. persicae* Clone 'O' v2.0 genome (140) using tBLASTn on 'SapFeederHub'. From the JH pathway, homologues of *Met*, *SRC*, *MCM3*, *MCM4*, *MCM7*, *Cdc6*, *Orc5*, *Cdc2*, *FoxO*, *LCMT1*, *PP2A*, *Hairy*, and *Groucho* were found. From the 20E pathway, homologues of *EcR*, *USP*, *BC*, *E75*, *E78C*, and *HR3* were found. Later, BLASTp (NCBI) was used to search for *kr-h1*, and *TOR* homologues, both of which were also identified. *SRC*, *USP* and *HR3* are located on the X chromosome, while all other homologues are autosomal (Table 2.1).

Table 2.1: A table of homologues of Vg-signalling associated genes, the queries used to find them, the E values and Total scores. Finally, the *M. persicae* 'O' v2.0 gene ID and associated scaffold is shown, except for Kr-h1 and TOR which were only searched for by BLASTp (NCBI) and not subject to further analysis.

Gene	Query	Result	E value	Total score	Mp O v2.0 ID	Scaffold
Met	AAX55681.1	AYI50057.1	1.00E-58	213	g10557	4
SRC	ANG56297.1	XP_022177405.1	4.00E-144	484	g7438	1 (X)
Kr-h1	QAA13014.1	XP_022178368.1	8.00E-147	443	N/A	N/A
MCM3	AIP98399.1	XP_022173910.1	0.00E+00	1067	g15158	5
MCM4	AHA42533.1	XP_022176756.1	0.00E+00	1055	g25220	2
MCM7	AHA42534.1	XP_022160872.1	0.00E+00	1030	g10962	4
Cdc6	ALO23489.1	XP_022180861.1	8.00E-99	304	g26967	2
Orc5	QKG02512.1	XP_022180676.1	2.00E-41	150	g20732	6
Cdc2	QKG02511.1	XP_022166469.1	1.00E-140	398	g14199	5
FoxO	QJX15634.1	XP_022179991.1	1.00E-103	317	g24925	2
LCMT1	QJX15635.1	XP_022163684.1	1.00E-122	355	g10557	2
PP2A	XP_021700500.1	XP_022182715.1	0.00E+00	619	g24583	5
Hairy	XP_001662100.1	XP_022162089.1	1.00E-65	208	g15123	6
Groucho	XP_021709493.1	XP_022168813.1	0.00E+00	1075	g21637	2
EcR	KAL2713188.1	ABN11289.1	2.00E-32	120	g27247	2
USP	XP_011493175.2	XP_022172713.1	7.00E-127	374	g24080	1 (X)
BC	AAS80329.1	XP_022164841.1	4.00E-70	296	g6795	2
E75	XP_001652743.3	XP_022162357.1	0.00E+00	661	g18236	3
E78C	JAC46196.1	XP_022161333.1	4.00E-95	519	g19248	3
H23	XP_021702101.1	XP_022180447.1	2.00E-77	487	g5901	1 (X)
TOR	AAR97336.1	XP_022175703.1	0.00E+00	2853	N/A	N/A

2.2.7 Morph-specific RNAseq analysis reveals MpVg is most expressed in males

Using previously generated RNAseq data from asexual winged, and wingless females, males and nymphs (141), the expression patterns of MpVg, MpVgR and the identified Vg-signalling associated homologues were analysed (Figure 2.8 and 2.9). Surprisingly, this revealed that MpVg is most expressed in males. Further, VgR is expressed at extremely low levels throughout (Figure 2.8, 2.9, 2.10, 2.11). Out of the JH pathway associated genes, *Met*, *FoxO*, and *PP2A* all share similar expression patterns to MpVg (Figure 2.9). However, polyploidy associated genes *Cdc2*, *MCM3*, *4* and *7* appear to be expressed in an inverse pattern, with their expression being least in males (Figure 2.9). Out of the 20E associated genes, *EcR*, *USP* and *E75* appear to be expressed in a similar pattern to MpVg (Figure 2.11). Summary statistics for these analyses are available in the appendix.

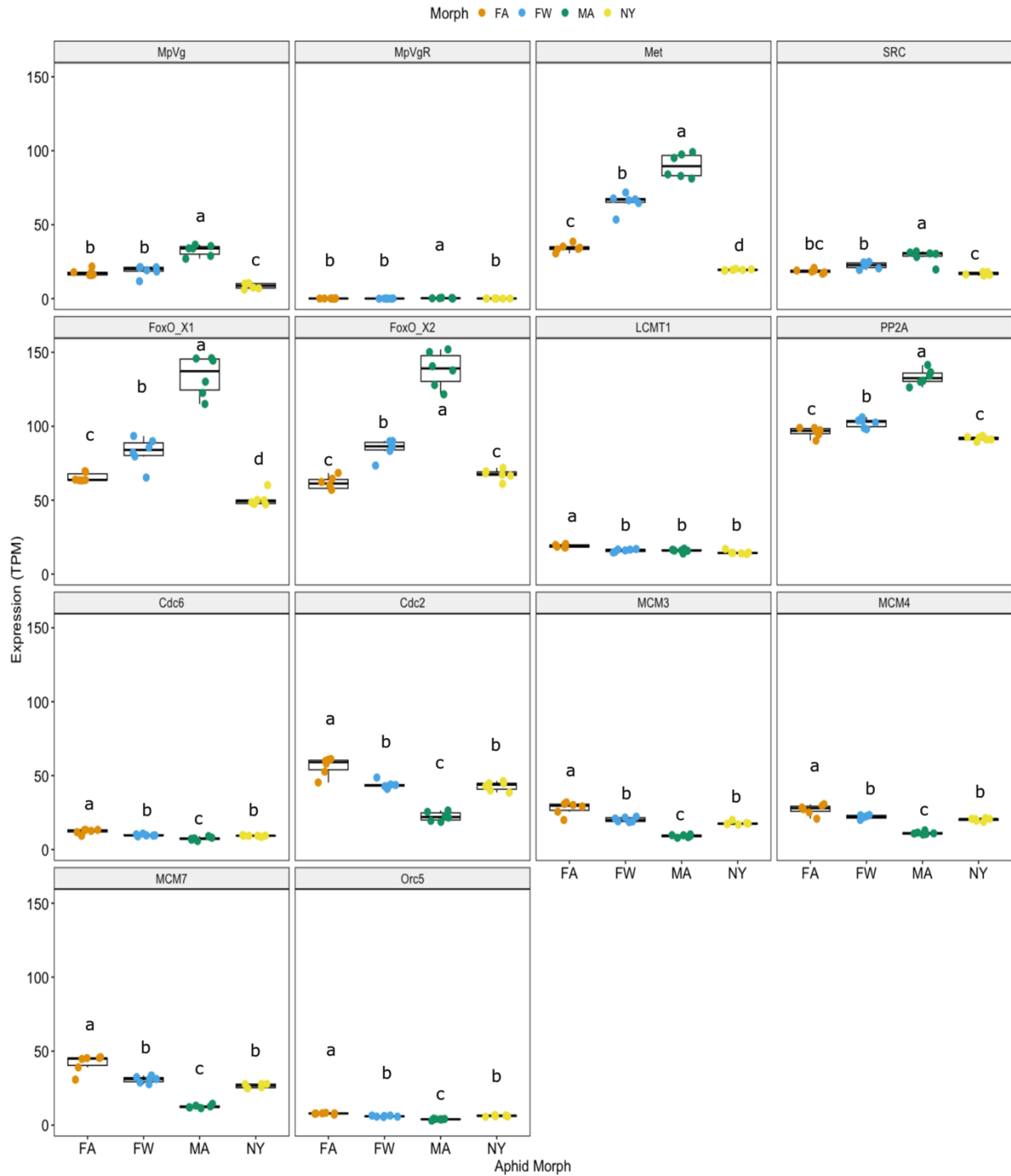


Figure 2.8: Expression patterns of MpVg, MpVgR and JH-pathway associated genes (Met, SRC, FoxO [isoforms X1 and X2], LCMT1, PP2A, Cdc6, Cdc2, MCM3, 4 and 7, and Orc5). For each morph, RNAseq data for 6 individuals was used. 'FA' is asexual wingless females (orange); 'FW' is asexual winged females (blue); 'MA' is males (green); 'NY' is nymphs (yellow). Significance was calculated via ANOVA with post-hoc analysis using Tukey test.

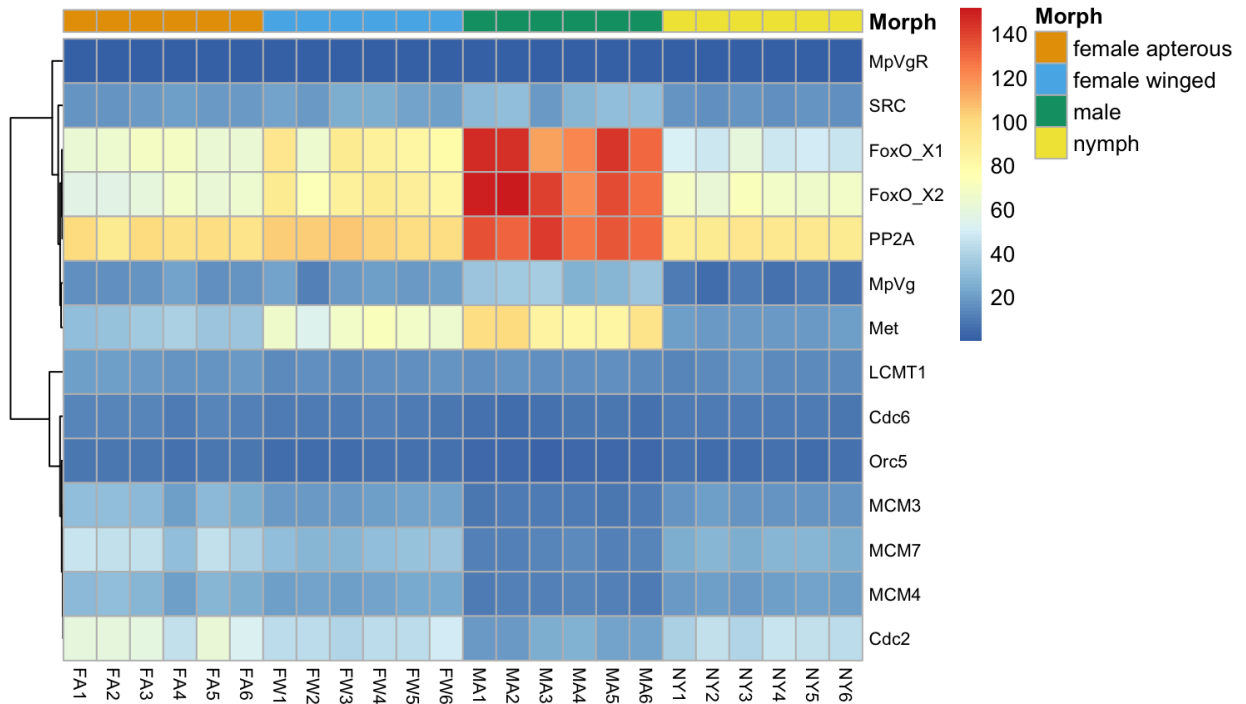


Figure 2.9: A heatmap with clustering to show expression levels of each JH pathway associated gene across all individuals of all 4 morphs. Clustering of the tree (left) shows similarity in expression patterns. Blue indicates a low expression level, while red indicates a high expression level.

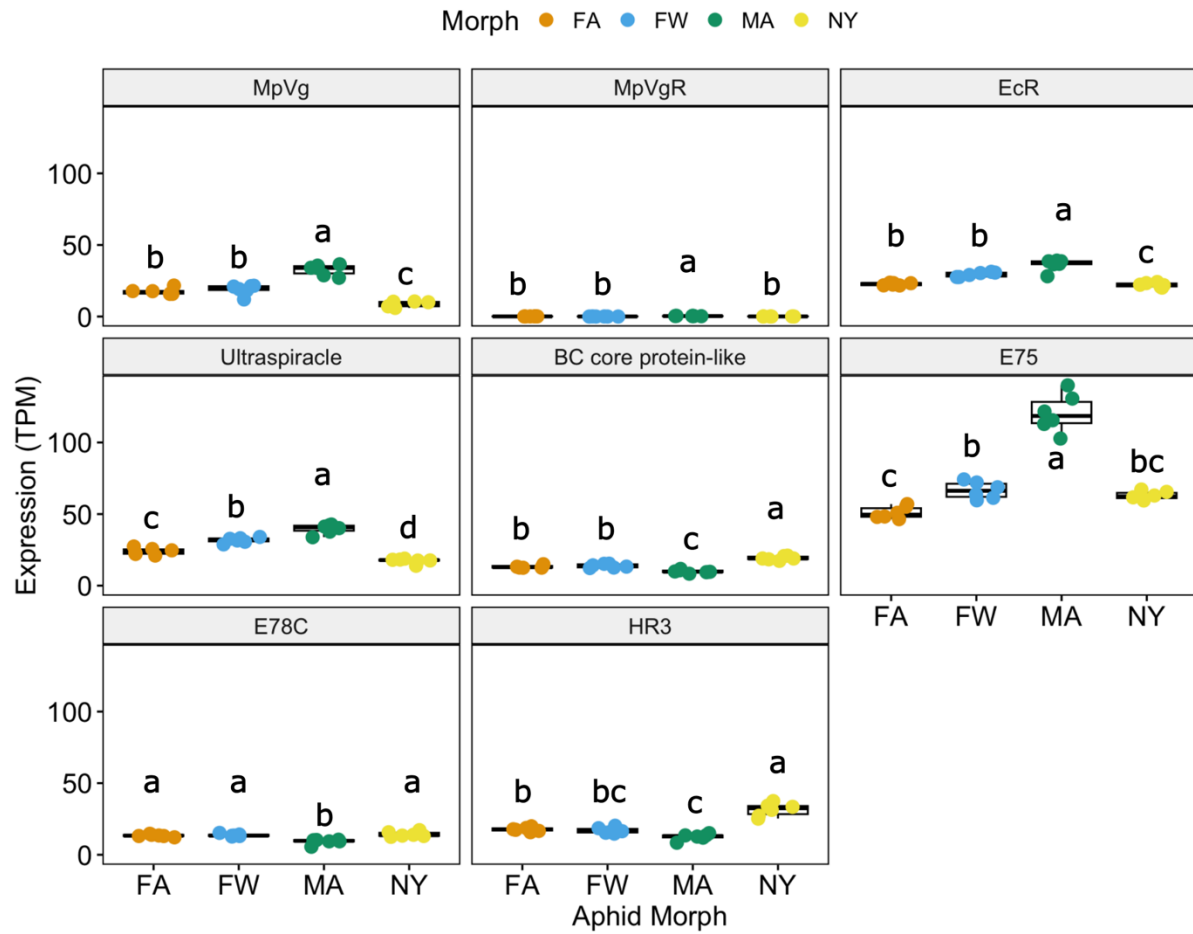


Figure 2.10: Expression patterns of MpVg, MpVgR and ECR-pathway associated genes (EcR, Ultraspiracle, BC core protein-like, E75, E78C, and HR3). For each morph, RNAseq data for 6 individuals was used. 'FA' is asexual wingless females (orange); 'FW' is asexual winged females (blue); 'MA' is males (green); 'NY' is nymphs (yellow). Significance was calculated via ANOVA with post-hoc analysis using Tukey test.

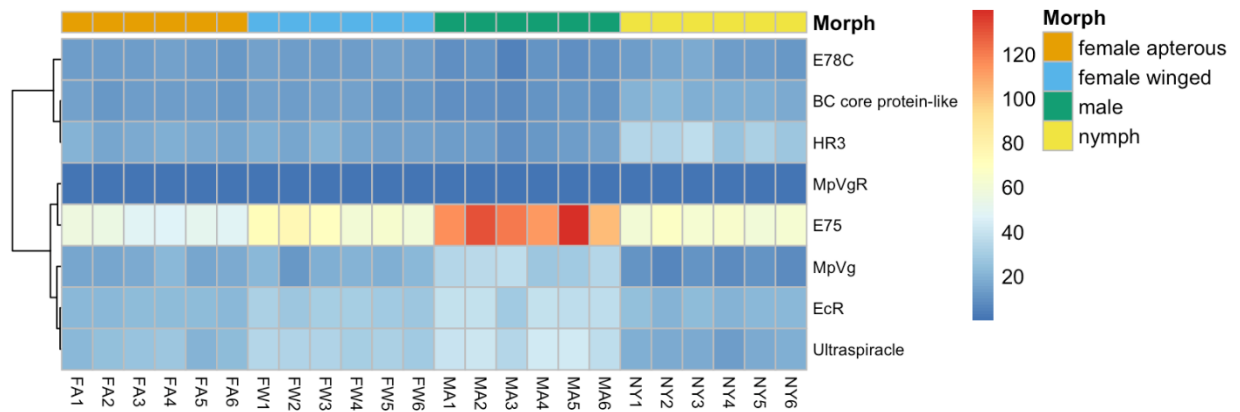


Figure 2.11: A heatmap with clustering to show expression levels of each 20E pathway associated gene across all individuals of all 4 morphs. Clustering of the tree (left) shows similarity in expression patterns. Blue indicates a low expression level, while red indicates a high expression level.

2.3 Discussion

2.3.1 Aphid Vgs are Vg-like orthologues, which likely carry canonical Vg functions

To derive a VgR binding peptide from MpVg, I first wanted to characterise MpVg, to decide whether this protein was likely to be similar in function to BtVg, from which a successful BtVgR binding peptide was derived (BtKV) (49).

Canonical Vgs contain 3 conserved structural domains and a signal peptide (103, 104). These are the Vg-N domain, the DUF1943, and the VWD. The Vg-N domain is involved in VgR binding (102, 103), while the DUF1943 and VWD are both involved in coordinating the lipid binding cavity. Further, the DUF1943 and VWD have been shown to have roles in innate immunity (129-131, 142). These domains are conserved in canonical Vgs and some Vg-like proteins (128).

Whilst there are many studies on insect Vgs, there is only one study to my knowledge which attempts to characterise an aphid Vg. In this study on *A. citricidus* Vg (AcVg), the expression characteristics of AcVg suggest that it carries the expected reproductive function; AcVg is highly expressed in the fat body of adult females, while *A. citricidus* VgR (AcVgR) is highly expressed in ovaries (105). However, in a phylogeny in the same study aphid Vgs do not group with other hemipteran Vgs and have apparently lost the DUF1943 (105). This appears to suggest that the lack of the DUF1943 is likely the cause of aphid Vgs grouping away from other hemipteran Vgs (105). I subjected Vg sequences from 19

aphids to functional domain prediction via InterProScan (143). Contrary to the study on AcVg (105) most of these sequences were predicted to have all three of the canonical Vg domains: Vg-N; DUF1943; VWD. However, in *M. persicae*, *P. humuli*, *M. varians*, *B. klugkisti*, and *A. thalictri*, the DUF1943 was not identified. Upon investigation via MSA, I could see no distinct amino acid changes that would indicate a reason for the lack of a DUF1943 prediction in these sequences (Figure 2.2). Therefore, in MpVg, the DUF1943 was predicted annotating DUF1943 on the MSA, and annotating the conserved region in MpVg. The finding that aphid Vgs contain all 3 canonical Vg domains suggest that they likely carry out similar functions. Further, the revelation that aphid Vgs do indeed contain DUF1943 reopens the possibility that aphid Vgs could have functions in innate immunity. Further experiments, such as expression of the individual Vg domains followed by coimmunoprecipitation after incubation with potential ligands, could help to further understand the role of aphid Vgs.

In asexual reproducing aphids, which are viviparous, the need for Vg-associated reproductive and developmental functions is likely decreased, or even abolished, as the embryos develop inside the mother in filamentous ovaries comprised of ovarioles containing multiple embryos at different developmental stages (28). Stage 1-4 represent oocytes that are pre-blastoderm-formation and are connected to the germaria (stage 0) by a trophic chord. It has been shown that the trophic cord in sexual ovaries is much larger than in asexual ovaries (29), perhaps to accommodate the greater need for Vg uptake and transport to sexual oocytes, destined to become overwintering eggs. It has been shown that aphids express VgR, which is conserved from other hemipteran species (105). Therefore, it is possible that in asexual females, the first 4 stages of embryos are

fed nutrition via Vg uptake by VgR and transport along the trophic chord. In sexual females, it is possible that the egg is fed from the larger trophic cord. If all nutritive yolk is delivered by transport along the trophic cord, it is possible that Vg is only taken up in germaria, and oocytes receive it solely from the trophic cord. For the purposes of gene editing by ReMOT control, the Cas9 cargo needs to be taken up by pre-blastoderm oocytes to enable germline gene editing (49, 89, 90, 94, 96, 98, 101, 144). Therefore, if MpVg is taken up in the germaria of asexual ovarioles and transported to pre-blastoderm oocytes via the trophic cord, then a peptide derived from MpVg that binds to MpVgR should allow for transduction of cargo into the appropriate stage of *M. persicae* oocytes.

The finding that aphids do indeed contain DUF1943 reopens the question of why aphid Vgs appear to group separately from other hemipteran Vgs (105). Another study on Vg and Vg-like genes in *N. lugens*, aphid Vgs are grouped with Vg-like genes of other species (128). Therefore, I hypothesised that aphids may have lost ‘canonical’ Vg and instead retained a Vg-like gene.

The Vg-associated pathways signalled by JH and 20E were analysed. The *M. persicae* genome was searched for homologues of genes related to vitellogenesis signalling. Homologues of the genes involved in JH, 20E and AA/TOR signalling were identified, showing that *M. persicae* has the capacity for Vg regulation through these pathways. Morph specific RNAseq data analysis on these homologues, MpVg and MpVgR, revealed unexpected expression patterns. MpVg is most expressed in males. This may indicate that MpVg has adopted an alternative role in males which remains unclear. Some regulatory genes share the same morph specific pattern as MpVg. *Met*, *SRC*, *FoxO* and

PP2A all share expression patterns with MpVg, suggesting regulation may occur through JH signalling causing FoxO dephosphorylation, as in *L. migratoria* (107). 20E-pathway related genes *EcR*, *USP* and *E75* also share the same expression pattern as MpVg, suggesting their involvement in regulation also. However, with the hypothesis that *M. persicae* has lost canonical Vg, it is possible that MpVg is signalled in a different way. Further study on the signalling pathways, and the promotor regions of the genes involved, is necessary to further characterise these pathways. The revelation that MpVg is most expressed in males in this dataset was surprising. However, a dataset including sexual females would reveal important information. It is possible that MpVg is upregulated in all sexual morphs, including males and females, due to signalling pathways involved in morph switching in response to photoperiod. This would support the hypothesis that MpVg carries out canonical Vg functions in sexually reproducing aphids. However, the findings here may also point to a non-canonical role of MpVg in males.

2.3.2 Aphidomorpha and Coccoidae lost canonical Vg when they diverged from psyllids

I constructed a phylogeny of hemipteran Vg and Vg-like protein sequences (Figure 2.1). This showed that all transcripts from 19 aphid species grouped with Vg-like proteins of other species, and not canonical Vgs. It appears that aphid Vgs have diverged most recently from phylloxera Vg-like. This provides evidence that the reason for aphid Vgs grouping away from hemipteran Vgs is due to the loss of canonical Vg, and not a lack of DUF1943.

I have found that aphid Vgs are Vg-like orthologues, rather than canonical Vg orthologues. The reproductive and developmental functions of Vg, Vg-like1 and Vg-like2 of the brown planthopper, *N. lugens*, have been studied (128). RNAi knockdown showed that *N. lugens* Vg (NlVg) caused fatty body build-ups in adult females, leading to a swollen abdomen, as well as inhibited oocyte growth, but knockdown of NlVg-like1 and 2 did not show these phenotypes, suggesting that canonical NlVg alone is important for oogenesis and oocyte development (128). RNAi knockdown against the same genes in nymphs revealed a lethal phenotype when NlVg is knocked down, causing nymphs to exhibit a much thinner body, while NlVg-like1 and 2 knockdown caused no significant effect on mortality, suggesting that canonical NlVg is also involved in nymph development (128). NlVg-like 1 and 2 were shown to play a role in egg hatching. RNAi knockdown of NlVg-like1 in adult females resulted in 18% of eggs failing to hatch or the offspring being dead before hatching. Knockdown of NlVg-like2 65% of eggs failing to hatch. This indicates NlVg-like1 and 2 have roles in embryo maturation and nutrition absorption by oocytes and embryos (128). It is possible that aphid Vgs have adopted some, or all of these roles. If this is the case, it is likely to be most important in the sexual reproductive cycle of aphids, as this involves oviposition of overwintering eggs. Indeed, eggs of multiple aphid species have been described as being filled with yolk upon deposition (28). In most insects, yolk is primarily made up of Vn (103). Therefore, it is possible that aphid Vg performs roles in embryogenesis and nutrition supply for these eggs.

The grape phylloxera, *D. vitifoliae*, like aphids, undergoes asexual and sexual reproductive cycles (145). I found that like aphids, they appear not to have a canonical Vg. Phylloxeroidae are a family in the same infraorder as Aphidoidea (Aphidomorpha) which diverged from a common ancestor shared with Coccidae, which diverged from psyllids (146). I found that it appears that, like aphids and phylloxera, the European fruit lecanium, *Parthenoclecanium corni* (Coccidae), has also lost canonical Vg. Coccidae can reproduce sexually or parthenogenetically, with some species undergoing both modes (147). Some species are also hermaphroditic (148). However, some species lay eggs, while others undergo viviparity (149). Most aphids, phylloxera and Coccidae lay eggs during their lifecycle, suggesting that yolk is required. Unlike aphids, phylloxera are oviparous throughout both their asexual and sexual reproductive cycles (145). Coccidae exhibit oviparity and viviparity like aphids, but some species are obligately oviparous (149). The finding that these families have lost canonical Vg suggests that either phylloxera, aphids and Coccidae have an alternative yolk protein, or that Vg-like has adopted the nutrition-providing functions of canonical Vg. Further study is required to test this hypothesis. My findings suggest that canonical Vg was lost after psyllids diverged.

2.3.3 The peptide MpRV is a promising candidate peptide for ReMOT control in aphids

MpVg was further characterised, and compared to BtVg-like and BtVg, using structural predictions done using Alphafold2 (Figure 2.3) (150). ChimeraX was used to label predicted domains on these predictions. These models confidently predict that the canonical domains Vg-N, DUF1943, and VWD are structurally conserved across all three proteins. This further suggests that MpVg does indeed contain DUF1943, despite its lack of prediction by InterPro. These structural predictions also provide further evidence that MpVg is in fact a homologue of BtVg-like, rather than canonical BtVg. Canonical BtVg contains long, poorly predicted linker regions between the β -barrel and α -helical domains of the Vg-N domain, which are not conserved in BtVg-like and MpVg. Further, structural alignment and foldseek analysis suggest a higher similarity between MpVg and BtVg-like (Figure 2.5).

To my knowledge, no full-length crystal or CryoEM structures of Vg exist. However, there is a crystal structure of its lipid binding product – lipovitellin – *I. unicuspis* (Uniprot: Q91062) (138). In the lamprey, lipovitellin is derived after cleavage of the Vg precursor, and binding of multiple lipids in the lipid-binding cavity (138). Cleavage is a common occurrence in Vgs, occurring in most insects at an RXXR motif, most commonly found in the Vg-N domain (103, 104). In other insects, such as the honey bee, *Apis mellifera*, the cleavage site is found in the polyserine linker between the two subdomains of Vg-N (136). The structure of *A. mellifera* Vg has been well characterised by homology modelling, and Alphafold prediction followed by mapping to an EM map, however, this is a model of the

uncleaved Vg (136). The oligomerisation state of Vgs appears to be varied. The lamprey structure appeared as a dimer of two chains both originating from the Vg precursor, while the honey bee structure was shown to be monomeric, as the EM map only allowed for fitting of a single monomer of full length Vg (136, 138). BN-PAGE was used to check for other oligomerisation states of *A. mellifera* Vg, and it was found that a small portion of the product matched to 345 kDa dimer of Vg (136). The oligomerisation state of MpVg remains unclear, but I identified three RXXR cleavage site motifs, suggesting that the protein can be cleaved.

I also compared my Alphafold model of MpVg to the structures of lamprey lipovitellin and *A. mellifera* Vg. I observe that the domain structures present are similar in all three structures, although, as the lamprey lipovitellin is solved as a dimer, the VWD domain is not present (Figure 2.4). This provides evidence, along with good pLDDT scores, that the predicted MpVg structure is accurate.

ReMOT control in *B. tabaci* was enabled by the Vg derived peptide 'BtKV', which was found by MSA against other insect Vgs, and the VgR binding peptide of *M. rosenbergii* previously determined by peptide array (49, 102). The findings presented in this chapter that MpVg shares structural domains and features of canonical BtVg, as well as the previous finding that *M. persicae* express a conserved VgR (105), suggest that a peptide derived from MpVg may enable ReMOT control in asexually reproducing *M. persicae*. However, the finding that MpVg is in fact a Vg-like homologue may suggest otherwise. Therefore, I investigated whether there is a suitable peptide in MpVg that would enable ReMOT control.

A MSA based on the one used to derive BtKV (49) revealed that ‘MpRV’ (RPSFAAQETGV), at amino acid position 135-145 of MpVg, aligns with BtKV and the VgR binding region of *M. rosenbergii* (Figure 2.6). Whilst there are conserved 136-proline, 142-glutamate and 145-valine residues, with more conserved residues in flanking regions, this sequence is quite dissimilar to both BtKV and the VgR binding region of *M. rosenbergii*. I therefore decided to map MpRV and BtKV to my MpVg and BtVg AlphaFold predictions respectively, to see whether the peptides reside in the same position of the Vg structures (Figure 2.7). Indeed, both MpRV and BtKV reside in the same region of the Vg-N domain, which has been associated with VgR binding (102-104). This provided confidence that MpRV is a promising candidate peptide for enabling ReMOT control in *M. persicae*. Other findings presented and discussed in this chapter shed light on the possibility that MpVg is taken up in the germaria of *M. persicae* ovarioles and delivered to stage 1-4 pre-blastoderm via a trophic cord. However, whether MpRV will enable uptake of cargo in this way remains unclear. In previous ReMOT control studies, the efficacy of the receptor binding peptides used have been shown by tagging fluorescent proteins with these peptides, injecting the chimeric proteins into adult mothers, and using fluorescence microscopy to determine uptake into oocytes (49, 89, 90, 94, 96, 98, 101, 144). Therefore, experiments of this nature testing whether MpRV can enable ovary transduction of cargo are presented and discussed in chapter 3.

2.4 Materials and Methods

2.4.1 Genome sequences used in the study

All aphid genomes used in this study were downloaded internally via the NBI HPC (140, 151-154).

2.4.2 Creation of aphid protein database compiling protein sequences from 18 aphid genomes

To identify Vg sequences in previously assembled aphid genomes (140, 151-154), I aimed to create a file compiling protein sequences annotated in these genomes to serve as a BLAST database. To achieve this, I located the FASTA files containing annotated protein sequences for each of the 18 aphid genomes. I then renamed the entries in each protein sequence file to include species-specific prefixes in the gene IDs (e.g., *D. noxia* gene IDs were modified to begin with 'Dn'). After renaming the entries, I concatenated the sequences from all species into a single file containing the protein sequences from all 18 species. This process was performed using BASH on the high-performance computing cluster. The protein annotations for *A. pisum* were included later, so its proteome was searched separately from the combined database.

2.4.3 BLAST searches for Vg homologues

To identify Vg sequences in the *M. persicae* genome, I used command line BLAST to search the *M. persicae* braker 2.1 annotation (154) in the John Innes Centre high performance computing cluster environment. BLAST+-2.2.30 was used with the default filters. Sequences with a e-value of 0.0, and a minimum blast score of 2000 were selected for further analysis. For later analysis, *M. persicae* clone 'O' v2.0 (140) gene IDs for MpVg were found by using BLASTn with sequences identified in the v2.1 annotation (154). The general code used for command line BLAST is available in the appendix (General code for command line BLAST).

This BLAST tool was also used to find homologues of Vg in other aphids from the compiled database of proteins from 18 aphid genome annotations. The *A. pisum* proteome was also searched in the same way. The *M. persicae* Vg sequence (MYZPE13164_O_Elv2.1_0213490.1) was used as a query. Sequences with a e-value of 0.0, and a minimum blast score of 2000 were selected for further analysis and used to generate amino acid sequence alignments and phylogenetic trees. The code used to perform the blast against the 18 genomes is available in the appendix (BLAST for Vg sequences in 18 aphid genomes). The sequences found and their parent genome annotation information are found in the appendix (Supplementary table S2).

Coccoidae were also searched for Vg and Vg-like homologues. NCBI BLASTp was used to search for Vg and Vg-like homologues using MpVg and BtVg as queries for Vg-like and canonical Vg respectively.

2.4.4 Construction of a Vg, Vg-like and aphid Vg phylogeny

To assess the relationships of canonical Vgs, Vg-likes and aphid Vgs, I constructed a phylogenetic tree of hemipteran Vg and Vg-like sequences Vg. Beyond the aphid Vg sequences identified as per above, I added other hemipteran Vg and Vg-like sequences identified via NCBI BLAST, using BtVg (ADU04392.1) and BtVg-like (BtabMEAM1_Bta14071/1-1330 (40)) respectively as query sequences. All sequences used in the phylogenetic analysis are listed in Supplementary Tables S1 and S2.

To generate the phylogeny, I aligned the Vg sequences using MUSCLE (155) in Geneious Prime. Incomplete sequences were deleted, and large gaps were removed from the alignment. The final tree was constructed using the Geneious Prime RAxML plugin (156), with the protein model set as 'GAMMA BLOSUM62', the algorithm set as 'Rapid Bootstrapping and search for best-scoring ML tree', the number of bootstraps set as 100, and the parsimony random seed set as 2. The resulting tree was visualized in iTol (157).

2.4.5 Vg structural domain prediction with InterPro

The structural domains of Vg proteins were predicted using InterProScan (143). These domains were labelled on Alphafold predictions using UCFC Chimera X. The protein sequences were also annotated with these domains in CLC Main Workbench (Qiagen).

2.4.6 Construction of an aphid Vg alignment

To investigate differences between the protein sequences of aphid Vg proteins, an alignment containing all the aphid Vg sequences used in the phylogenetic tree was produced. This was done using MUSCLE in Geneious prime, with the default settings.

2.4.7 Identification of an RXXR cleavage site motif in MpVg

The sequence of MpVg was manually searched for an RXXR motif in CLC Main Workbench (Qiagen).

2.4.8 Vg structure predictions in alphafold2

The structures for MpVg, BtVg, and BtVg-like were predicted using Alphafold2 (150) on a HPC. The script used for this work is included in the appendix (Alphafold script). The pLDDT of these predictions was visualised in ChimeraX by colouring the structure by the B-factor. This was also plotted as a graph in R. The R script used is included in appendix (R script for generation of pLDDT plots) and uses the packages ggplot2, dplyr (158) and svDialogs (159).

2.4.9 Comparison of MpVg, with silver lamprey lipovitellin and honeybee Vg structures

To check the accuracy of my Alphafold prediction of MpVg, I visually compared it with the structures of silver lamprey, *Ichthyomyzon unicuspis*, lipovitellin solved by X-ray crystallography (Uniprot: Q91062), and *Apis mellifera* Vg, predicted by homology modelling, Alphafold, and fitting to an EM map (Uniprot: Q868N5) (136, 138). The

structure associated sequences for *I. unicuspis* Vg (GenBank: AAA49327.1), and *A. mellifera* Vg (GenBank: CAD56944.1) were subject to InterProScan to predict structural domains. Where possible, these domains were labelled on the structures using ChimeraX.

2.4.10 Foldseek analysis of MpVg vs BtVg-like and BtVg

The alphafold2 ‘unrelaxed_model_1’ predictions of MpVg, BtVg-like 1 and BtVg were subject to foldseek (160) analysis against one another using the command line on the HPC at JIC using default arguments.

2.4.11 MSA to find *M. persicae* alternative to ‘BtKV’

To find a potential VgR binding peptide in MpVg, a multiple sequence alignment was constructed comparing MpVg to known VgR binding sequences from *Macrobrachium rosenbergii* and *B. tabaci*, as well as some other insect Vgs. The alignment was constructed using CLC Main Workbench 21 (Qiagen Digital Insights). The settings used were ‘gap open cost’ of 10, ‘gap extension cost’ of 1, ‘end gap cost’ set to ‘as any other’ and the ‘very accurate’ alignment setting. MpVg (MYZPE13164_O_Elv2.1_0213490.1) was aligned to the vitellogenin of two other aphid species (*D. noxia*, XP_15366382.1 and *A. citricidus*, AVP41182.1), one sequence from *B. tabaci* ADU04392.1), one from *A. gambiae* (AAF82131.1), and one from *A. aegypti* (AAA99486.1). Also aligned was ‘BtKV’, and the VgR targeting peptide from *M. rosenbergii* identified previously (102).

2.4.12 Identification of Vg-signalling associated gene homologues in *M. persicae*

To identify *M. persicae* homologues of genes associated with the JH and 20E signalling pathways, BLASTp (NCBI) was used. Query sequences were selected from *L. migratoria* or *A. aegypti* where possible as the pathways are well characterised in these insects. For each query, the most similar result was taken forward. These sequences were then inputted into tBLASTn on 'Sap Feeder Hub' to find the associated genes in the *M. persicae* clone 'O' v2.0 annotation (140). Scaffold location was found and labelled according to the scaffold labelling in the assembly (140).

2.4.13 Morph specific RNAseq analysis of MpVg, MpVgR, and Vg-signalling associated genes

RNAseq data for asexual winged and wingless females, males, and nymphs was readily available (141), and had been mapped to the *M. persicae* clone 'O' v2.0 annotation (140). The data corresponding to the gene IDs of MpVg, MpVgR and the Vg-signalling associated genes identified was analysed in R studio. Significance in expression level for each gene between different morphs was calculated using ANOVA followed by post-hoc adjustment using Tukey test. To analyse clustering of genes with similar expression patterns, clustered heatmaps were generated in R using 'pheatmap' (161). The scripts used are available in the appendix (Scripts for Morph Specific RNAseq analysis in R).

3 Generation and validation of embryo-localising fluorescent protein chimeras

3.1 Introduction

Gene editing by ReMOT control relies on the use of chimeric embryo-localising proteins. The embryo-localising peptides in these chimeric proteins are derived from YPPs such as *Drosophila melanogaster* yolk precursor protein 1 (DmYP1) or vitellogenin (Vg) (49, 89, 90, 94-98, 100, 101). By far the most often used is P2C, derived from DmYP1. The P2C receptor binding region was determined by deletion assay (90). Different fragments of purified recombinant GFP-fusions of DmYP1 (P1, P2 and P3) were injected into adult female mosquitos 24 hours after a blood meal. After this, the ovaries were removed, and embryos were checked for fluorescence. The most fluorescent came from P2, which was further deconstructed into P2A, P2B and P2C. These parts were subject to the same fluorescence experiment, with P2C being the most effective (90). P2C, fused to CAS9, has been effective in ReMOT control studies on a variety of insects including multiple mosquitos, *N. vitripennis*, and *T. castaneum*, and even one arachnid – the black legged tic *I. scapularis* (89, 90, 94-97, 101). However, in *B. tabaci* P2C was unsuccessful, so the alternative ‘BtKV’ was found, derived from Vg using MSA against a known VgR binding region of *M. rosenbergii* Vg (49, 102). This peptide’s embryo-localising ability was verified in a similar way to P2C; a chimeric protein BtKV-mCherry was purified and injected into *B. tabaci* mothers, after which ovaries were excised and examined for fluorescence (49). BtKV has successfully enabled ReMOT control in two hemipteran species – *B. tabaci* and

R. prolixus (49, 98). In *B. mori* a different Vg derived peptide was used called BmOTP (100). These cases demonstrate that P2C is not a one-size fits all peptide, and sometimes a taxa-specific peptide is required.

In Chapter 2, I presented the identification of MpRV, derived from MpVg, as a possible aphid specific embryo-localising peptide. While this peptide looks promising in terms of sequence and structural properties, further experimentation is required to validate its ability to direct localisation of proteins to the embryo. This can be done in the line with the previously used methodology: production of an MpRV-fluorophore chimera, injection into aphid mothers, and inspection of ovaries for fluorescence. It is also possible that P2C could enable ReMOT control in aphids, and the same experiment can be used to check this.

Therefore, the aims of this chapter are to: identify a suitable fluorophore for use in *M. persicae* embryos by analysing aphid embryo autofluorescence; express and purify a chimeric MpRV-fluorescent protein in *E. coli*; analyse the localisation pattern of MpRV- and P2C- fluorescent protein chimeras in early-stage *M. persicae* embryos by confocal microscopy and image analysis.

3.2 Contributions to this chapter

P2C-mCherry and untagged mCherry were provided by Dr. Grant Hughes, Dr. Ian Bennet, and Dr. Mukund Madhav (Liverpool School of Tropical Medicine). Injection procedure optimisation was also done with advice from Dr. Grant Hughes, Dr. Ian Bennet, and Dr. Mukund Madhav.

3.3 Results

3.3.1 mCherry is the best fluorophore for localisation experiments in *M. persicae* embryos

To assess localisation of embryo-localising tags MpRV (derived from MpVg) and P2C, (derived from *Drosophila melanogaster* DmYPP1 (90)), I designed an experiment wherein adult females would be injected with fluorescent chimeric MpRV and P2C tagged proteins, and 24 hours later their ovaries would be dissected and checked for embryo-localised fluorescence.

To decide which fluorophore was best for this experiment, the autofluorescence of *M. persicae* embryos were analysed. This was done by confocal imaging of a *M. persicae* embryo under excitation wavelengths of 458nm, 488nm, 561nm, and 633nm, used for the four fluorophores CFP, GFP, mCherry and mPlum respectively (162). The resulting images were inspected visually (Figure 3.1), and mean fluorescence intensity across the embryo was measured in Fiji (163). This revealed that autofluorescence is most prevalent

under 488nm excitation, used for GFP, and least prevalent under 688nm excitation, used for mPlum (Table 3.1). This would appear to suggest that mPlum would be the most effective fluorophore to use. However, mCherry also showed a low fluorescence intensity, which is less spread across the embryo. Further, a comparison on FPbase (162) revealed that mCherry is 3 times brighter and more commonly used than mPlum. Therefore, I decided to use chimeric MpRV- and P2C-mCherry for my embryo-localisation experiments.

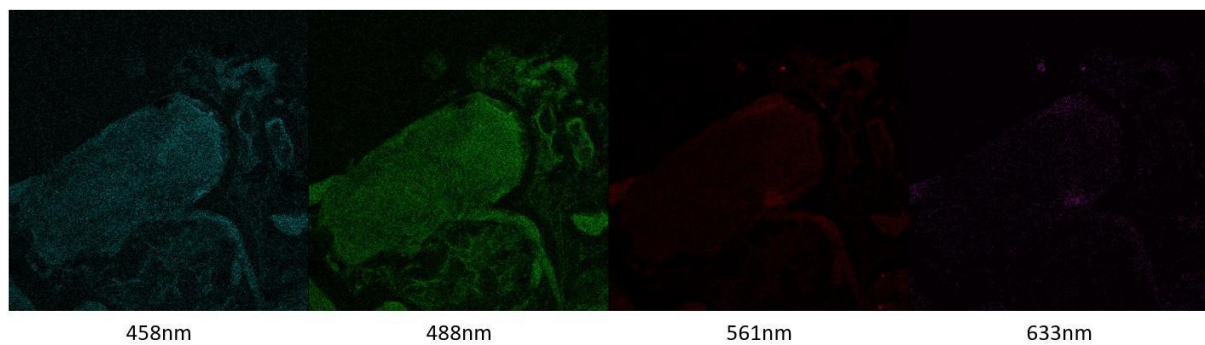


Figure 3.1: An *M. persicae* embryo imaged under four excitation wavelengths, corresponding to the fluorophores CFP, GFP, RFP, and mPlum.

Table 3.1: Mean fluorescence intensity across the *M. persicae* embryo under different excitation wavelengths.

Fluorescent protein	Wavelength	Area Examined	Mean Intensity
CFP	458 nm	156191	14.642
GFP	488 nm	162998.9	32.416
mCherry	561 nm	155468.5	14.054
mPlum	633 nm	143860	2.025

3.3.2 MpRV-mCherry is expressed and purified from *E. coli* on a large scale

To generate a construct for MpRV-mCherry expression, golden gate cloning was used to insert 6xHis, MpRV and mCherry modules into pOPIN-F5-RFP-K, one of the pOPIN-GG suite of expression vectors (164). This method successfully resulted in the generation of the plasmid pOPIN-F5-6His-MpRV-mCherry-6His-K (Figure 3.2), verified with sanger sequencing (Eurofins). The pOPIN-GG vectors allow for expression under induction with IPTG. Following IPTG induced expression, the protein was purified by IMAC and gel filtration using an AKTA Express system with a 5 mL His-Trap column and an S200 16/60 gel filtration column with no detergent. The recovered sample was analysed by SDS-PAGE (Figure 3.3). This revealed a strong band slightly larger than the expected size of 28 kDa (as calculated by ProtProgram by ExPasy (165)).

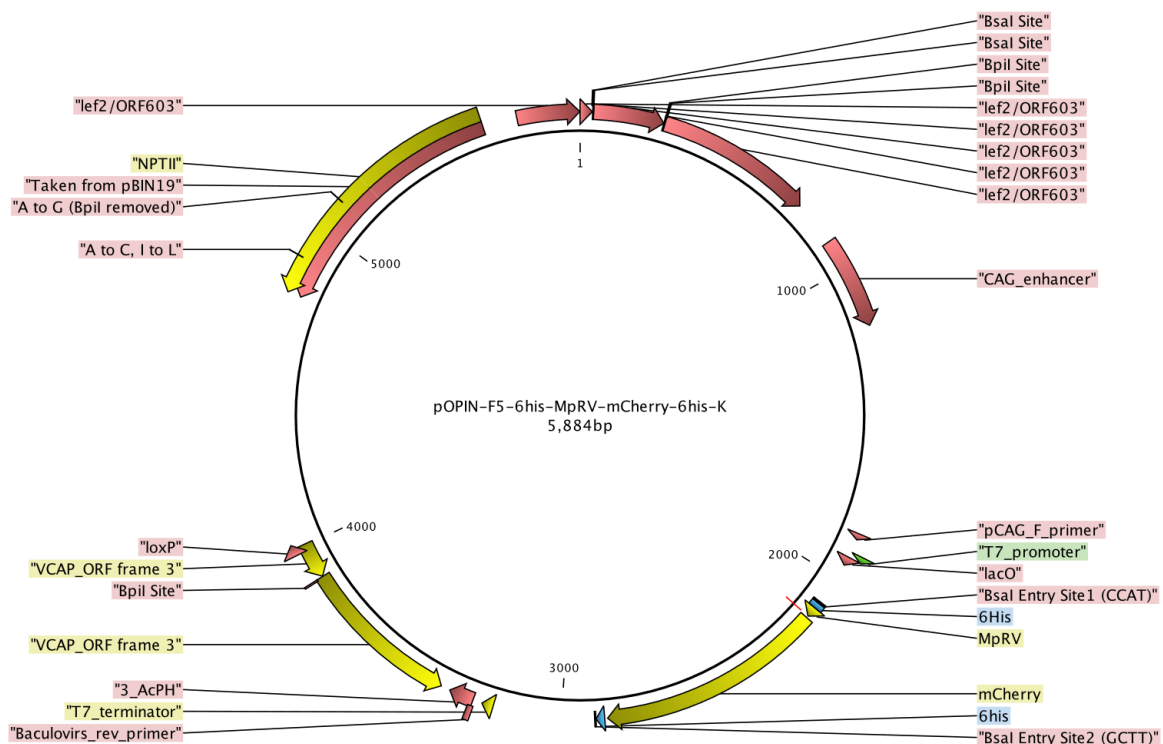


Figure 3.2: A plasmid map of pOPIN-F5-6his-MpRV-mCherry-6his-K, constructed using golden gate assembly.

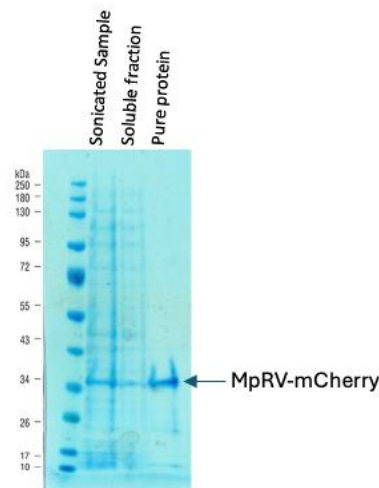


Figure 3.3: SDS-PAGE gel showing the sonicated sample, soluble fraction, and pure protein samples of MpRV-mCherry obtained by large scale purification. A single strong band in the 'Pure protein' sample represents MpRV-mCherry and is labelled with an arrow.

3.3.3 Optimisation of *M. persicae* injections

Over the course of injection experiments to analyse MpRV- and P2C-mCherry localisation, an injection procedure for *M. persicae* adult injections was optimised. Initially, aphids were immobilised using a vacuum pump attached to a plastic tube with a small pipette tip on the end. A vacuum was applied, and an aphid's dorsal side was placed onto the end of the pipette tip. Then, the ventral side of the abdomen was injected using a nanoject II (Drummond) fitted with a borosilicate micropipette. The injection mixture was ejected for 1-2 seconds, or until the aphid body visually expanded. This method was slow, as applying each individual aphid to the vacuum was a slow process. Further, injections lacked fine control of injection volumes and injection rate. Therefore, optimisations were made to increase speed and control.

Initially, micropipettes were pulled using a gravity-based micropipette puller, which heated the borosilicate capillary until gravity caused the bottom half to pull away from

the top, leaving two micropipettes with long sharp tips. This resulted in inconsistent needle lengths and shape. Therefore, in later experiments, micropipettes were pulled using a Sutter P-97 micropipette puller. The programme used was heat=500; pull=150; velocity=90; delay=235. This programme was optimised to give a long needle, which could be easily opened without damaging the tip. This was largely achieved by optimising the delay parameter: a short delay would result in a shorter point, while a longer delay gave the desired long, thin tip. This optimisation enabled consistent shape and sharpness in micropipettes used, ultimately causing less damage to aphids when injecting.

Next, the immobilisation of *M. persicae* adults was optimised. Visiting with Dr. Grant Hughes, Dr. Ian Bennet, and Dr. Mukund Madhav (Hughes lab, Liverpool School of Tropical Medicine) to learn about *Anopheles* mosquito injections, gave great insight. Mosquitos were first pre-chilled in a fridge (~4°C) for ~10 mins. Then, they were transferred to a cold plate under a light microscope for injection. For *M. persicae* a similar method was adopted. *M. persicae* adult females were pre-chilled in a petri-dish on ice for ~5-10 mins before injection. A folded sheet of paper towel was placed between the ice and the underside of the Petri dish to evenly distribute the cold temperature. Then, *M. persicae* were moved to a pre-chilled (~4°C) steel block under a microscope for injection. This successfully caused aphids to remain docile for injections and was a faster procedure than the use of a vacuum pump.

Finally, the injection equipment and procedure were optimised to gain fine control. A NANOLITRE 2020 (WPI) was used for injections, which allowed for specific injection rate

and total volume settings. Aphids were injected with 60 nL of injection mixture at a rate of 15 nL/s. This enabled consistency and reduced the likelihood of accidentally bursting the aphids by injecting with too much mixture. Ultimately, these optimisations resulted in a faster workflow.

3.3.4 MpRV-mCherry and P2C-mCherry localise to early-stage embryos 24 hours post injection of adult female mothers

To analyse embryo-localisation of our MpRV-mCherry and P2C-mCherry (gifted by Grant Hughes at Liverpool School of Tropical Medicine), groups of age-matched adult females were injected with the proteins. Each of 4 groups was injected with MpRV-mCherry, P2C-mCherry, untagged mCherry (gifted by Grant Hughes at Liverpool School of Tropical Medicine) or IMAC buffer A4. Untagged mCherry and IMAC Buffer A4 were used as controls; Untagged mCherry without an embryo-localising tag would be expected not to localise specifically to the embryo, and IMAC buffer A4 was the buffer in which the proteins were purified and should show no fluorescence. Ovaries from surviving injected females were extracted, and confocal microscopy was used to analyse red fluorescence intensity across 'early-stage' embryos. In this experiment, 'early-stage' refers to the earliest observable stage embryo in the image. The experiment was repeated 4 times. All experiments were done under the same conditions, and confocal microscopy images were taken with the same settings. Therefore, data from all four experiments could be compared. Further, a linear mixed effects model was used for statistical analysis, which accounted for multiple data points per ovary, and date of experiment.

Visual observation of the images from ovaries of MpRV-mCherry and P2C-mCherry injected mothers revealed fluorescence in early-stage embryos. I also observed some fluorescence in early-stage embryos of ovaries originating from untagged mCherry injected mothers. Very little fluorescence was observed in ovaries from IMAC buffer A4 injected mothers (Figure 3.4).

The data collected is presented in supplementary table S3. This data was plotted as a box plot, and significance values were calculated by linear mixed effects model in R (Figure 3.5). This revealed that fluorescence across early-stage embryos in MpRV- and P2C-mCherry groups was significantly greater than the IMAC buffer A4 negative control ($p = 0.0116$ and 0.0124 respectively), but not the untagged mCherry control.

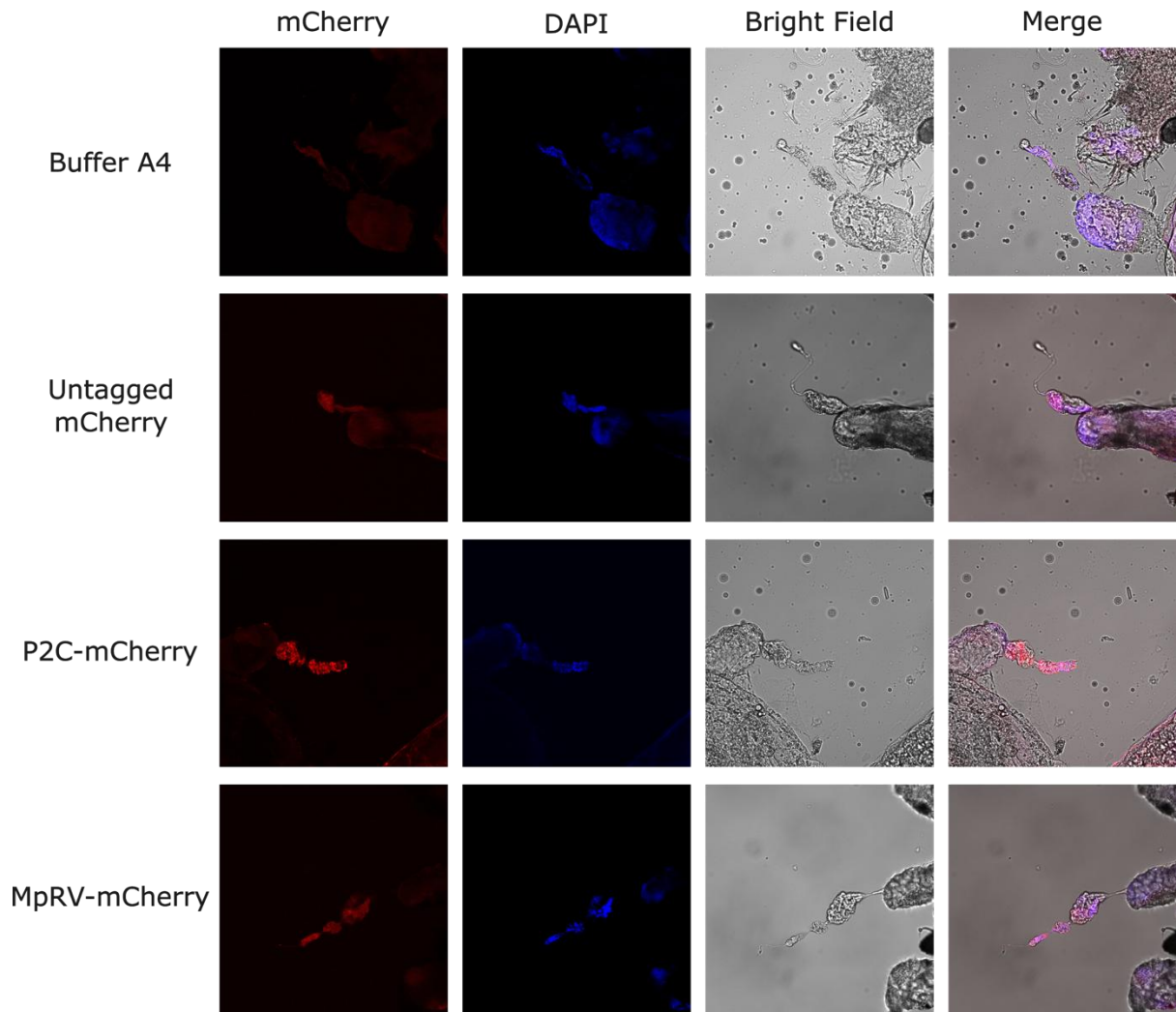


Figure 3.4: Example confocal microscopy images of early-stage embryos in ovaries originating from injected mothers. Images shown from left to right are mCherry signal, DAPI signal, bright field signal, and a merge. Microscope power and gain settings were kept consistent to enable comparison of fluorescence intensity between samples. A total of 8 images were taken from 5 ovaries dissected from mothers injected with buffer A4, 6 images from 4 ovaries from mCherry injected mothers, 11 images from 5 ovaries from MpRV-mCherry injected mothers, and 15 images from 7 ovaries from P2C-mCherry injected mothers.

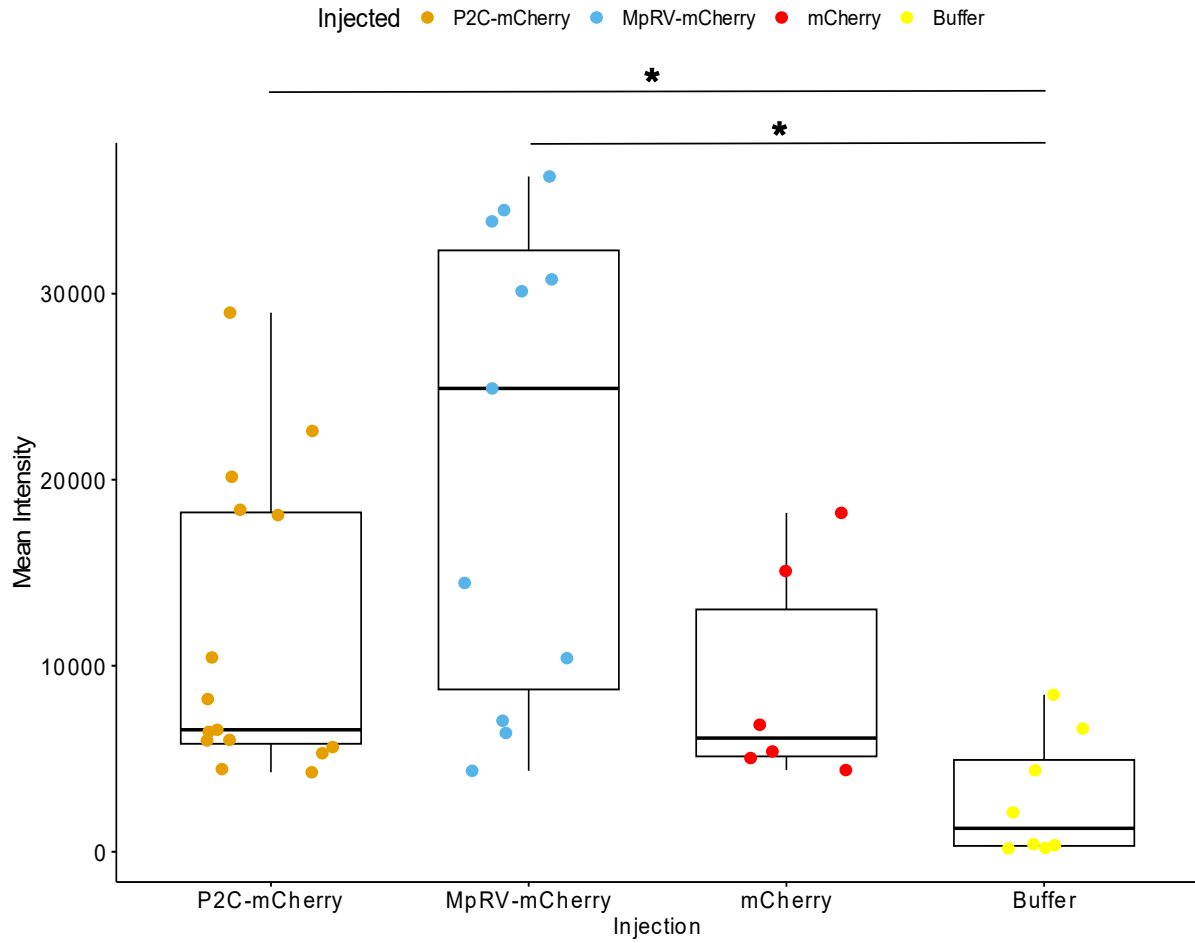


Figure 3.5: A Box plot showing mean fluorescence intensity of 'early-stage' embryos of ovaries from *M. persicae* adult females injected with P2C-mCherry (orange), MpRV-mCherry (blue), untagged mCherry (red) or IMAC buffer A4 (yellow). Statistical significance was calculated by a linear mixed effects model and is shown here as * = $p < 0.05$. A total of 8 images were taken from 5 ovaries dissected from mothers injected with buffer A4 (yellow), 6 images from 4 ovaries from mCherry injected mothers (red), 11 images from 5 ovaries from MpRV-mCherry injected mothers (blue), and 15 images from 7 ovaries from P2C-mCherry injected mothers (orange).

3.4 Discussion

We established and optimised an experimental procedure for adult *M. persicae* injection. Using this method, we injected adult female aphids with recombinantly expressed MpRV- and P2C-mCherry chimeras to analyse their embryo-targeting capacity. Dissection of ovaries out of injected mothers ~24 hours post injection, followed by confocal microscopy and fluorescence intensity analysis, we show that MpRV- and P2C are both capable of targeting early-stage embryos after injection into asexual adult females.

We optimised an injection procedure for use on adult *M. persicae* females. By, using cold treatment to immobilise the aphids, we were able to inject adult females at a good pace. Borosilicate micropipettes were suitable for our purposes, but it is possible that quartz micropipettes would further increase survival rate. Indeed, quartz needles were used to inject *B. tabaci* in a previous ReMOT control study (49). However, a Sutter-P2000 would be required to prepare these needles. Further, the use of a high precision NANOLITRE 2020 (WPI) injector, we achieved consistency in injection volumes and reduced the chances of aphid damage during injection.

Fluorescence microscopy shows that upon injection, recombinantly expressed MpRV- and P2C-mCherry can localise cargo to ‘early-stage’ *M. persicae* embryos. This methodology follows previous studies. In the initial ReMOT control study, the peptide P2C was derived from the *D. melanogaster* protein DmYPP1 by deletion assay (90). To establish whether this peptide could localise cargo to embryos, a fluorescence chimeric

protein P2C-EGFP-Cas9 was made and injected into adult females. The ovaries were then checked for fluorescence (90). This has become a standard experiment in other ReMOT control studies (49, 89, 94, 96-99, 101, 144). P2C has successfully enabled ReMOT control in most arthropods to which it has been applied (89, 90, 94-98, 100, 101, 144, 166). Interestingly, in *B. tabaci*, P2C was unsuccessful and so BtKV was found to be an effective alternative, which also enabled ReMOT control in *Rhodnius prolixus* (hemiptera) (49, 98). In other species, Vg derived peptides have also been used; in black legged ticks, *Ixodes scapularis*, a peptide derived from *I. scapularis* Vg8 was used, and in the silkworm, *Bombyx mori*, a Vg derived peptide, found both by MSA and deletion assay, called 'BmOTP' was successful (100, 101). I therefore decided to test the more general P2C, and an aphid-specific Vg-derived peptide (MpRV) in *M. persicae*, both of which were effectively taken up by 'early-stage' *M. persicae* embryos in asexually reproducing aphids. Analysis shows that 'early-stage' embryos extracted 24 hours post-injection of MpRV- and P2C-mCherry display significantly higher fluorescence intensity than the buffer A4 control. However, this experiment could be improved by using better fixing methods, such as formaldehyde, to limit distortion of tissues for analysis. Further, this would enable accurate staging of aphid embryos; in this experiment, fluorescence intensity was measured across the earliest available stage embryo in each image, but the specific embryo stage is not clear. By accurately staging the embryos, and better tissue integrity in images, the uptake of P2C- and MpRV-mCherry could be localised to specific stages, or cells.

Fluorescence microscopy showed a strong signal from early embryos after injection with the MpRV or P2C tagged mCherry, but the quantitative data did not show a significant

difference from free mCherry. This may be due to uptake by chance, where mCherry molecules are taken up along with Vg molecules, due to being close to the sites where this takes place. This mechanism is the basis for DIPA-CRISPR, a method similar to ReMOT control, but lacking a localisation tag (167). Commercial Cas9 is complexed with a sgRNA before injection into adults, where it is taken up randomly by endocytosis. The injection is timed to coincide with high vitellogenesis, so that much Vg induced receptor mediated endocytosis is occurring at the time, increasing the chance of uptake of Cas9-sgRNA RNPs (167). I have stipulated that vitellogenesis is likely to be very low in asexual *M. persicae*, so the uptake of mCherry molecules would also be low, which may explain why, in my results, there is a non-significant decrease in fluorescence when compared to MpRV- and P2C-mCherry. This may also suggest that DIPA-CRISPR could be a viable method for CRISPR-Cas9 gene editing in aphids but shows lower potential than ReMOT control using P2C or MpRV.

To my knowledge only one other attempt to find a similar system in aphids has been done, in which highly conserved cell penetrating peptides (CPPs) were used in similar experiments, where they were tagged to mVenus and injected into adult female pea aphids (*A. pisum*) (168). The CPP 'PEN', derived from a homologue of *D. melanogaster* antennapedia, was fused to mVenus at the C-terminus (168). PEN had previously been shown to cross membranes in an energy-independent manner, but no specific embryo-targeting has been reported (169). The protein mVenus-PEN was shown to localise to various tissues, including embryos, and the bacteriocyte within these embryos (168). However, there was no specific localisation to early-stage (pre-blastoderm) embryos, which is key to heritable germline gene editing in aphids. To this date there is no

publication to my knowledge which uses PEN to enable ReMOT control. The finding that P2C and MpRV can transduce cargo to early-stage embryos shows their promise in enabling gene editing in pre-blastoderm embryos, which could give rise to heritable gene edits.

The peptides P2C and MpRV could be used to target other cargo to embryos, including smaller Cas proteins such as Cas ϕ -2 (71) and Cas12f (72), who's smaller size could enable higher efficiency gene-editing by ReMOT control. Further, their requirement for an AT rich PAM site enables more varied target selection in AT rich organisms such as aphids (71, 72). One drawback of gene editing by ReMOT control so far is that it only facilitates gene knockouts, with no capacity for knock-in using a HDR template. One method of inserting transgenes which has a long history in the field of gene therapy is the use of viral particles such as adeno-associated viruses (AAVs) (170). These particles have been shown to be modifiable using peptides to target specific tissues; modification of the capsid peptides in AAVs has allowed transduction across the blood-brain barrier in mice (171). Similar engineering approaches could potentially be used to deploy targeted AAV gene delivery in insects, including via the use of peptides such as P2C and MpRV for embryo targeting. Another virus-like particle (VLP) system that has been developed is plant expressed CPMV VLPs, which can be used to present epitopes of disease, creating vaccines, and to package custom RNA (172, 173). This system could be adapted for embryo targeting in aphids using MpRV and P2C to introduce CPMV particles containing RNA for RNAi, or for protein expression of CRISPR machinery. Aphids vector multiple viruses, some of which are taken up by the aphid upon phloem feeding (10). *Beet western yellows virus* (BWYV), a luteovirus, is transmitted by aphids in a persistent, circulative

manner (174). The virions are taken up upon feeding and cross the intestinal membrane into the haemolymph, before crossing into the salivary glands for further transmission (174). Understanding the mechanisms by which the virions cross the intestinal membrane could inform design of VLPs that could be expressed and packaged *in planta* and delivered to aphids via feeding, before crossing the intestinal membrane and localising to embryos.

For our purposes, MpRV and P2C can be used, as in other ReMOT control studies, to target Cas9-sgRNA RNPs to embryos for gene editing. From previous studies we can be confident that P2C-Cas9 carries out its nuclease function with no issue, and is easy to express and purify (89, 90, 94-97, 101, 144), but the same remains to be seen for MpRV-Cas9. The expression, purification and in vitro validation of MpRV-Cas9, and sgRNAs to a marker target are presented in Chapter 4. Also presented is in vitro testing of P2C-Cas9-sgRNA RNPs against our marker gene of choice.

3.5 Materials and methods

3.5.1 Preparation of *M. persicae* ovary samples for confocal microscopy for autofluorescence analysis

Adult female *M. persicae* were dissected in PBS using pipette tips with pins taped to the thin end, under a Leica M125 C dissection microscope. The underside abdomen of each aphid was pierced to release the developing embryos. The samples were then fixed in 4% formaldehyde in PBS at pH 7 for two hours. After this, the formaldehyde solution was removed and the samples were washed using PBS followed by water, and then transferred to slides. The samples were then soaked in Prolong glass (Invitrogen P36984), covered with a cover slip and allowed to harden overnight. The following day, slides were sealed using nail varnish and kept at room temperature until examination.

3.5.2 Confocal microscopy and fluorescence intensity analysis for autofluorescence analysis

One sample was examined using a Zeiss LSM780 confocal microscope. Embryos were viewed in each of four channels corresponding to four potential fluorophores which are excited at different wavelengths. These were CFP, GFP, mCherry and mPlum, which are excited at 458 nm, 488 nm, 561 nm and 633 nm respectively (162). A mean fluorescence intensity across the embryo under each channel was then calculated using Fiji (163). The

channel resulting in the lowest intensity compared to the respective fluorophores brightness would be suitable for use.

3.5.3 Generation of an MpRV fragment for golden gate assembly

An MpRV fragment suitable for golden-gate assembly was generated by ordering synthetic oligos of the forward and reverse sequence. The acceptor plasmid and other parts are cut in the cloning reaction with BsaI. BsaI recognition results in a staggered cut downstream of the recognition site. Golden-gate cloning exploits this; the BsaI recognition sites cause cleavage downstream, therefore the insert itself will not contain the recognition site, therefore after insertion, the DNA will not be cut by BsaI again. Golden-gate cloning relies on BsaI generated overhangs for assembly of fragments in the correct order; the 4bp overhangs caused by these staggered cuts can be chosen to be complementary to the previous, and the next fragment or acceptor in the sequence (175). Therefore, written into these oligos were 4bp overhangs at each end. The 5' overhang was 5'-CCAT-3' on the forward strand, complementary to the entry site 1 of pOPIN-F5-RFP-K. The 3' overhang was 5'-CATT-3' on the reverse strand, complementary to the 5' overhang of the next fragment, which in this case was mCherry. Also written into these oligos was a coding sequence for a 6xHis tag upstream of MpRV. This places 6His-MpRV at the N-terminus of chimeric protein. Finally, an extra two base pairs, 'GC', were written into the 3' end of MpRV to maintain framing. These oligos were ordered from Merck and are presented in Table 3.2 The two oligos were mixed at equal concentrations (100µM) and mixed with 10x annealing buffer followed by dilution with nuclease free water (Merck) to achieve a 1x annealing buffer working concentration. The annealing buffer consisted of

100 mM TRIS at pH7.5, 500 mM NaCl and 10 mM EDTA. The mixture was heated to 95°C for 2 minutes, followed by 70 cycles of decreasing temperature at increments of 1°C per cycle, starting at 95°C and ending at 25°C. Then 5 µL of the reaction product was visualised by gel electrophoresis on a 1% agarose gel in 1x TAE. The concentration and purity of the product was determined by Nanodrop (ThermoFischer). The concentration of the product was 1960.3 ng/µL and had A260/280 and A260/230 ratios of 1.84 and 2.14 respectively. This was diluted stepwise 10x and then 13x to achieve a product at approximately 15 ng/µL.

Table 3.2: Forward and reverse oligonucleotides ordered for generation of a golden-gate suitable MpRV fragment.

Oligo ID	Sequence
GG_6his-MpRV_F	CCATATGCATCATCATCATCACCGCCCGAGCTTTGCGGCGCAAGAAACCGGCGTGTA TGGCAAATGCAACGTGCAGTATCTGGTGGC
GG_6his-MpRV_R	CATTGCCACCAGATACTGCACGTTGCATTTGCCATACACGCCGGTTTCTTGCGCCGCAAA GCTCGGGCGGTGATGATGATGATGATGCAT

3.5.4 Generation of an mCherry fragment for golden gate assembly

The fragment for mCherry was generated by PCR. The mCherry fragment required BsaI recognition sites at each end, with 4bp overhangs at the cut sites complementary to the previous and the next overhang. These were written into the primers. The forward primer contained the overhang 5'-AATG-3', complementary to the 3'-TTAC-5' overhang of the reverse strand of the MpRV fragment. The reverse primer contained the overhang 5'-

CGAA-3', complementary to the 5'-TTCG-3' overhang on the forward strand of the 6xHis tag part. The annealing section of the primers were designed against the 5' and 3' ends of the mCherry coding sequence of the plasmid 'pET mCherry LIC cloning vector (u-mCherry)' in CLC main workbench. pET mCherry LIC cloning vector (u-mCherry) was a gift from Scott Gradia (Addgene plasmid # 29769 ; <http://n2t.net/addgene:29769> ; RRID:Addgene_29769). The plasmid was received as a bacterial stab. This was grown overnight in 10 mL LB with ampicillin, in a 37°C shaking incubator. From this culture, 500µL was mixed with 1 mL 40% glycerol and stored at -80°C. The rest was subject to miniprep using a QIAprep spin miniprep kit (QIAGEN, #27106) to extract the plasmid. As the entire coding sequence was required, the annealing section of the primers were designed from the 5' and 3' end of the sequence and have G-C clamps at their 3' ends, with the last 6bp containing 50% GC content. The reverse primer also contained an extra 'GC' to maintain framing. These primers are presented in Table 3.3.

PCR was carried out using Phusion DNA polymerase (ThermoFischer, #F530S). To ensure enough fragment was retrieved, 450 µL reactions were run. In each reaction was: 10 µL 5x Phusion HF buffer; 1 µL 10 mM dNTPs; 2.5 µL of each primer at 10 µM, giving a working concentration of 0.5 µM; 1 µL of the plasmid at 1 ng/µL as template DNA; 1.5 µL of DMSO; 0.5 µL Phusion HF enzyme; 31 µL nuclease-free water (Merck). The cycling conditions used were: initial denaturation at 98°C for 30s; 35 cycles of denaturation at 98°C for 10s, annealing at 50°C for 30s, and extension at 72°C for 30s; final extension at 72°C for 5 minutes; hold at 10°C. The products of these reactions were analysed by gel electrophoresis on a 1% gel in 1x TAE. The products were purified using a QIAquick PCR purification kit (QIAGEN, #28104), and final product concentration and purity was

measured using a Nanodrop (ThermoFischer). The final product had a concentration of 12.8 ng/μL, an A260/280 ratio of 2.26, and an A260/230 ratio of 1.91.

Table 3.3: Primers used to amplify an mCherry fragment for use in a golden gate cloning reaction.

Primer ID	Sequence
AATG-mCherry_F	GGTCTCAAATGGTGAGCAAGGGCGAGGAGGATAACATGGC
CGAA-mCherry_R	GGTCTCACGAAGCCTACTTGTACAGCTCGTCCATG

3.5.5 Golden gate assembly

The cloning reaction was performed as a one pot reaction. The inserts are usually mixed at a 2:1 molar ratio, using 200 ng of the acceptor plasmid. The MpRV insert is especially small, so a 7:1 ratio was desired for this insert. The required mass of each insert was calculated using NEBioCalculator. This revealed that 13.61 ng of the MpRV insert was required for a 7:1 molar ratio against 200 ng pOPIN-F5-RFP-K. Therefore, 1 μL of the 15 ng/μL stock was added to the one pot reaction. For a 2:1 molar ratio of the mCherry insert to the acceptor, 45.89 ng was required. Therefore, 3.6 μL of the 12.8 ng/μL stock was added to the one pot reaction.

A C-terminal 6xHis tag was incorporated in the cloning reaction directly from plasmid pICSL50025 from TSL SynBio. Upon cutting with BsaI in the one pot reaction, the 6xHis tag coding sequence is excised, leaving a 5'-TTCG-3' overhang on the forward strand, complementary to the 5'-CGAA-3' overhang on the reverse strand of the mCherry fragment, and a 5'-AAGC-3' overhang on the reverse strand, complementary to the

forward strand overhang of entry site 2 of pOPIN-F5-RFP-K. For a 2:1 molar ratio of pICSL50025 to 200 ng of pOPIN-F5-RFP-K, 354.5 ng of the plasmid was required as calculated with NEBioCalculator. Therefore, the 1026 ng/ μ L stock of pICSL50025 was diluted to 354.5 ng/ μ L, and 1 μ L of this was added to the one pot golden gate reaction.

Along with the insert fragments and pICSL50025, 5 units of type IIS restriction enzyme BsaI, 200 units of T4 DNA ligase, 1.5 μ L T4 DNA ligase buffer (NEB) and 1.5 μ L 10X bovine serum albumin (BSA) (NEB). This was subject to thermal cycling in the following programme: 20 seconds at 37°C, then 26 cycles of 3 minutes at 37°C followed by 4 minutes at 16°C, then 5 minutes at 50°C and 5 minutes at 80°C, followed by holding at 16°C.

3.5.6 Heat-shock transformation of DH5 α *E. coli* with the MpRV-mCherry construct and screening for successful clones

For transformation, 5 μ L of the golden gate reaction product was transformed into DH5 α *E. coli* (NEB, #C2987) via heat shock. The cells (50 μ L per reaction) were mixed with 5 μ L of the ligation on ice for 30 minutes before 30 seconds of heat shock at 42°C, followed by 2 minutes cooling on ice. Then 950 μ L SOC was added, and the cells were placed in a 37°C shaking incubator for recovery before being spread on LB agar plates containing kanamycin for selection. Plates were left in a standing incubator at 37°C overnight. The acceptor plasmid contained an RFP cassette which would be removed if cloning had been successful. Therefore, colonies that were pink could be ignored, and only white or translucent colonies were screened for the correct sequence. These colonies were grown overnight in liquid culture and plasmids were purified by Miniprep (Qiagen) followed by a diagnostic digestion reaction with HindIII, and sanger sequencing (Eurofins) using the forward and reverse primers ‘pCAG_F_primer’ and ‘Baculovirus_rev_primer’ respectively, which were previously annotated on the pOPIN-F5-RFP-K plasmid map. Colonies containing successfully cloned constructs were grown in 5 mL LB overnight shaking at 37°C. 500 μ L of this culture was mixed with 1 mL 40% glycerol and stored at -80°C, while the rest was subject to miniprep.

3.5.7 Expression and purification of MpRV-mCherry

MpRV-mCherry-6xHis was obtained via large scale protein expression in *E. coli*. The cloned construct pOPIN-F5-MpRV-mCherry-6xHis-K was transformed into BL21(DE3)

cells (NEB, #2572) by heat shock according to the manufacturers protocol. The cells were then incubated shaking at 37°C for 2 hours. Finally, 100 µL was spread onto LB agar plates containing kanamycin. These plates were incubated at 37°C in a standing incubator overnight. As BL21 cells express the protein, successfully transformed colonies which appeared pink were deduced to have good expression levels. One transformed colony was used to inoculate 2 starter cultures of 100 mL LB with kanamycin, which were incubated shaking at 37°C overnight. The next morning, 4x 500 mL cultures of LB with kanamycin were inoculated with 20 mL each of starter culture. These were incubated shaking at 37°C until an OD600 of 0.6-0.8 was reached, at each point the cultures were stored at 4°C until induction with IPTG. Expression was induced by addition of IPTG to a working concentration of 1 mM. The cultures were then incubated shaking at 18°C overnight. The next morning, the cultures were centrifuged at 24,000 g for 30mins to collect the pellet. Each pellet was then resuspended in 12.5 mL of IMAC buffer A1 (50 mM Tris-HCL, 50 mM glycine, 0.5 M NaCl, 20 mM imidazole, 5% glycerol) and combined to achieve a 50 mL sample, to which one 'c0mplete, EDTA-free protease inhibitor cocktail tablet' (Roche, #11873580001) was added. Cell lysis was done by sonication (instrument) with 1s on, 3s off pulses for a total of 20 mins, after which 100 µL was collected as a 'total sonicated' sample. After this, the 50 mL sample was split into two centrifuge tubes and centrifuged at 30,000 g for 30 mins. From the lysate, 100 µL was taken as a 'soluble fraction' sample. The rest of the lysate was combined in a 50 mL falcon tube. This was subject to purification using an AKTA express system to perform IMAC with buffer A1 and elution with buffer B1, which was immediately subject to gel filtration using a 5 mL HisTrap High Performance column (Merck) and an S200 16/60 gel filtration column with buffer A4 and no detergent. The purified protein was eluted in IMAC

buffer A4 into a 24-well plate. The wells that matched the elution peak were combined, and 100 μ L was taken as a 'pure protein' sample for analysis.

Protein samples were analysed using SDS PAGE. Precast protein gels were used (NuPAGE 4-12% Bis-Tris, Invitrogen, #NP0321BOX). 2 μ L of each sample was mixed with 5 μ L 4x NuPAGE LDS sample buffer (Invitrogen, NP0007), 12 μ L buffer A1 and 1 μ L IDT, boiled for 5 minutes for denaturing, and loaded onto the gels, in tanks containing NuPAGE MOPS SDS running buffer (Invitrogen, #NP0001). The proteins were separated at 100 volts for 10 minutes followed by 200 volts for ~20 minutes. Gels were then moved to square plates and stained with ReadyBlue Protein Gel Stain (Sigma-Aldrich, #RSB-1 L) via shaking for ~1 hour. This was then washed off with distilled water. Gels could then be imaged by scanning. Upon verification of a pure protein product, aliquots of 50 μ L were snap-frozen in liquid nitrogen and stored at -80°C for later use.

3.5.8 Experimental design for assessment of uptake of chimeric embryo localising mCherry proteins by early-stage *M. persicae* embryos

To assess the ability of early-stage *M. persicae* embryos to uptake tagged mCherry proteins, an experiment was designed where age-matched, 7-day old asexual females were injected with each chimeric embryo-localising protein, followed by removal of ovaries from surviving females 24 hours later, and then confocal microscopy to assess fluorescence intensity across early-stage embryos. For this experiment, the proteins tested were MpRV-mCherry, purified as described, and P2C-mCherry. Buffer A4, and

untagged mCherry were used as controls. P2C-mCherry and untagged mCherry were both gifted by Grant Hughes at Liverpool School of Tropical Medicine.

3.5.9 *M. persicae* age-matched colony preparation

For consistency, and to coincide with maximum fecundity, adult females were injected at the same age. To achieve this, age-matched colonies were set up. Adult females were taken from a stock colony and placed on *Brassica rapa* in clip cages, each containing 15 adult females. A total of 150 adults were used. These adult females were allowed to reproduce for 24 hours before they were removed from the clip cages. The offspring produced in this period were then allowed to develop on the same plant in clip cages until injection on the 7th day.

3.5.10 *M. persicae* adult female injection procedure

3.5.10.1 Micropipette pulling

Initially borosilicate micropipettes were pulled using a vertical puller which used a heating element to heat the centre of the capillary. The capillaries used in this case were Drummond Microcaps (Drummond, # 1-000-0001/CA). Once the glass was sufficiently melted, gravity would cause the bottom part of the machine (holding the bottom end of the capillary) to drop, causing the formation of two micropipettes. These were useable but would often be too blunt for aphid injection after opening against a slide. Therefore longer, sharper needle was desired. As a result, we decided to use a Sutter P-297 micropipette puller. The programme used was: heat=500; pull=150; velocity=90; delay=235. Further micropipettes were gifted by Grant Hughes at Liverpool School of

Tropical Medicine. The length of the needle tips meant that they could be opened by cutting with scissors, allowing for a more specific opening site in a sharper part of the needle.

3.5.10.2 Needle preparation with injection mixture

Pulled needles were backfilled with mineral oil (Sigma-Aldrich, #M5904-5 ML), using a syringe with a long stainless-steel tip. In initial experiments, the backfilled needle was then inserted into a Drummond Nanoject II injector. The needle was then emptied until the injector would empty no more. A 20 μ L drop of the desired injection mix was placed on parafilm. The tip of the needle was carefully inserted into the drop and then was filled with the injection mix. The result should be that the injection mix is at the front (tip) end of the needle. In later experiments, a World Precision Instruments 'NANOLITRE2020' was used for injection. Micropipette filling was done in a similar way, but after filling with oil and insertion into the system, 2,500 nL of oil was ejected at 50 nL/s. The injection mixture was then sucked up at the same volume and rate. The injection mixtures used were Buffer A4, MpRV-mCherry, purified as described, P2C-mCherry and untagged mCherry, both gifted by Grant Hughes at Liverpool School of Tropical Medicine. The proteins were normalised to 3 mg/mL using Vivispin 500 columns (Sartorius, #VS0112), centrifuging for ~30mins at 6000 g.

3.5.10.3 Aphid immobilisation

For effective microinjection the aphids needed to be immobilised. To begin with, aphids were held in place using a vacuum applied to a tube which was capped with a small pipette tip. The aphid was placed on its back on the pipette tip hole before being injected. Whilst this effectively immobilised the aphids, it was inefficient as only one aphid could be injected at a time. Therefore, we decided to immobilise the aphids by using cold. A frozen cold (-20 °C) aluminium block was placed under the microscope with a glass slide on top. Aphids were pre-anesthetised on ice. Once ready for injection, 5 aphids would be lined up on the slide, remaining anaesthetised due to the cold block.

3.5.10.4 Aphid injection

Aphids were injected on the ventral side of the abdomen. The needle was held at ~45° by a micromanipulator. The wheels on the micromanipulator were turned by hand to move the needle to inject the aphid. Once the needle was inside the aphid, when using the NanojectII, the 'empty' button was held for 1-2 seconds. When using the NANOLITRE2020, aphids were injected with 60 nL at a rate of 15 nL/s. Forceps were used to move or hold the aphid when necessary. The injected aphids were then moved onto a *Brassica rapa* leaf in a petri-dish at room temperature and allowed to recover before being moved to a controlled environment for 24 hours. After this the injected aphids were ready for dissection.

3.5.11 Sample preparation for MpRV- and P2C-mCherry embryo localisation analysis

To prepare samples for microscopy, ovaries were dissected out of mothers on slides in PBS under a Leica M125 C dissection microscope, 24 hours post injection. This was done using two pairs of forceps, using one to clamp the head, and the other to pull from the lower abdomen to release the ovaries. The bodies were then removed, and samples were stained with DAPI at 1 mg/mL. This was then washed off with PBS. As much PBS as possible was removed by pipetting, then wax was placed around the sample to slightly raise the cover slip, minimising squashing. To fix the samples, ~40 μ L ProLong glass (Invitrogen, #P36984) was added to the sample. The 1.5 gauge cover slip was then placed on top. The samples were then kept at room temperature overnight to allow the ProLong glass to harden.

3.5.12 Confocal microscopy for MpRV- and P2C-mCherry embryo localisation analysis

Confocal microscopy was used to visualise injected protein uptake into early-stage embryos. For this, a Zeiss LSM880 was used. Each sample was visualised at 20x magnification under three channels. These were bright field, 561nm laser for mCherry visualisation, and 405nm laser for DAPI visualisation. To produce comparable images, settings were kept consistent throughout. Under the bright field channel, the master gain was set to 240, and the digital gain was set to 1.0. The 561nm laser was set to a power of 3.0 and a pinhole of 33.6. The master gain was set to 875, and the digital gain was set to 1.0. Then 405 nm laser was set to a power of 0.4 with a pinhole of 25.4. The master gain

was set to 750, and the digital gain was set to 1.0. Images were taken on a single plane, at a resolution of 1024 x 1024.

3.5.13 Fluorescence intensity analysis across early-stage embryos

Confocal microscopy images were analysed in Fiji (163). The images were imported one by one. For each image, the colour channels were split, and the 'red' channel image was saved as a .tiff file to retain true fluorescence intensity values. A composite image was then constructed by re-merging the channels. This composite image was used to draw an area around the early-stage embryos, which was stored in the 'ROI manager'. This same area was then applied to the 'red' channel image, and mean fluorescence intensity was measured. A total of 8 images were taken from 5 ovaries dissected from mothers injected with buffer A4, 6 images from 4 ovaries from mCherry injected mothers, 11 images from 5 ovaries from MpRV-mCherry injected mothers, and 15 images from 7 ovaries from P2C-mCherry injected mothers.

Statistical analysis was performed in R Studio and a box plot was constructed. A linear mixed effects model was used to account for multiple data points from the same injected mother. This also accounted for the different dates of the experiments. Mean fluorescence intensity was plotted for each injection mixture. The script for this analysis is included in the appendix (Script to generate fluorescence intensity Box plot in R). The R packages used were readxl, ggplot2, dplyr, stringr, svglite (158), ggpubr (176), and lme4 (177).

4 Generation and validation of embryo-targeting Cas9 fusion RNPs

4.1 Introduction

In Chapter 2 I presented work finding a candidate VgR binding peptide called MpRV, derived from MpVg. Then, in Chapter 3 I presented work showing that MpRV and the DmYP1 derived peptide P2C can localise cargo to early-stage *M. persicae* embryos after injection into adult females. The next step is to produce MpRV-Cas9 (and P2C-Cas9) chimeras in *E. coli* and assess if the chimeras are functional as nucleases using *in vitro* cutting assays. To conduct these assays, it will also be necessary to design sgRNAs for target genes, to test if the MpRV-Cas9 chimera cleaves the target genes in the presence of the sgRNAs.

In vitro nuclease assays have been a standard method of testing Cas9 nuclease function since Jinek et al., 2012 showed CIRPSR-Cas9 use as a programmable guided nuclease (58). These are assays involving the generation of a complex of Cas9 and guide RNAs (also known as crRNA:tracrRNA, or sgRNA), and incubating the complex with a fragment of DNA containing a sequence complementary to the sgRNA and protospacer-adjacent motif (PAM) (58). The Cas9 protein scans for PAMs in the DNA, then cuts DNA if the upstream ~20bp sequence is complementary to the sgRNA (58). Similar assays were used to determine the activities of the chimeric Cas9 proteins P2C- and BtKV-Cas9 towards cleaving marker genes in the presence of sgRNAs, in previous ReMOT control

studies (49, 90). Hence, it should be feasible to heterologously produce and purify MpRV- and P2C-Cas9 in *E. coli*. To test the activity of these two Cas9 chimeras, it will also be necessary to select *M. persicae* genes that are targeted in the ReMOT control experiments, and design suitable sgRNAs for these targets.

In previous insect ReMOT control studies, pET28a-(P2C/BtKV)-Cas9-Cys constructs were used to produce the chimeras with a T7 leader peptide and an SV40 NLS (49, 90). This was done using standard protocols designed for pET28a-based expression in *E. coli*. The bacteria were grown in TB with 9.4 g/L K₂HPO₄, 2.2 g/L KH₂PO₄ for producing BtKV-Cas9, whereas standard LB medium was used for P2C-Cas9 production (49, 90). These conditions do not deviate much from those used for producing wild-type/unfused Cas9 in *E. coli* (49, 58, 90), and show that the fusions of small peptides, such as P2C and BtKV, to Cas9 have minor impacts on the efficiency of Cas9 production by *E. coli*.

To optimize CRISPR-Cas methods, including the ReMOT control technology, it is essential to target genes whose knockout results in clear and distinct phenotypes in the organism (49, 89, 90, 94-101). Nonetheless, the phenotype should not be lethal to the organism. For this reason, genes involved in determining eye colour are often targeted for optimization. For example, introducing mutations that prevents expression of *kynurenine monooxygenase (kmo)* in *A. aegypti*, *A. sinensis*, and *C. pipiens* *kynurenine monooxygenase (kmo)* and results in white eye phenotypes (90, 94, 95). Along the same line, mutations of *cinnabar (cin)* in *T. castaneum*, *N. vitripennis*, and *Blattella germanica*, cause bright red eye phenotypes (97, 144). In hemipteran insects, such as *D. citri*, *R.*

prolixus and *B. tabaci*, mutations in the gene *white* give rise to varied eye discolorations, ranging from bright red to white (49, 98, 99).

The goal herein is to optimise protocols to purify P2C- and MpRV-Cas9 chimeras. Furthermore, I wish to identify a *M. persicae* gene that may be used for optimisation of the CRISPR-CAS/ReMOT control technology. Finally, I aim to test if the purified Cas9 chimeras cleave *M. persicae* DNA in the presence of sgRNAs in *in vitro* reactions.

4.2 Results

4.2.1 Optimisation of protocols for producing the MpRV-Cas9 chimera in *E. coli*

To obtain the MpRV-Cas9 chimera, MpRV was synthesised and cloned into pET28a-Cas9-Cys (by NBS biologicals) upstream of the Cas9 sequence but downstream of a sequence corresponding to an N-terminal His tag. *E. coli* carrying the resulting pET28a-His-MpRV-Cas9-Cys plasmid were grown to produce His-MpRV-Cas9. Whereas the bacteria are likely to produce His-MpRV-Cas9, the protein was not detected upon purification with Ni-NTA agarose beads (Figure 4.1A). Sequencing of the entire plasmid did not reveal any obvious cloning mistake. The most likely explanation for the failure to purify the His-MpRV-Cas9 protein is that the His tag, which is essential for binding to the Ni-NTA agarose beads, is not adequately exposed. This lack of exposure likely prevents efficient extraction of the chimera from the *E. coli* protein extract.

Next it was attempted to add a His tag at the C-terminal end of Cas9. Fortunately, the plasmid sequence uncovered a His tag sequence downstream of a Cas9 stop codon in the pET28a-His-MpRV-Cas9-Cys plasmid. Using SDM, a new plasmid was generated in which the 3' stop codon was removed, resulting in a plasmid that produces His-MpRV-Cas9-His. *E. coli* cells successfully produced the His-MpRV-Cas9-His protein (Figure 4.1B). However, most of the protein ended up in the insoluble fraction that is not available for binding to the Ni-NTA agarose beads, as shown by the large band between the ~130 and 180kDa markers on the SDS-page gel (Figure 4.1B).

To resolve the problem of the His-MpRV-Cas9-His protein being insoluble, a Gb1 solubility tag was incorporated into the plasmid, located at the N-terminus, between the N-terminal His tag and MpRV, via the Gibson assembly method (178). This method relies on the incorporation of ~20 bp sequences at the 5' and 3' ends of the insert, which are complementary to the sequences flanking the cut site in the vector. A T5 5' exonuclease turns these complementary sequences into overhangs, which anneal, are extended by a DNA polymerase, and are ligated together by a Taq ligase to fit the insert into the vector. This design facilitates precise integration of the insert into the vector through homologous recombination. Between Gb1 and MpRV, a 3C protease sequence is incorporated. This enables removing the His-Gb1 part of the His-Gb1-3C-MpRV-Cas9-His protein via 3C protease treatment to generate the MpRV-Cas9-His chimera. In its final state, after SDM and insertion of Gb1-3C, the expression plasmid was named pET28a-His-Gb1-3C-MpRV-Cas9-His. This plasmid also encodes a C-terminal NLS from SV40, and a HA tag (Figure 4.2)

The His-Gb1-3C-MpRV-Cas9-His protein was successfully purified from *E. coli* and the Ni-NTA agarose beads, as evidence by a clear band of the correct size in the 'Pure' lane of the protein gel (Figure 4.1C). Based on this promising result, a large-scale purification procedure was carried out using IMAC and gel filtration. Initial IMAC and gel filtration resulted in high amounts of eluted protein of the correct size (Figure 4.3A). The protein preparation was subjected to 3C cleavage overnight, followed by a second IMAC and gel filtration to separate the His-Gb1 and MpRV-Cas9-His parts of the protein. Fractions coming from the IMAC were collected in wells of elution plates, and small amounts of each well were analysed on an SDS-PAGE gel. Fractions in wells C3-C6 and D3-D6 contain a band corresponding to the MpRV-Cas9-His protein of slightly higher than 180kDa (Figure 4.3B). Lower molecular weight proteins were also detected, as was observed for MpRV-mCherry in chapter 3. However, the MpRV-Cas9-His protein purification was considered sufficiently pure for planned experiments. Proteins of wells C3-D3 were combined and concentrated to 2.48 mg/mL using a 50kDa gated Vivispin column. The concentrate was split into 20 μ L aliquots, which were snap frozen in liquid nitrogen and stored at -80°C for later use.

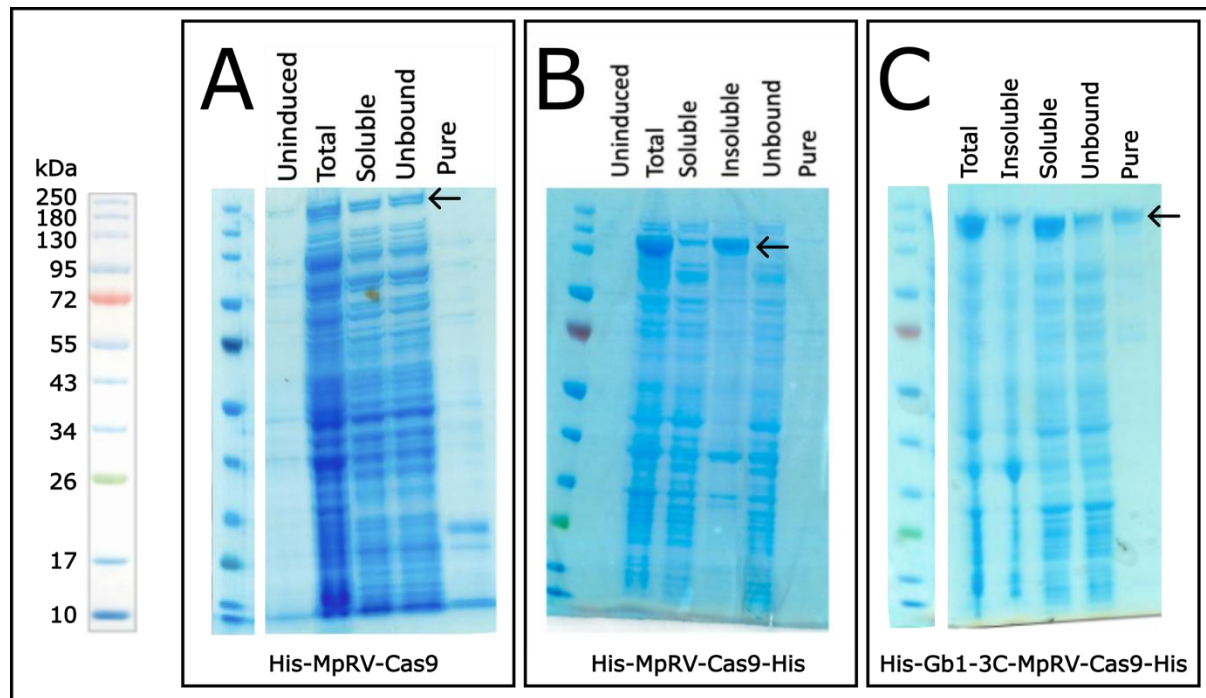


Figure 4.1: SDS-Page gels showing different stages of optimisation of purification of MpRV-Cas9. (A) Initial purification attempted after expression from the plasmid pET28a-MpRV-Cas9, giving rise to the protein His-MpRV-Cas9. The arrow shows most of the desired protein in the 'unbound' fraction. (B) Purification attempted after SDM removal of a 3' stop codon which had excluded the C-terminal His tag. The plasmid used was named pET28a-MpRV-Cas9-His and gave rise to the protein His-MpRV-Cas9-His. The arrow shows most of the desired protein in the 'insoluble' fraction. (C) Successful purification of the protein His-Gb1-3C-MpRV-Cas9-His, after expression from the plasmid pET28a-Gb1-3C-MpRV-Cas9-His. The arrow indicates the desirable protein is recovered in the 'pure' fraction.

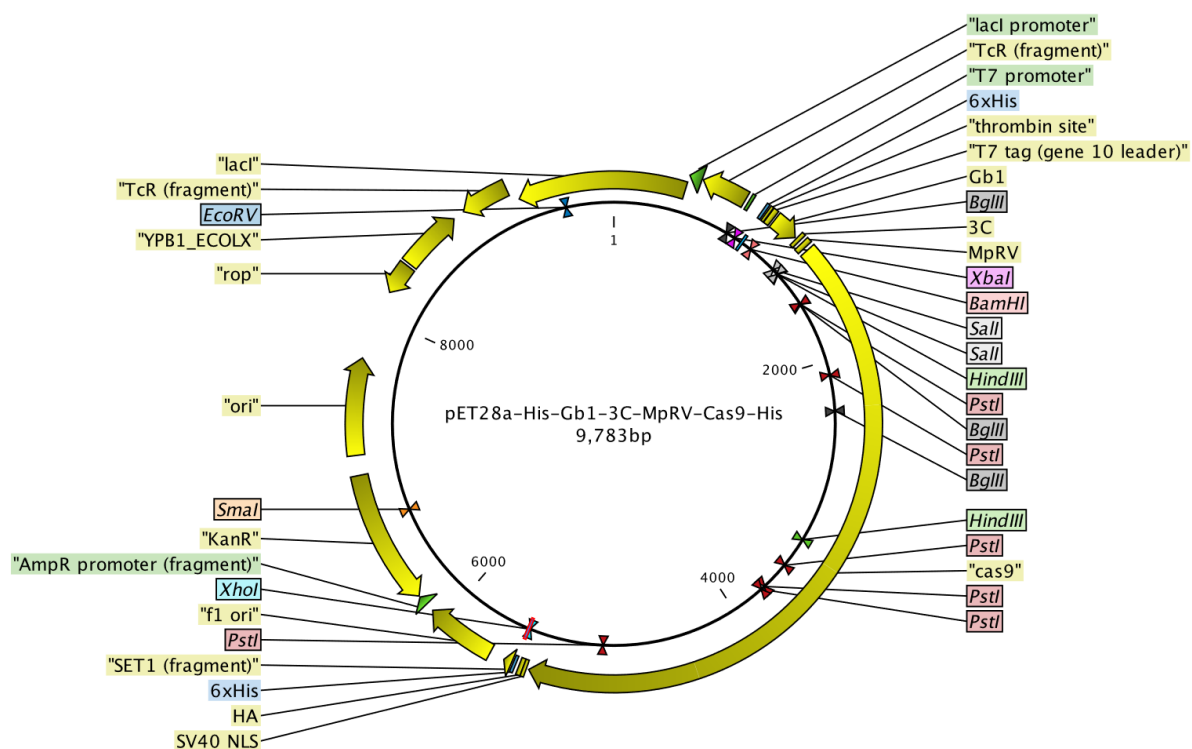


Figure 4.2: A plasmid map for the final version of the MpRV-Cas9 expression plasmid. This is a variation of pET28a-MpRV-Cas9-Cys, edited by SDM to add a C-terminal His tag, and by Gibson cloning to add an N-terminal Gb1 solubility tag with a 3C protease recognition site.

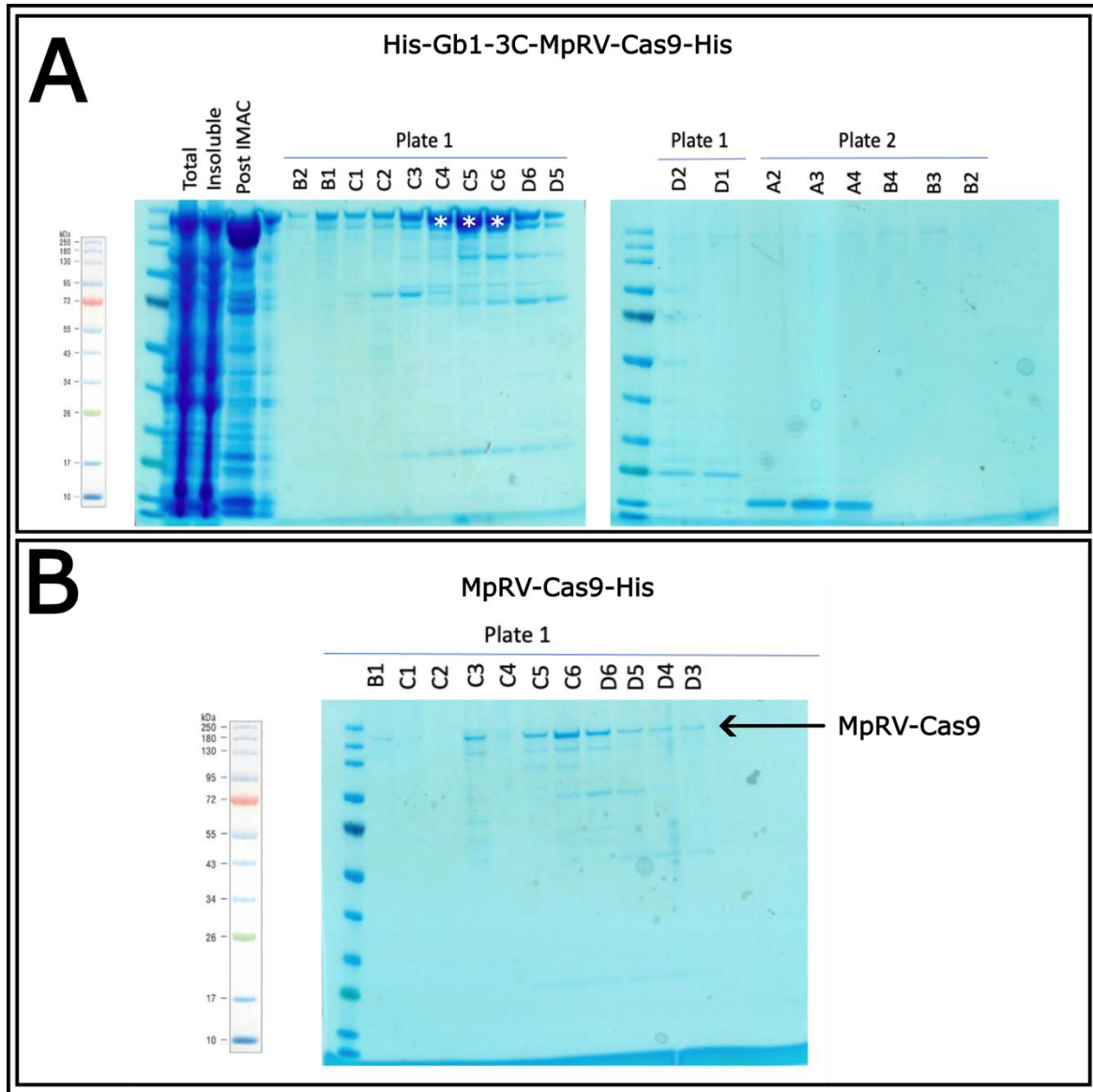


Figure 4.3: SDS-PAGE gels showing two IMAC stages (pre and post 3C cleavage) of MpRV-Cas9 purification. (A) After the first IMAC and gel filtration, large bands (labelled by white *) of ~160-190 kDa were observed in elution wells C4-6. These elutions were collected and subject to 3C cleavage followed by a 2nd IMAC and gel filtration. (B) After a 2nd IMAC and gel filtration, strong bands of ~180kDa were observed in wells C3-D3, which correspond to MpRV-Cas9, as indicated by the arrow.

4.2.2 Producing the P2C-Cas9 chimera in *E. coli*

Production of P2C-Cas9 was only conducted on a large scale as it had been successfully purified previously (90). The plasmid pET28a-P2C-g4sx3-Cas9 was constructed and gifted by Ian Bennet and Grant Hughes at Liverpool School of Tropical Medicine (Figure 4.4). Upon receipt of this plasmid, it was successfully transformed into *E. coli* both for storage and expression. Expression was carried out using IMAC and gel filtration on an AKTA express system. After initial IMAC and gel filtration, SDS-PAGE revealed many elution wells contained strong bands between ~160-180kDa (Figure 4.5A). This corresponds to the expected size of the translation region of the plasmid (T7 leader-His-thrombin recognition site-P2C-Cas9) which was ~170kDa. However, many strong bands of lower molecular weight were observed. Therefore, using the combined elutions from the wells containing P2C-Cas9, a second round of IMAC and gel filtration was run to further clean up the protein. SDS-PAGE revealed a band corresponding to P2C-Cas9 at ~180kDa (Figure 4.5B). The wells containing this band were combined and concentrated to 1.48 mg/mL. This was then split into 20 μ L aliquots, snap frozen in liquid nitrogen, and stored at -80°C for later use.

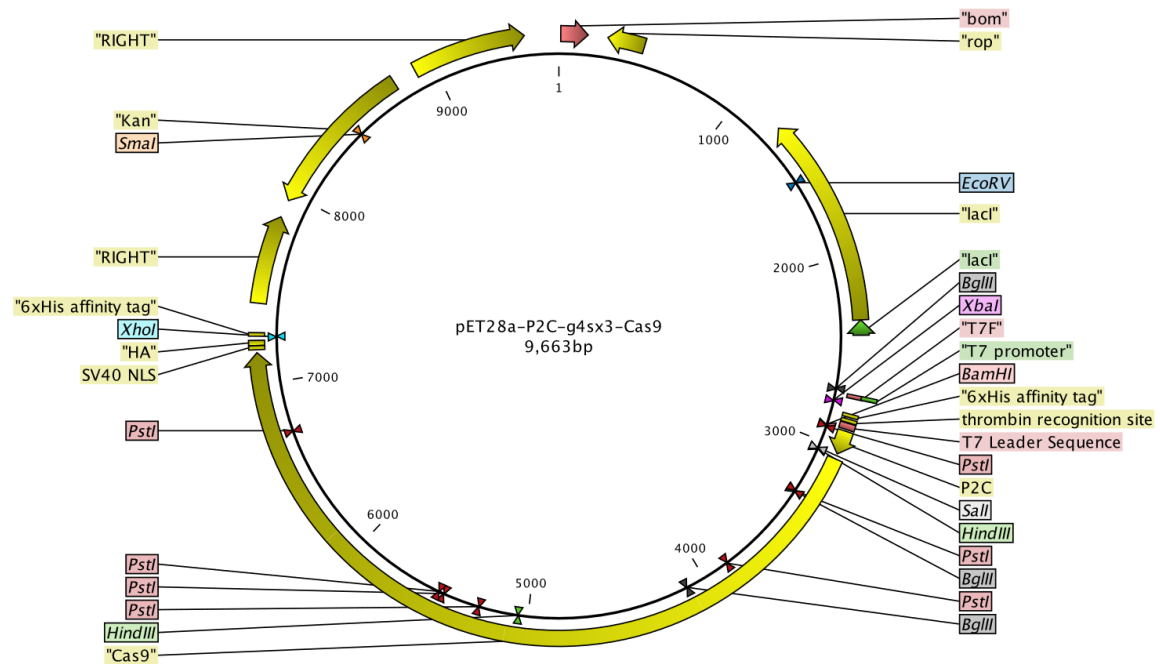


Figure 4.4: A plasmid map of pET28a-P2C-g4sx3-Cas9, created and gifted by Ian Bennet and Grant Hughes at Liverpool School of Tropical Medicine.

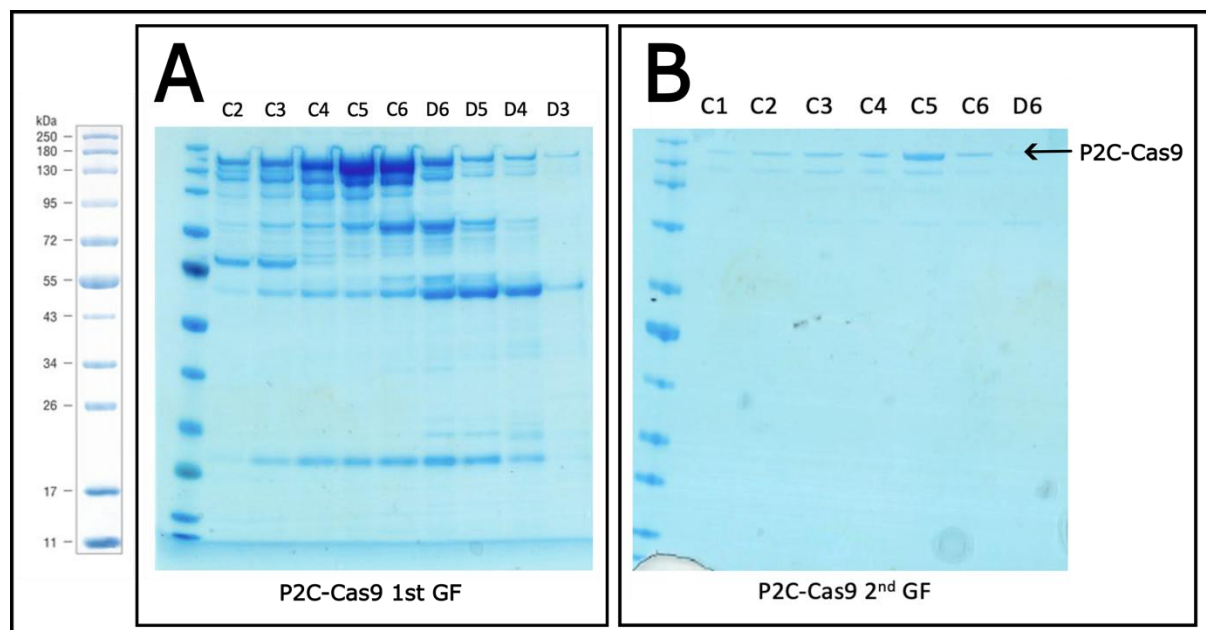


Figure 4.5: SDS-PAGE gels showing elutions from the first and second IMAC and gel filtration for purification of P2C-Cas9. (A) After the first IMAC and gel filtration, multiple wells contained bands corresponding to P2C-Cas9, but there were many contaminants. (B) After the second IMAC and gel filtration, much fewer contaminants were observed, but P2C-Cas9 remained, as marked by an arrow.

4.2.3 Design of sgRNAs for the *M. persicae* white gene

Existing literature on aphid gene functions was reviewed to identify potential targets for optimization using the CRISPR-Cas system. In the whitefly *B. tabaci*, mutation of the *white* gene was shown to have significant effects on eye colour (49). Since few genes have been functionally characterized through knock-out or knock-down in aphids, we selected a homolog of the *white* gene in *M. persicae* as the most promising candidate for further study. Translated protein sequences of genes annotated in the *M. persicae* genome assembly v2.1 (154) were searched for proteins with high level of sequence similarity to the deduced protein sequence of the *B. tabaci* *white* gene (49) using BLASTp. This resulted in 197 candidates of which one has good alignment throughout the full length protein and had 61% identity and an E value of 0.0 (Figure 4.6). Beyond this candidate, the next *M. persicae* protein had only 33% similarity and was only partially aligned to the deduced protein sequence of *B. tabaci* *white*. Therefore, *M. persicae* is likely to have only one homologue of the gene *white*. To further investigate this, the corresponding mRNA sequence was aligned to the *M. persicae* genome assembly, and this identified one gene region encoding the *white* gene, indicating that *M. persicae* has one copy of the gene *white*.

The *M. persicae* *white* gene model comprised 13 exons that covered 21.2 kb of genomic DNA. These exons are sequentially transcribed to form a 2001 bp mRNA which is translated into a 667 amino acid protein (Figure 4.7). InterPro was then used to predict the key domains of *M. persicae* protein white (Figure 4.7). These regions were also mapped to the coding sequence (not shown), and sgRNAs were designed with a view to

disrupting these domains. Three sgRNAs were designed and named sgRNA1, 2 and 3, which targeted the walker A/P-loop (exon 3), the ABC transporter signature motif (exon 5), and the ABC transporter type 2 domain (exon 9) respectively (Figure 4.6). These sgRNAs were also subject to BLASTn against the coding sequences of the *M. persicae* clone 'O' v2.1 genome annotation, to check for possible off target effects. In all three cases, no similar sequences were found.

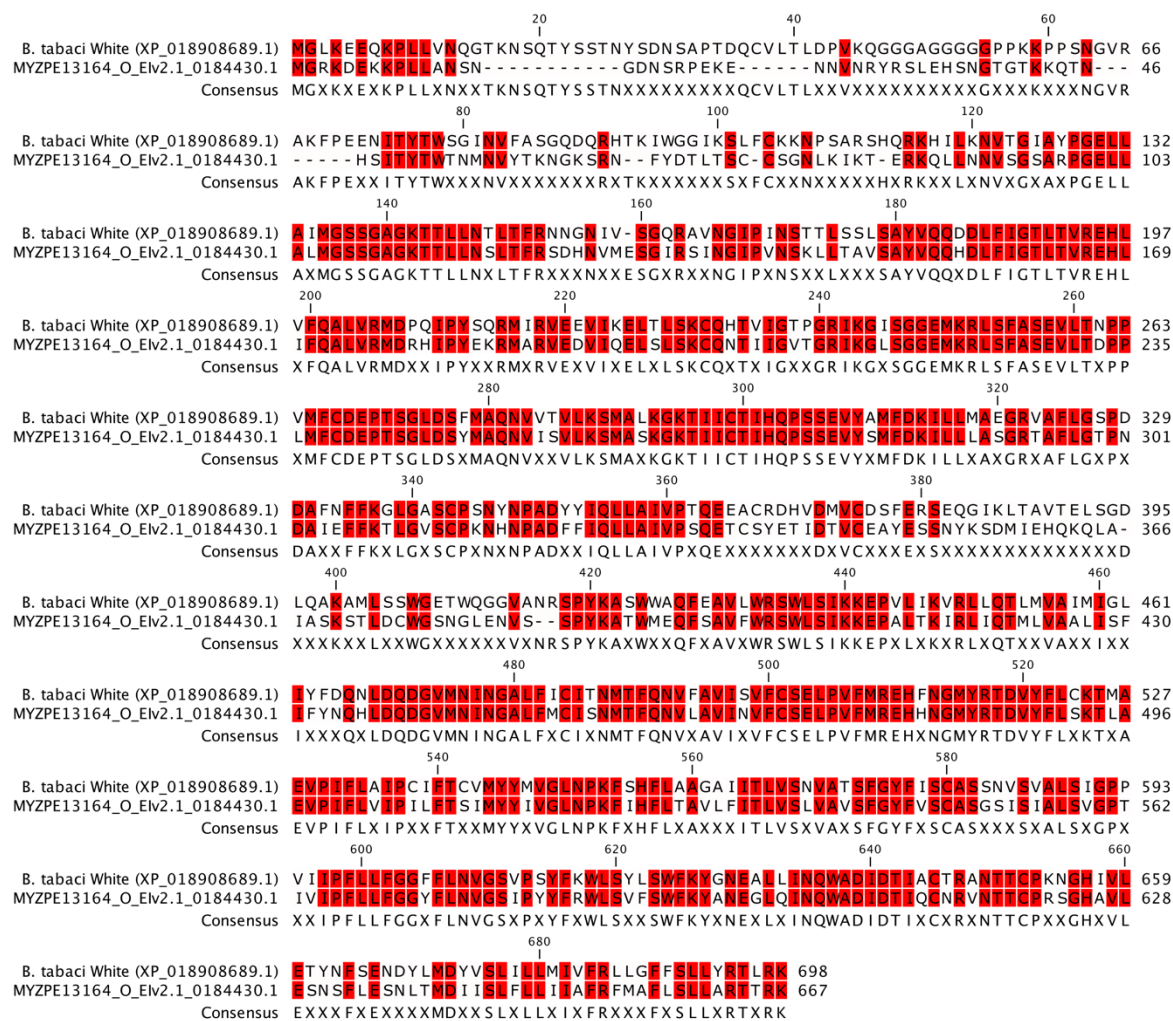
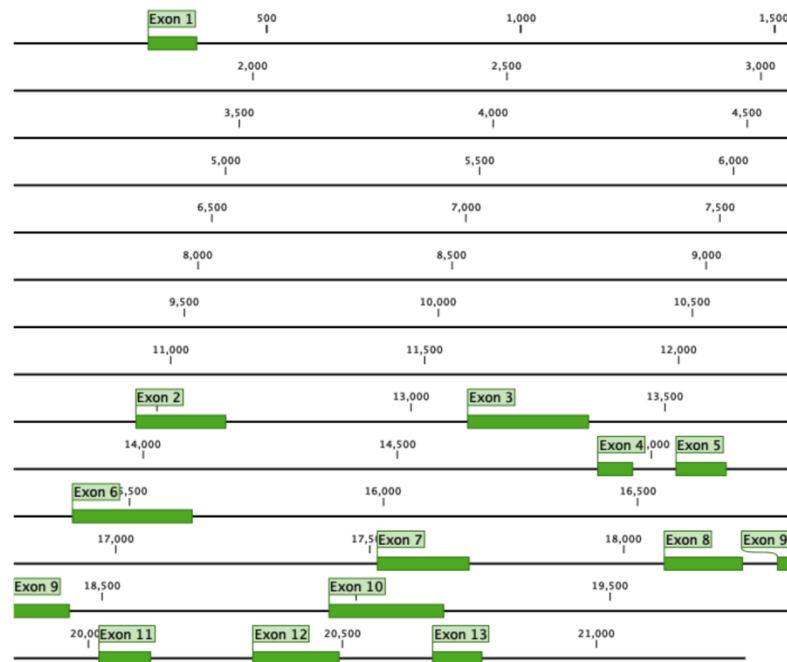
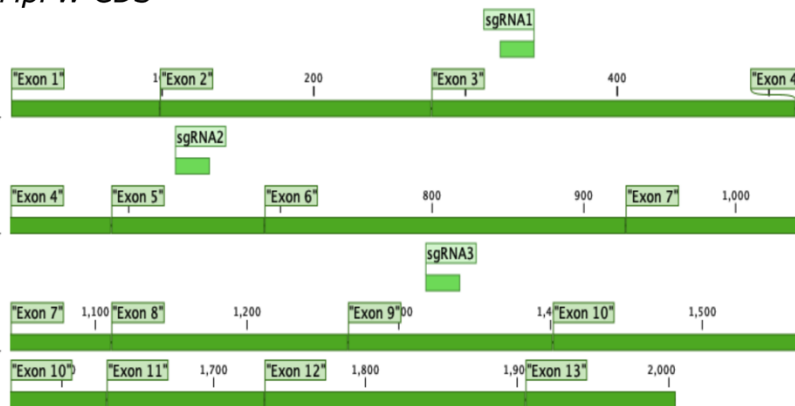


Figure 4.6: A recreation of the alignment of *B. tabaci* Protein White against the top hit when subject to BLASTp against the *M. persicae* Clone O v2.1 (154) proteome, made in CLC main workbench (QIAGEN). Conserved residues are coloured red.

MpPW gene



MpPW CDS



MpPW protein

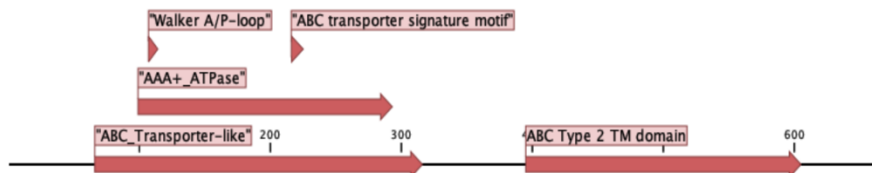


Figure 4.7: A map of the *M. persicae* white gene, coding sequence (CDS) and protein. The gene map shows the exons on the gene. The CDS shows the exons, and the three sgRNAs designed to those key domains. sgRNA1 lies in exon 3, targeting the Walker A/P-loop, cutting at position 328 of the coding sequence. sgRNA2 lies in exon 5, targeting the ABC transporter signature motif, cutting at position 631 of the coding sequence. sgRNA3 lies in exon 9, targeting the ABC type 2 transporter domain, cutting at position 1318 of the coding sequence. The protein map shows the key structural domains of White as predicted by InterProScan.

4.2.4 *In vitro* assays to assess if MpRV-Cas9-His and P2C-Cas9 chimeras cut *M. persicae* white gene fragments in the presence of sgRNAs

Next, I wished to assess if the MpRV-Cas9-His and P2C-Cas9 chimeras show endonuclease activity. It was previously shown that purified Cas9, and P2C-Cas9 cleave DNA fragments in the presence of guide RNAs *in vitro* (58, 90). Here, I setup a similar assay. First, three fragments of the *white* gene, each containing an sgRNA target site, were amplified using genomic *M. persicae* DNA as a template. Amplifications generated single fragments of 2.6kb (Fragment White 1), 5.5kb (Fragment White 2) and 1.6kb (Fragment White 3) in size (Figure 4.8 and 4.9). Then, I generated RNP complexes by combining MpRV-Cas9 or P2C-Cas9 solutions with sgRNA corresponding to Fragment White 1,2 or 3. Upon incubation of these mixtures with their target fragments for 30 mins and at 37°C, the mixtures were loaded onto a 1% agarose gel. A positive control using SpCas9 (NEB) was included.

The positive control (SpCas9) cleaved the DNA in the presence of sgRNAs. Fragment sizes were as expected, with 1.2 and 1.4 kb for Fragment White 1, 3.4 and 2.1 kb for Fragment White 2, and , 1.0 and 0.6 kb for Fragment White 3 (Figure 4.8 and 4.9). This indicates that the sgRNAs were functioning as expected. The sgRNA1 was most efficient at cleaving the DNA among the three sgRNAs, whereas sgRNA3 was the least efficient. I was pleased to find that chimeric P2C-Cas9 also cleaved the DNA into predicted fragment sizes in the presence of all three sgRNAs, indicating that the P2C-Cas9 chimera has endonuclease activity (Figure 4.8). Similar results were obtained for MpRV-Cas9 (Figure 4.9). As may be expected, P2C-Cas9 and MpRV-Cas9-His appeared less efficient

at cutting the DNA compared to SpCas9. It is possible that the fusions at the N-terminus of Cas9 reduces enzyme activity or that the chimera concentrations and/or purity may have to be further optimized. Nonetheless, these results show that the MprV-Cas9 and P2C-Cas9 purified from large scale *E. coli* expression are effective nucleases when complexed with sgRNAs that target *M. persicae* white gene fragments.

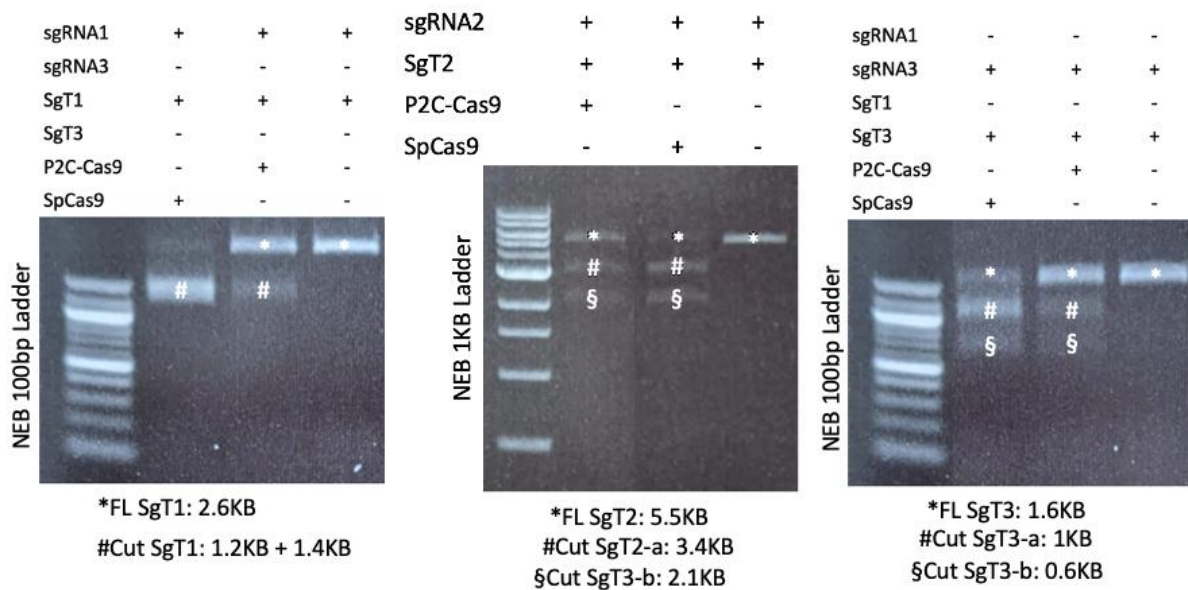


Figure 4.8: In vitro nuclease assays to assess P2C-Cas9 nuclease action with each of the white targeting sgRNAs. SpCas9 was used as a positive control. a negative control containing no Cas9 variant is also shown. In each image, the full-length fragment is denoted by '*' and the smaller bands observed post cleavage are denoted by '#' and '§'. In the case of sgRNA1, the bands expected are only 200bp apart in size, so one thicker band is observed corresponding to both the 1.2kb and 1.4kb fragments expected post-cleavage. In the cases of sgRNA2 and 3, two distinct bands are observed. For sgRNA2 these bands are 3.4kb and 2.1kb in size, while for sgRNA3, the bands are 1kb and 0.6kb in size.

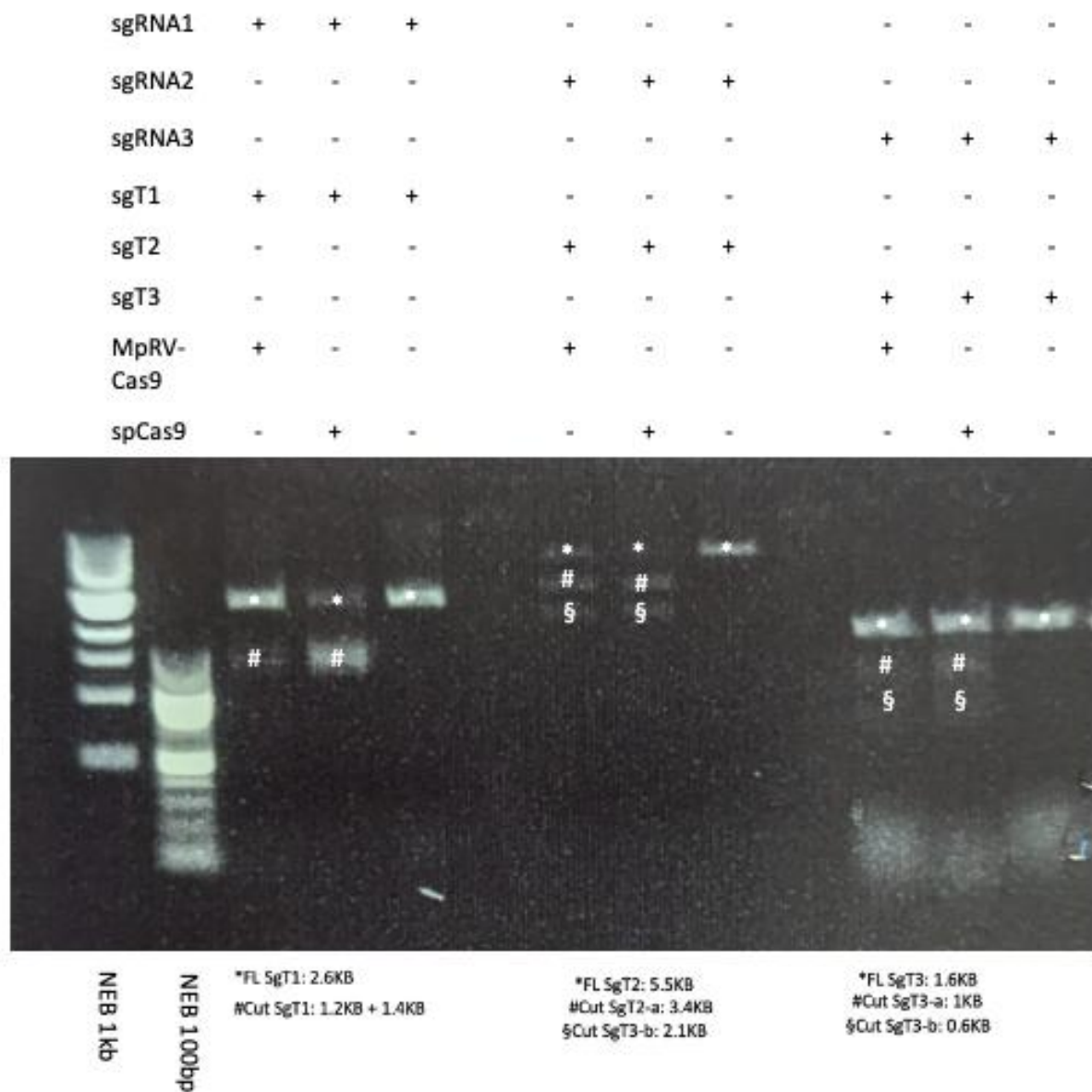


Figure 4.9: A single electrophoresis gel showing the results of an *in vitro* nuclease assay testing MpRV-Cas9 RNPs. The original DNA fragment is labelled with ‘*’ while the larger and smaller cleavage products are labelled with ‘#’ and ‘\$’ respectively. Similarly to P2C-Cas9, MpRV-Cas9 was able to cut all 3 fragments into the expected cleavage products when complexed with their respective sgRNAs. On testing sgRNA1 and 2, similar amounts of original fragment remained when compared to SpCas9 as were observed for P2C-Cas9. On testing sgRNA3, neither MpRV-Cas9, nor SpCas9 appear very efficient, both leaving much of the original fragment intact.

4.3 Discussion

In this chapter, I have optimized protocols to successfully purify P2C- and MpRV-Cas9 chimeras from *E. coli*. Furthermore, I mined the *M. persicae* genome sequences for a homologue of the gene *white*, which is likely to affect eye colour in aphids. Finally, I was able to generate complexes of the chimeras and sgRNAs to cleave aphid *white* gene fragments *in vitro*.

Purifying His-MpRV-Cas9 from *E. coli* was a challenge, likely because the His tag was not sufficiently exposed to bind the Ni-NTA agarose beads. This result was unexpected, because His-P2C-Cas9 and His-BtKV-Cas9 fusions were readily purified (49, 90). It is possible that structure of MpRV is somewhat rigid with the N-terminal His tag folding away from the exterior, thereby preventing it from binding the beads. Another possibility is that the MpRV fragment is bulky causing steric hindrance. It was also observed that the His-MpRV-Cas9 protein did not accumulate at high levels in *E. coli*. After introduction of a C-terminal His tag, the His-MpRV-Cas9-His protein ended up mostly in the insoluble fraction of the *E. coli* protein preparation. This suggests that the MpRV chimera may be somewhat toxic to the *E. coli* cells, as bacteria often deal with toxic proteins by depositing them in inclusion bodies, which generate protein aggregates that are insoluble (179). Moreover, protein insolubility can be caused by extrinsic factors such as pH and solvent additives, and intrinsic factors to do with peptide sequence and protein folding (180). Before the re-incorporation of the C-terminal His tag, the protein appeared to be soluble, therefore I thought it likely that the re-incorporation of the C-terminal His tag caused insolubility due to toxicity rather than an extrinsic factor. This situation was

somewhat rescued by adding a Gb1 solubility tag to the chimera. Gb1 is derived from the B1 domain of *Streptococcus* Protein G. *Streptococcus* Protein G is unusually stable for its size and lack of stabilising components such as disulfide bridges and prosthetic groups (181). Gb1 readily enables the expression and purification of a number of proteins in *E. coli* (182) and *Nicotiana benthamiana* (183). The His-Gb1-MpRV-Cas9-His chimera was found in the soluble fraction of the *E. Coli* protein prep, leading to its purification at sufficient quantities for the in vitro nuclease experiments. Whereas the amount of purified His-Gb1-MpRV-Cas9-His was high, the amount of purified MpRV-Cas9-His upon cleavage of the Gb1-His tag with the 3C protease was quite low compared to expectations, and moreover small protein fragments were detected in the purified fractions. It is possible that beyond putting proteins in inclusion bodies, the *E. coli* bacteria also cleave a part of the proteins to prevent toxicity. Altogether, this suggests that the production of MpRV-Cas9 requires some optimisation, and this could include adding a larger linker between the N-terminal His tag and MpRV, as well as optimisation of the *E. coli* growing conditions in such a manner that inclusion body formation and protein cleavage is prevented. Nonetheless, sufficient MpRV-Cas9-His was purified to proceed with downstream assays in this chapter, and it was satisfying to find out that the MpRV-Cas9-His chimera has nuclease activity.

Selecting a gene in *M. persicae* whose knockout leads to clear and observable phenotypes poses a significant challenge. To date, only a limited number of genes associated with important phenotypes in aphids have been functionally characterised. The major reasons are that, firstly, RNAi-based knock-down methods are highly inefficient in aphids. Furthermore, genetic modification techniques have been

developed exclusively for the pea aphid (26), and these methods present technical challenges that limit their widespread adoption across multiple laboratories. Consequently, potential target genes (commonly referred to as marker genes) for optimizing CRISPR-Cas-based knockout methods, which result in clear phenotypic changes, are scarce in aphids.

The *M. persicae white* gene was selected as a phenotypic marker for gene editing by ReMOT control. In *M. persicae* we considered using a carotenoid synthesis gene as a phenotypic marker. The most common *M. persicae* chromosome variant arises from a reciprocal translocation in autosomes 1 and 3. This rearrangement causes upregulation of carotenoid synthesis genes, which result in a pink-red body colour (184). Therefore, knocking out these genes may cause a return to a green body colour. These carotenoid cyclase/synthase genes are thought to have arisen by an ancient acquisition of fungus genes, which were duplicated in the aphid genome (185). This has the effect of the cyclase/synthase genes in *M. persicae* being multi-copy (184). I decided that it was likely we would have to knock out all copies of the gene to observe a phenotypic change, which would be more challenging if the ReMOT control process had low efficiency. Therefore, I decided to find a marker gene that is single copy in *M. persicae*. Many examples of marker genes have been found in *D. melanogaster*. The developmental gene *wigless* (*wg*¹), can result in *D. melanogaster* lacking wings and/or halteres when a recessive or homozygous mutation is introduced (186-188). Another such gene is *distal-less*, which causes malfunction in limb development in mutant alleles (189, 190). One of the best characterised marker genes is *white*, which encodes an ABC transporter protein responsible (along with *brown* and *scarlet*) for transport of tryptophan and guanine into

pigment cells in the eyes, which then give rise to the eye colour of *D. melanogaster* (191, 192). Loss of function mutations of *white* give rise to flies with altered eye colour, ranging from a lighter red to white (192). A *B. tabaci* homologue of this gene was used for the initial ReMOT control study in whiteflies, which gave rise to whitefly with bright red eyes when successfully knocked out (49). I therefore identified an *M. persicae* homologue of *white* as a suitable marker gene for ReMOT control experiments in *M. persicae* and designed sgRNAs targeting its signature domains.

In vitro nuclease assays using P2C- and MpRV-Cas9 showed that both proteins can cut the *white* gene fragments when complexed with each of the three sgRNAs designed against the signature domains of *M. persicae white*. These results show that my purified P2C- and MpRV-Cas9 samples were of decent quality and that the correct, functional protein was eluted. The bands observed from the in vitro nuclease assays showed successful cutting at the sgRNA target sites. Further, dialysis into an alternative buffer to IMAC A4 is unnecessary for the function of the Cas9 fusions in the presence of the cutting buffer NEB buffer r3.1. Therefore, P2C-Cas9 and MpRV-Cas9, both functional targeted nucleases, were ready for use in ReMOT control experiments in *M. persicae*. The experiments, optimisations and results will be discussed in Chapter 5.

Despite various challenges, this chapter demonstrates that purified P2C- and MpRV-Cas9 chimeras, when complexed with sgRNAs, successfully cleave *M. persicae white* gene fragments *in vitro*. This achievement provides a strong foundation to begin conducting experiments directly with *M. persicae*, as per work described in the next chapter (Chapter 5).

4.4 Materials and Methods

4.4.1 Generation of a construct for expression of MpRV-Cas9

4.4.1.1 Synthesis of pET28a-MpRV-Cas9-Cys

MpRV was synthesised into pET28a/Cas9-Cys. pET28a/Cas9-Cys was a gift from Hyongbum Kim (Addgene plasmid # 53261 ; <http://n2t.net/addgene:53261> ; RRID:Addgene_53261 (193). Synthesis was done by NBS Biologicals with MpRV being synthesised into the Sall cut site at the N-terminus of Cas9.

4.4.1.2 Whole plasmid sequencing

Sequencing of the plasmid pET28a-MpRV-Cas9-His was carried out by Plasmidsaurus, and results were analysed in CLC Main Workbench (QIAGEN).

4.4.1.3 Site-directed mutagenesis for removal of C-terminal stop codon

Site directed mutagenesis (SDM) was used to change the premature stop codon in pET28a-MpRV-Cas9-Cys from TAA to TAT (encoding tyrosine). This was done using the Q5 site directed mutagenesis kit (NEB, #E0554S) according to the manufacturers protocol. Primers for this were designed using NEBaseChanger, which also provided the recommended annealing temperature of 72°C. The primers were ordered from Merck and are presented in table 4.1. When used in the PCR step of the protocol, with an

extension time of 5 mins (30s/kb), this was successful in amplifying the full-length plasmid. The product of this PCR was subject to ligation followed by heat shock transformation into DH5 α cells (all following NEB's protocol).

Table 4.1: Primers used for SDM of pET28a-MpRV-Cas9-Cys to remove a C-terminal stop codon. The codon change is lowercase and written into the forward primer.

Primer ID	Sequence
Q5SDM_pETRV9_F	CTACGCCTGTtatGCGGCCGCAC
Q5SDM_pETRV9_R	TCGGGCACGTCGTAGGGG

4.4.1.4 Gibson cloning of Gb1 solubility tag to create pET28a-His-Gb1-3C-MpRV-Cas9-His

I designed a Gibson cloning strategy to insert a Gb1-3C sequence into the EcoRI cut site of the SDM treated pET28a-MpRV-Cas9-His, placing the Gb1-3C tag downstream of the N-terminal His tag. The insert sequence was generated by PCR using 'pGNC-6xHis-Gb1-Amsh2 level 1', generated previously by Matteo Gravino (a post-doc in the Hogenhout lab) as a DNA template. The annealing sections of the primers were designed to amplify the Gb1-3C part of this plasmid, and the overlapping flanking sequences of the EcoRI cut-site of pET28a-MpRV-Cas9-His were added. One extra base was added to the reverse primer to maintain framing. These primers were ordered from Merck and are presented in Table 4.2.

PCR was run using Q5 DNA polymerase (NEB, #M0491S) according to the manufacturers protocol, with the GC enhancer. To ensure enough product was retrieved, 4 50 μ L

reactions were set up consisting of 10µL 5x Q5 reaction buffer, 1µL 10 mM dNTPs, giving a working concentration of 200µM, 2.5µL of each primer at 10µM, giving a working concentration of 0.5µM, 1µL of 1 ng/µL stock 'pGNC-6xHis-Gb1-Amsh2 level 1' template DNA, 10µL 5x Q5 high GC enhancer, 0.5µL Q5 high-fidelity DNA polymerase (0.01U), and 22.5µL nuclease-free water. Cycling conditions used were initial denaturation at 98°C for 30s, followed by 30 cycles of denaturation at 98°C for 10s, annealing at 60°C for 30s, and extension at 72°C for 10s, followed by a final extension at 72°C for 2 mins. The expected product size was 195bp. 5µL of each product was analysed by electrophoresis on a 1% agarose gel in 1x TAE. The remaining 45µL of each product were combined for purification by QIAquick PCR purification kit (QIAGEN, #28104). The concentration and purity of the final product were analysed by Nanodrop (ThermoFischer).

For Gibson assembly, pET28a-MpRV-Cas9-His was cut using EcoRI (ThermoFischer, #FD0274). 5µL of the product was analysed by gel electrophoresis before the rest was subject to purification by QIAquick PCR purification kit (QIAGEN, #28104). The concentration and purity of the final product were analysed by Nanodrop (ThermoFischer). The Gibson assembly reaction was carried out using the NEBuilder HiFi DNA assembly cloning kit (NEB, #E5520S). A high excess of insert to vector was required. To achieve this 8µL of linearised pET28a-MpRV-Cas9-His at 17.4 ng/µL, and 2µL of the Gb1-3C insert at 48.1 ng/µL were added to the reaction. This gave a molar excess of insert to vector of 67.9:1. The DNA was mixed with 10µL 2x NEBuilder HiFi DNA Assembly Master Mix and incubated at 50°C for 30 mins. The product was then transformed into

DH5 α *E. coli* (NEB, #C2987, provided in the NEBuilder HiFi DNA assembly cloning kit) by heatshock according to the manufacturers protocol.

Colony PCR was carried out using primers consisting of the annealing sections of the Gb1-3C Gibson insert primers, using the same PCR conditions used to generate the insert. For this, 8 colonies were picked at random and named A-H. The same primers were run on 1 ng pGNC-6xHis-Gb1-Amsh2 level 1 as a positive control. The primers used were ordered from Merck and are listed in Table 4.3. 5 μ L of each product was analysed by gel electrophoresis. The remaining PCR product from colonies F, G and H, who's bands matched the positive control, were purified by QIAquick PCR purification kit (QIAGEN, #28104) and sequenced by sanger sequencing (Eurofins). To check the cloning was successful, the returned sequences were aligned to the Gb1-3C sequence of pICSL30028 (TSL SynBio), which is the plasmid from which the Gb1-3C sequence used in pGNC-6xHis-Gb1-Amsh2 level 1 originated.

Table 4.2: Primers used to generate an insert of Gb1-3C for Gibson assembly into the *EcoRI* cut site of pET28a-MpRV-Cas9-His. The parts of the primers overlapping with the vector sequences flanking the *EcoRI* cut site are lowercase. An additional base to maintain framing is underlined in the reverse primer.

Primer ID	Sequence
Gb1-3C_GIB_F	cagcaaattgggtcgcg gatccgaaTACAACTGATCCTGAACGG
Gb1-3C_GIB_R	gcgccgtcgacggagctcgaATGGGCCCTGAAACAGAACTTCC

Table 4.3: Primers used in colony PCR to analyse the product of Gibson assembly to insert Gb1-3C into pET28a-MpRV-Cas9-His.

Primer ID	Sequence
Gb1-3C_F	TACAAACTGATCCTGAACGG
Gb1-3C_R	TGGGCCCTGAAACAGAACTTCC

4.4.2 Small scale expression and purification of MpRV-Cas9

To check that our constructs expressed MpRV-Cas9 successfully, small scale protein expression and purification was carried out. This method was used multiple times for optimisation purposes. Initially, all constructs were transformed by heat shock into BL21 (DE3) (Merck, #2572) *E. coli* for expression. All were subject to the same pET28a vector expression protocol designed for induction of expression by addition of IPTG. Cultures of the BL21 cells containing the construct were grown overnight in 10 mL LB with kanamycin in a 37°C shaking incubator. The following morning, 4 mL of this culture was added to 100 mL LB with 50 µg/mL kanamycin. This culture was incubated, shaking at 37°C, until an OD600 value of 0.6 - 0.8 was reached. At this stage 2 µL of the culture was taken as an uninduced control. Then the culture was cooled in a cold room for 20 minutes before addition of IPTG, followed by overnight shaking incubation at 18°C. The following morning 2 µL of this culture was taken for comparison to the uninduced control. The remaining culture was centrifuged at 4000 g, and the pellet was kept at -20°C for purification.

For protein purification the pellet was resuspended in 2 mL IMAC buffer A1 with 2 µL 100 mM PMSF protease inhibitor and left on ice for ~30 mins. Then cells were lysed by

sonication for a total of 20 cycles of 5 seconds on and 5 seconds off, after which 2 μ L was set aside as a total protein sample. After this the sample was centrifuged at 4°C at max speed (~17,900 g) for 20 minutes, after which 2 μ L supernatant was taken aside as a soluble fraction sample, and 2 μ L resuspended pellet was taken aside as an insoluble fraction sample. The remaining steps were carried out in a 4°C cold room. The supernatant was applied to 100 μ L Ni⁺ agarose beads (Ni-NTA His-Bind Resin, Merck, #70666-10 ML) and allowed to bind with slow rotation for ~3 hours. After this the beads were centrifuged at 300 g for 1 minute and 2 μ L of the supernatant was taken as an unbound sample. The supernatant was removed and 500 μ L IMAC buffer A1 was added as a wash buffer, and this mixture was then centrifuged at 300 g for 1 minute, after which the supernatant was carefully removed. This was repeated 3 times. Finally, 200 μ L IMAC buffer B1 (recipe as for A1 but with 0.5 M imidazole) was applied to the beads, which were then centrifuged at 300 g for 1 minute. The supernatant was then taken which contains the purified protein sample. All the collected samples were subject to SDS-PAGE using precast gels (NuPAGE 4-12% Bis-Tris, Invitrogen, #NP0321BOX). The samples were mixed with 4x NuPAGE LDS sample buffer (Invitrogen, NP0007), buffer A1 and IDT, boiled for 5 minutes for denaturing, and loaded onto the gels, in tanks containing NuPAGE MOPS SDS running buffer (Invitrogen, #NP0001). Also loaded was colour protein standard size marker (NEB, #P7719S). The proteins were separated at 100 volts for 10 minutes followed by 200 volts for ~20 minutes. Gels were then moved to square plates and stained with ReadyBlue Protein Gel Stain (Sigma-Aldrich, #RSB-1 L) via shaking for ~1 hour. This was then washed off with distilled water. Gels could then be imaged by scanning.

4.4.3 Large scale expression and purification of P2C-Cas9

Large scale expression and purification of P2C-Cas9 was carried out similarly to MpRV-mCherry, described in Chapter 3, with a few amendments.

The plasmid pET28a-P2C-g4sx3-Cas9 was constructed and gifted by Ian Bennet and Grant Hughes at Liverpool School of Tropical Medicine. After transformation into BL21(DE3) *E. coli* (NEB, #2572), a single colony was used to inoculate 6 100 mL starter cultures in TB with 9.4 g/L K_2HPO_4 , 2.2 g/L KH_2PO_4 , and 50 μ g/mL kanamycin, which were grown overnight shaking at 37°C. This growth medium recipe is taken from the purification procedure of BtKV-Cas9 (49). From these cultures, 40 mL was used to inoculate 6 1 L cultures of the same TB recipe, which were grown to OD600 0.6-0.8 by shaking incubation at 37°C. The cultures were then kept in the fridge (at ~4°C) for at least 20 mins before expression was induced by addition of IPTG to a working concentration of 1 mM. The cultures were then incubated shaking at 18°C overnight. The next morning, the cultures were centrifuged at 24,000 g for 30 mins to collect the pellet. The supernatant was removed, and each pellet was resuspended in 25 mL IMAC buffer A1. These resuspensions were combined to form 2 150 mL suspensions, to which 3 complete, EDTA-free protease inhibitor cocktail tablets (Roche, #11873580001) were added. Sonication, IMAC and gel filtration were carried out as described for MpRV-mCherry purification in Chapter 3. Samples from each of the wells on the elution plate that matched a peak on the AKTA Express system trace were analysed by SDA-PAGE. A 2nd IMAC and gel filtration was run to further clean up the elution. The wells corresponding

to successful purifications as shown by the SDS-Page were combined and concentrated using 50 kDa gated vivispin columns to 1.48 mg/mL as measured by Qubit.

4.4.4 Large scale expression and purification of MpRV-Cas9

Large scale expression of MpRV-Cas9 was done similarly to P2C-Cas9 large scale expression but scaled up to 8 L of expression culture. Further, 3C cleavage was carried out, after initial IMAC and gel filtration, to remove the His-Gb1 solubility tag at the N-terminus. This was carried out using 3C enzyme prepared previously by Matter Gravino. 100µg of 3C enzyme was added to the His-Gb1-3C-MpRV-Cas9-His and incubated at 4°C for ~18 hours. The next day, a second IMAC and gel filtration using AKTA express was carried out. The two peaks observed corresponded to the His-Gb1 and MpRV-Cas9-His. Samples were taken from the wells in the elution plate corresponding to these peaks and analysed by SDS-PAGE. Wells corresponding to MpRV-Cas9 were combined and concentrated using 50 kDa vivispin columns to 2.48 mg/mL as measured by Qubit.

4.4.5 Identification of an *M. persicae white* homologue as a target gene

A *M. persicae white* homologue was found via BLASTp on the HPC using the *B. tabaci* Protein White sequence (XP_018908689.1), used in the ReMOT control study as a marker (49), as a query against *M. persicae* clone 'O' annotation v2.1 (154). From the sequence identified (MYZPE13164_O_Elv2.1_0184430.1), the corresponding gene sequence was

found. The general code used for BLAST in the command line is in the appendix (General code for command line BLAST).

4.4.6 Identification and generation of sgRNAs targeting *white*

To identify sgRNA target sites, I first analysed the Protein White peptide sequence (MYZPE13164_O_Elv2.1_0184430.1) using InterProScan (194). This revealed the ABC transporter-like region (region 66-316), the AAA+ ATPase region (region 99-293) and the ABC transporter family G domain (region 273-371). The conserved Walker A/P-loop and ABC transporter signature motifs (GSSGAGKT and LSGGEMKRLS respectively) were also revealed. These regions were then matched up to their coding sequence in the *white* coding sequence. sgRNAs were designed in CLC Workbench 21 (QIAGEN) by searching for 'GNNNN NNNNN NNNNN NNNNN NGG'. Suitable sgRNAs that targeted key domains of the protein but did not overlap exons were selected. These sequences were then subject to BLASTn in the command line against the coding sequence database constructed from *M. persicae* O annotation v2.1, to search for the possibility of off-target effects. sgRNAs were ordered from Merck and are presented in Table 4.4.

Table 4.4: sgRNAs designed against *M. persicae* *white* coding sequence.

sgRNA	Start position on <i>white</i> CDS	Sequence
1	328	GGTCGTTTTGCCGGCGCCACTGG
2	631	GGCAGAATTAAAGGTTTATCTGG
3	1318	GACGGTGTGATGAACATCAACGG

4.4.7 Genomic DNA extraction from *M. persicae*

Template gDNA was extracted from *M. persicae* by grinding aphids in 250 µL extraction buffer (10 mM Tris HCl pH 7.0, 1 mM EDTA, 0.4 M NaCl), centrifuging at 17,900 g for 30 mins, and moving the supernatant to a new tube. To this 250 µL isopropanol was added to precipitate the DNA. The sample was centrifuged at 17,900 g for 30 mins, and the supernatant removed. The pellet was then washed in 75% ethanol, centrifuged at 17,900 g for 30 mins, and resuspended in 50 µL water. The concentration and purity were checked on a nanodrop.

4.4.8 PCR of *white* target fragments

Primers PWEX3F and PWEX9R were used to amplify a fragment containing all the sgRNA target sites. VeriFi (PCR biosystems) was used as the polymerase as it is a high-fidelity polymerase and sequence sensitivity is important for this experiment. PCR cycling conditions were based on the manufacturers protocol. A gradient PCR was first conducted to optimise annealing temperature. This revealed an annealing temperature of 53°C to be the most effective. Therefore, the final PCR conditions were 95°C for 1 min, followed by 40 cycles of 95°C for 15s, 53°C for 15s, 72°C for 3 mins, followed by a final extension of 2 mins 72°C. The product was subject to purification using the QIAGEN PCR purification kit. The final concentration of DNA recovered was 12.6 ng/µL as measured using a Nanodrop. This fragment was then named SgT2.

This amplicon was suitable for testing sgRNA2 but was unsuccessful in testing sgRNAs 1 and 2. Therefore, new primers were designed to amplify products containing the targets for sgRNA1 and 3. These primers were first subject to a gradient PCR, from which the annealing temperature 55°C was selected. The final PCR and purification were carried out as described for SgT2, but with an annealing temperature of 55°C. The products containing the targets of sgRNA1 and 3 were named SgT1 and SgT3 respectively. All the primers used were ordered from Merck and are listed in table 4.5. The final concentrations, measured by nanodrop, of SgT1 and SgT3 were 52.3 ng/μL and 36.0 ng/μL respectively.

Table 4.5: Primers used to generate fragments containing sgRNA1, 2 and 3 targets for use in in vitro cutting assays.

Primer ID	Sequence	Fragment produced
sgRNA1_targ_F	ATACTGCACACCAAAAATCG	SgT1
sgRNA1_targ_R	TTCCACATCAAACCTCCAGTA	SgT1
PWEX3F	CTTTCTCTCTCTCTAATAGA	SgT2
PWEX3R	GATTGGGAGGCTATAAGCTG	SgT2
sgRNA3_targ_F	GTGTTAGTTCTTTGCAAGTG	SgT3
sgRNA3_targ_R	TTGCTCAGGAAGTACACAT	SgT3

4.4.9 In vitro cutting assays

To assess whether sgRNAs could guide Cas9 to the target sequence, and to test whether the fusion proteins retained their nuclease ability, *in vitro* nuclease assays were conducted. All assays were performed using SpCas9 (NEB, #M0386) as a positive control and NEBuffer 3.1 as the cutting buffer. NEB's *in vitro* digestion with Cas9 protocol (195) was used as a guide. Briefly, SpCas9, MpRV- and P2C-Cas9 fusions and individual sgRNAs were incubated at room temperature in NEBuffer 3.1 and water to form RNPs for 10 mins at 25°C. The final working concentrations were 30 nM sgRNA and Cas9. Then 5 µL substrate DNA (SgT1, 2 or 3) was added, and the mixture was incubated at 37°C for 20 mins to allow cleavage to occur. Finally, 1 µL proteinase K (20 mg/mL, ThermoFischer, #EO0491) was added to degrade the proteins for 10 mins at room temperature. Gel electrophoresis on a 1% agarose gel in 1X TAE was used to analyse the results of the reaction.

5 Gene editing of *Myzus persicae* by ReMOT control

5.1 Introduction

Gene editing by CRISPR/Cas systems has revolutionised genetic studies in many fields. The first CRISPR/Cas-edited insect was the model organism *Drosophila melanogaster* (Diptera) (80, 81). Gratz et al; 2013 co-injected pre-blastoderm embryos of the flies with plasmids encoding Cas9 and an sgRNA against exon 1 of the *yellow* gene, causing a loss of function mutation (80). Further, specific insertions into this gene were obtained via injecting embryos with a homology directed repair (HDR) template, along with the Cas9 and sgRNAs (80, 81).

The most common method of choice for CRISPR/Cas transformations of insect lines is microinjection eggs containing preblastodermal embryos (26, 80, 81, 86, 87, 91, 92, 196). Unfortunately, this method of delivery is more challenging to achieve in aphids due to their predominantly asexual and viviparous reproduction habits. Aphids reproduce predominantly asexually, with the odd cycle of sexual morphs arising in the autumn, often in response to short photoperiodic and colder climate conditions (25). After mating, the sexual females lay overwintering eggs, which undergo a diapause period, before hatching in the spring (14, 25, 28, 197). The current method for CRISPR/Cas gene editing in aphids involves embryo injection (shortly after sexual morphs lay eggs) in the pea aphid, *Acyrtosiphon pisum*, and takes 7 months to complete (26). This is due to the time taken to obtain synchronous sexual male and female colonies, and the ~83-day

diapause the eggs must undergo (26). *M. persicae* produces sexual males and females under short photoperiods, with females laying eggs; however, obtaining many eggs remains challenging. Additionally, the eggs may require deposition on peach leaves for optimal survival and hatching. Furthermore, *M. persicae* appears to be susceptible to inbreeding depression, which can significantly reduce the survival rate of both eggs and hatchlings. Therefore, an alternative approach is needed to achieve effective CRISPR/Cas gene editing in *M. persicae*.

ReMOT control may be such an alternative methods, because it involves CRISPR/Cas delivery into females that carry the eggs/embryos. Therefore, such may be deployed to edit the female asexual aphids that carry various stages of embryos before viviparous birth. Gene editing by ReMOT control has been successful in many insects, and one arachnid (the black legged tick) (49, 89, 90, 94-98, 100, 101). This system relies on injection of the mother insect with a chimeric embryo-localising Cas9 ribonucleoprotein (RNP), made up of the Cas9 protein and an embryo localising peptide derived from yolk precursor proteins (YPPs), complexed with sgRNAs targeting a gene of choice (49, 89, 90, 94-98, 100, 101). In previous chapters of this thesis, it has been described that MpRV, derived from *M. persicae* vitellogenin (MpVg), and P2C derived from DmYP1 (90), act as embryo localising peptides in *M. persicae*, and enable uptake of cargo into oocytes. Further, this work demonstrated that chimeric MpRV- and P2C-Cas9 cleave amplicons from the *M. persicae* *white* gene *in vitro* when complexed with sgRNAs targeting 3 regions of the gene.

The efficiency of editing via ReMOT control may be improved by including an endosome escape reagent (EER) into the injection mixtures (49, 89, 90, 94-98, 100, 101). Saponin is one such EER. Various plants produce saponins as a defence against herbivory (198). The amphipathic properties of saponins allow them to penetrate membranes, achieved by complexing with sterols to form pores (198). Consequently, these proteins are effective as EERs in ReMOT control injection mixtures, facilitating RNP escape by forming pores in the endosomal membrane. Notably, they have successfully supported ReMOT control in species such as *N. vitripennis*, *A. aegypti*, *C. pipiens pallens*, *A. stephensi*, *I. scapularis*, and *B. tabaci* (49, 89, 90, 94, 96, 101). In most of these cases, saponin increased editing efficiency. However, it was reported that it had little effect on the efficiency of REMOT-mediated gene editing in *B. tabaci*, and above a concentration of 8ug/mL, it was inhibitory (49). The same properties that make saponin an effective EER also contribute to its potency as an insecticide. Indeed, multiple saponins reduce survival and reproduction in *A. pisum* (199-202).

It is crucial to rigorously screen the offspring of mothers injected with ReMOT control mixtures for potential gene-editing events. Traditionally, this has been achieved through a combination of phenotypic and genotypic analyses (49, 89, 90, 94-98, 100, 101). To evaluate the effectiveness of CRISPR/Cas ReMOT editing and facilitate further optimization, target genes are selected based on their ability to produce observable phenotypes upon loss of function. For example, changes in eye colour serve as a readily identifiable phenotype (49, 89, 90, 94-98, 100, 101). We have selected the *M. persicae white* gene for this purpose. As such, in a previous chapter we have optimized the design of sgRNAs targeting three regions of this gene. The *white* gene is a marker gene that was

initially characterised in *D. melanogaster* (191). Loss of function mutations in *white* have been shown to give rise to eye colour changes in multiple insects, including when induced by gene editing via ReMOT control (49, 98, 99). Since the *M. persicae* white homolog has not yet been studied in the context of loss-of-function mutations, there is a risk that its disruption may not result in an observable phenotype in the aphid. Additionally, somatic cell editing or heterozygous mutations of this gene are unlikely to produce observable phenotypes. To address these challenges, genotypic screening methods will need to be employed alongside phenotypic analysis. In previous ReMOT control studies, genotyping was conducted via gene-specific PCR followed by Sanger or next generation sequencing (49, 89, 90, 94-98, 100, 101).

The goals of this chapter are to (i) optimize methods to inject adult *M. persicae* females with REMOT control constructs; (ii) setup aphid rearing conditions to optimize the chance to detect changes in eye colour of progeny born from injected mothers, (iv) establish genotyping methods, and (v) use the optimized methods to obtain gene-edited *M. persicae*.

5.1.1 Contributions to this chapter

I conducted most of the work described in this chapter. Dr. Sam Mugford and Dr. Mar Marzo aided in *M. persicae* injections, and Dr. Marzo also aided in phenotypic and genotypic screening of offspring. Dr. Ioanna Morianou aided in genotyping procedure design by recommending the ICE programme by Synthego. The *white* SNP identification analysis across samples from 193 *M. persicae* populations was performed by Thomas Heaven.

5.2 Results

5.2.1 Selection and optimization of aphid injection method

Aphid injections were optimised from the procedures described in Chapter 3 where aphids were injected with mCherry chimeras. Micropipette preparation was not changed as borosilicate needles had been suitable for injections of mCherry chimeras. Micropipettes were pulled on a Sutter P-97. The programme used was also the same: heat=500; pull=150; velocity=90; delay=235. The NANOLITRE 2020 (WPI) was used over the nanoject II (Drummond) for finer control of injection volumes and rate. During injections described in Chapter 3, the method of aphid immobilisation was switched to cold treatment from a vacuum pump across a small hole. By continuing the use of cold treatment to anaesthetise the aphids, we were able to inject adult females at a good pace, with a good survival rate, which produced many offspring. Adult females were injected in the abdomen on the dorsal side, close to the location of growing embryos.

5.2.2 Optimization of rearing conditions to enhance likelihood of capturing edited individuals

Aphids were age matched to 7-days old on *Brassica rapa* prior to injection. It has been shown that on *Brassica campestris*, preadult development lasts ~7.5 days, and adult longevity is ~17.8 days (203). We housed *M. persicae* on *B. rapa* and assumed a similar timescale. Therefore, by injecting 7-day old females, we allowed the maximum possible time for post-injection reproduction.

Over the course of injection experiments, the conditions for housing and rearing progeny were optimised. Initially, aphid nymphs were screened for a phenotypic change approximately every 2 days post-injection. During this process, the 2-day old nymphs were moved away from the mothers with a fine paintbrush. This approach was taken so that clip cages did not become overcrowded with offspring over a longer period. However, we found that nymphs could be damaged by the process of moving them from clip cages too soon. Therefore, in later experiments, approximately every 2 days post-injection, the injected mothers were moved to a new clip-cage on a new *B. rapa* plant. The progeny were left in the original clip-cage and allowed to mature for a further 2 days before inspection. This allowed us to ensure that any potentially edited nymphs would not be easily damaged when screening for a phenotypic change in eye-colour.

5.2.3 Selection of genotyping procedure

We optimised a genotyping procedure to check injected aphid progeny for successful gene editing events. Due to the high number of aphid progeny, and the cost of sequencing, it is infeasible to screen progeny individually. Consequently, pooled samples of progeny were screened. Under advice from Dr. Ioanna Morianou (The Sainsbury Laboratory), pooled samples of progeny were first subject to Sanger sequencing (Azenta). The results were inputted into the inference of CRISPR edits (ICE) tool provided by Synthego. This programme analyses Sanger sequencing results to highlight presence or absence of potential edits. Any samples highlighted to contain potential edits were then subjected to amplicon sequencing by Illumina short read sequencing (Azenta). As part of this service, reads were aligned and analysed for insertions and deletions (indels).

5.2.4 Offspring of *M. persicae* females injected with MpRV-Cas9-sgRNA RNPs are unlikely to have edited target sites

Having optimized various key procedures, we proceeded with injecting a total of 255 aphids with the combined mixture of MpRV-Cas9 and all three sgRNAs, and 8 µg/mL saponin included in 155 of them. In each experiment ~60 aphids were injected with MpRV-Cas9, with ~30-50 surviving beyond 48 hours (barring cases of low survival due to improper plant care) and readily producing progeny. Screening of 2524 progeny revealed no obvious changes in eye colour. A total of 21 pooled samples – 7 of which from

Saponin-inclusive injections – were genotypically screened at all three target sites by Sanger sequencing. 14 pooled samples (no saponin) were sequenced in both forwards and reverse directions, while 7 (with saponin) were sequenced only in the forward direction. All the results from 105 total reactions were subject to analysis by ICE. 45 reactions returned errors due to poor sequencing quality and/or issues with guide sequence detection. All the 60 samples that were successfully sequenced gave outputs like the example shown in Figure 5.2. The sequences showed single peaks at each nucleotide around the site targeted by the sgRNAs (Figure 5.2A) and the discordance ratios around the targeted site were consistently low (Figure 5. 2B). Moreover, analysis of the possibility of indels at the target site revealed a model fit (R^2) score of ~ 100.0 (Figure 5.2C). A summary table of all these analyses can be found in the appendix (Table S27). Altogether, these data indicate that sites targeted by the sgRNAs did not result in nucleotide changes or indels in the pooled progeny from injected mothers.

MpRV-Cas9 sgRNA1 Injected 12/4/24 - Offspring collected 14/4/24

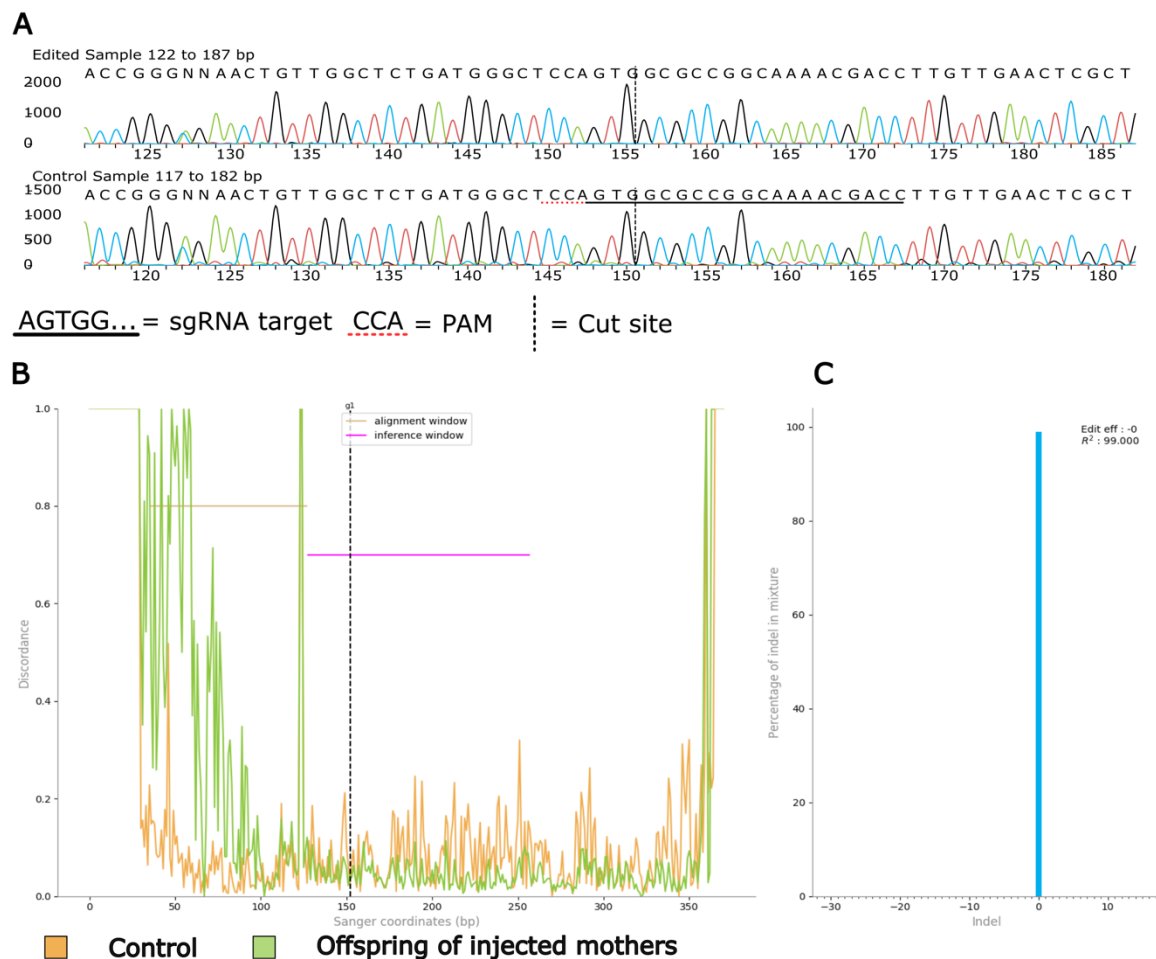


Figure 5.1: An example of ICE analysis results for offspring of mother injected with MpRV-Cas9 RNPs. In this case, mothers were injected on 12/4/24 with MpRV-Cas9 complexed with sgRNAs 1, 2 and 3. Offspring were collected and pooled on 14/4/24. These results are for sgRNA target 1. (A) The sanger sequencing trace of the test sample compared to an un-injected negative control. (B) the discordance between the test sample (green) and the negative control (orange) across the amplicon. (C) The percentage of reads containing indels. In this case 100% of reads contain no indels. The model fit score (R^2) of 99.000 and the editing efficiency of 0 are also displayed.

5.2.5 Evidence of target site editing in offspring of *M. persicae* females injected with P2C-Cas9-sgRNA RNPs

Next, we tested if the P2C-Cas9 protein may induce edits. To this end, a total of 320 aphids were injected with a mixture of P2C-Cas9 and three sgRNAs, and 8 µg/mL saponin was included in 80 of the aphids. Encouragingly, two individuals exhibited a ‘patchy eye’ phenotype, with part of one eye being white (Figure 5.3). These two aphids were found on the same plant, indicating that they may have been derived from a single injected female. Two aphids with the patchy eye phenotype were moved to new leaves, but unfortunately did not survive so adulthood or produce progeny. When discovered diseased the two aphids were stored frozen at -20°C. Unfortunately, the DNA extraction from these two individuals was unsuccessful, and therefore it was not possible to genotype these individuals to confirm the mutation of the *white* gene. Nonetheless, we proceeded with genotyping of pooled samples via amplifying the target region followed by Sanger sequencing and ICE (Synthego) analyses. One sample of offspring collected on 23/03/24 from mothers injected on 08/03/24 (P2C Group 1) showed low confidence/mixture of peaks for each nucleotide (Figure 5.4A), high discordance scores (Figure 5.4AB) and an indel efficiency score of 22 with the model fit (R^2) of 45.000 near the target site of sgRNA1 (Figure 5.4C). These patterns would be consistent with gene editing events at this site. A summary table of all ICE analysis on progeny of P2C-Cas9 injected aphids is available in the appendix (Table S27).

To further investigate if the P2C Group 1 includes individuals with gene edits in the sgRNA1 target site, the PCR product for this sample was also subjected to amplicon

sequencing via Illumina short-read sequencing. Moreover, to investigate if the poor results with Sanger sequencing is caused by alterations in the primer annealing sequence because of successful editing at the sgRNA1 target site, one sample of 70 pooled offspring collected on 20/03/24 from mothers injected on 08/03/24 (P2C group 2) was also subjected to amplicon Illumina (Azenta) sequencing of the sgRNA1 target region. Finally, as a control, amplicon Illumina (Azenta) sequencing of the sgRNA1 target region derived from a pool of ~50 nymphs from mothers that were not injected, was included. Aligning of the Illumina short read sequence data to the original sequence corresponding to the sgRNA1 target area of the *M. persicae white* gene revealed deletions at multiple positions, including at the cut site. These reads were counted as a percentage relative to the total number of reads successfully aligned. The control sample of non-injected females provided a baseline. This showed that the reads corresponding to mutations/deletions/insertions is less than 0.05% for the region between 205-155 nucleotides and then slowly increases to 0.1% in region between 155 - 129 nucleotides. This region includes the sgRNA1 target area, as well as the PAM and cut site (Figure 5.5 upper panel). There is a slight increase at nucleotide 163 at the cut size, and this coincides with small percentages of reads that have insertions in the two nucleotides 164 and 165 adjacent to the cut size. The most likely interpretation of this result is that there is variation in sequence in nucleotide 164 and 165 among individuals present in the pooled sample, and this is possibly related a higher number reads with deletions at nucleotide 163. Intriguingly, percentages of reads with deletions were profoundly higher in the P2C Group 1 sample, starting at 0.1% at base 205 and rising to 0.2% at base 129. As well, there was a more profound increase in reads with deletions at the cut site of nucleotide 163. There were no reads with insertions observed at nucleotides 164 and

165. The P2C Group 2 sample more resembled the control, but the number reads with deletions were overall higher than the control. These data are most consistent with the interpretation that individuals in the P2C Group 1 sample contain edits in the sgRNA1 region of the white gene, consistent with the Sanger sequence data. The P2C Group 2 sample may also include individuals with edits in this region, but the number of individuals may be lower than that of the P2C Group 1 sample.

In conclusion, detection of individuals with white eyes, and the Sanger and Illumina results indicate that individuals in the P2C Group 1 are likely to carry edits in the sgRNA1 target area of the white gene. This is an important finding, because it provides evidence that P2C-CAS9 protein has activity within aphids and that the REMOT control approach could work for aphids.

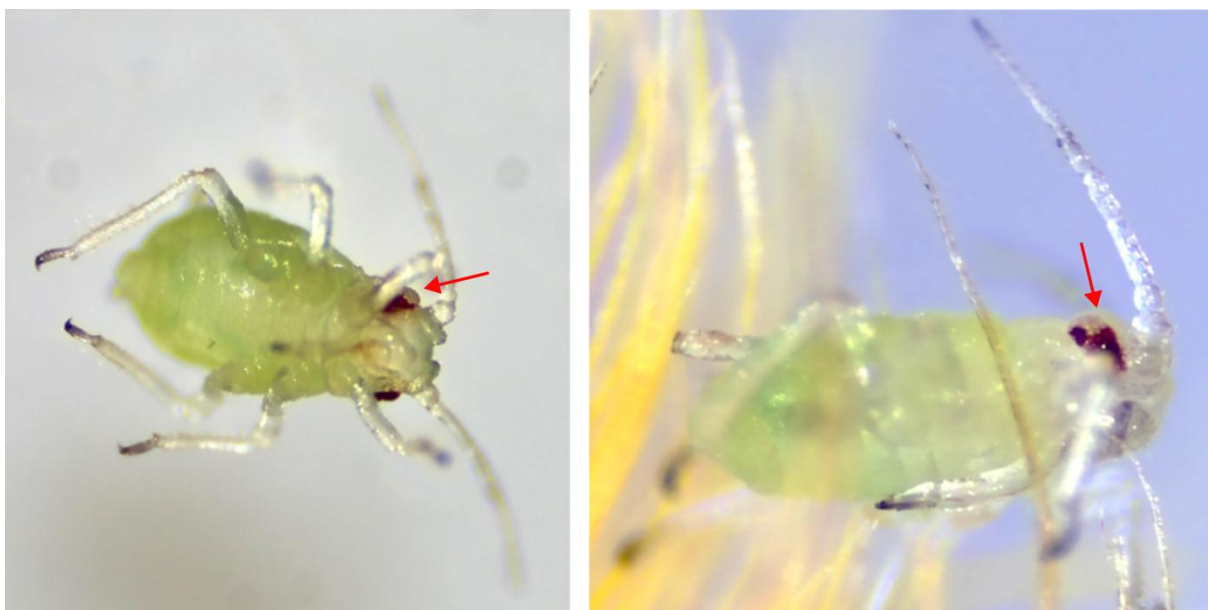
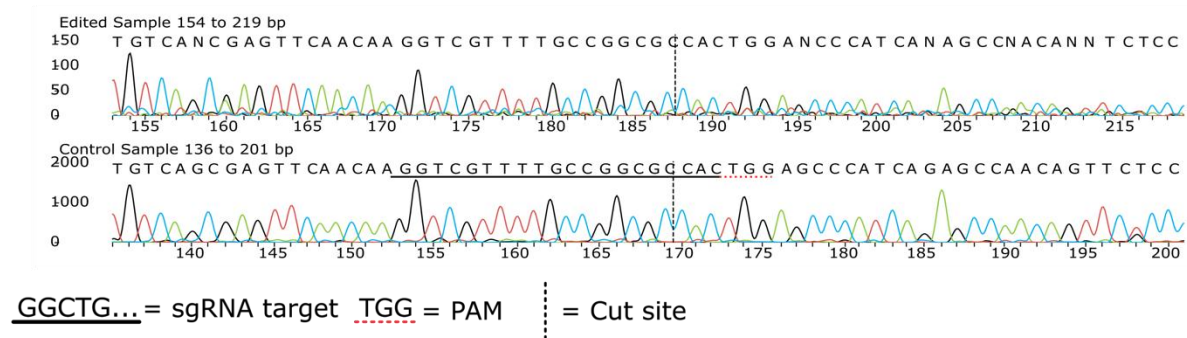


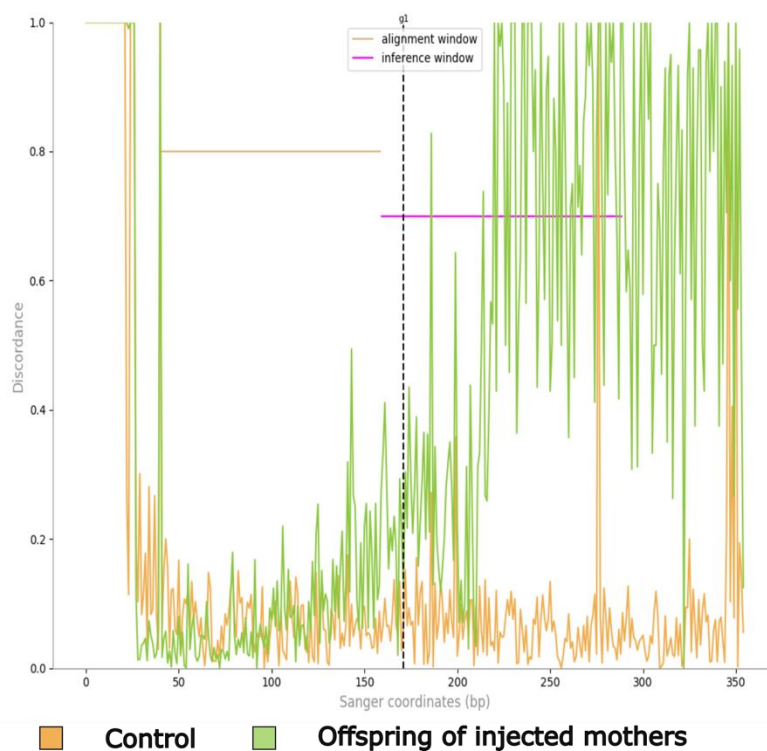
Figure 5.2: Two M. persicae offspring displayed phenotypic evidence of successful genome editing. Both presented a 'patchy eye' phenotype, marked here by red arrows.

P2C group 1 ICE Results

A



B



C

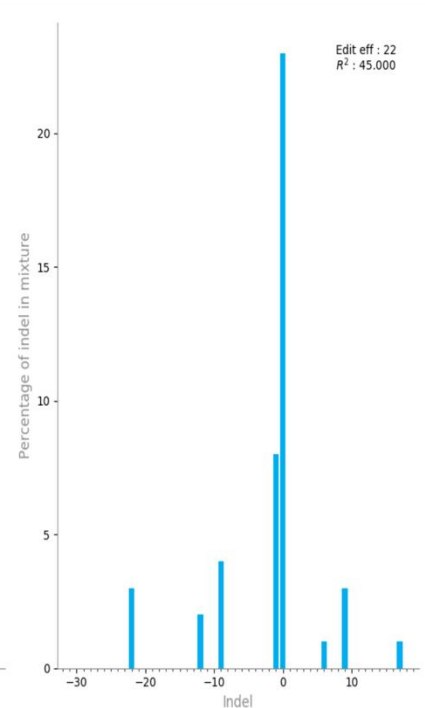


Figure 5.3: ICE results for the reverse sequencing reaction of sgRNA1 target region of offspring of mothers injected with P2C-Cas9 RNPs. Mothers were injected on 08/03/24 and offspring were collected on 23/03/24. (A) The sanger sequencing trace of the potentially edited sample compared to the negative control sample. (B) Discordance in base pairs between the injected sample (green) and the negative control (orange) in the sanger sequencing result. (C) The percentage of the mixture of sequencing reads made up by individual indels. Edit efficiency score of 22 and R² value of 45.000 is displayed. Adapted from results figures provided by Synthego ICE analysis.

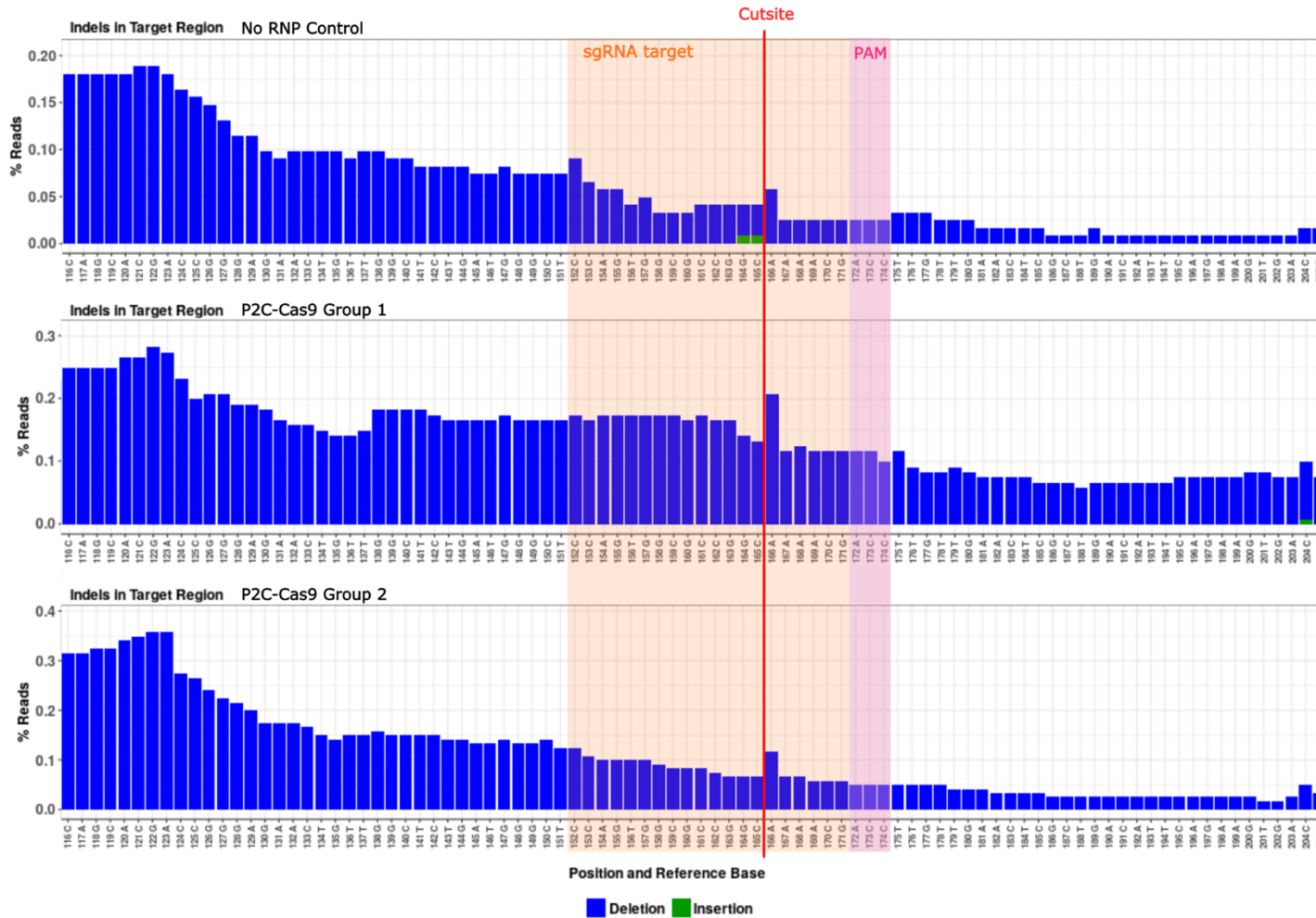


Figure 5.4: Amplicon sequencing results from a negative (un-injected) control sample and two samples injected with P2C-Cas9. The graph shows the percentage of Illumina reads containing a deletion (blue) or an insertion (green) at each base along the sequence. The sgRNA target is coloured orange, the PAM is coloured pink, and the cut site is represented by a red line. The two injected samples show increased deletions compared to the negative control, but the editing efficiency remains low. Adapted from Amplicon-EZ results figured provided by Azenta.

5.2.6 Approach to detect large deletions in the region targeted by sgRNA1, 2 and 3.

We wished to investigate if some individuals may carry large DNA deletions upon P2C-CAS9 editing of the sgRNA1, 2 and 3 target areas in the white gene. Given that the aphids were injected with all three sgRNAs, it is possible that large chunks of DNA are edited out due to double-stranded breaks occurring at multiple target sites. To investigate this possibility, PCRs to amplify regions between each combination of two target sites (i.e. target 1-2 [L1], 2-3 [L2], and 1-3 [L3]) were conducted. Any successful editing that would give rise to large deletions may be detected by running the PCR fragments on electrophoresis gels. Large deletions may result in a PCR product of ~400bp in length, as the primers were designed such that the sgRNA cut sites are ~200bp from either primer. We observed PCR fragments of ~400bp in the L2 combination that would indicate large deletions between sgRNA2 and 3 (Figure 5.6). However, fragments of similar sizes were also detected in non-injected control samples, indicating that the ~400bp may be the result of off-target amplification (Figure 5.6). To investigate this further, the 'L2' samples were submitted to Sanger sequencing. This generated poor sequence data further suggesting that the ~400bp L2 fragment may be derived from off target amplification.

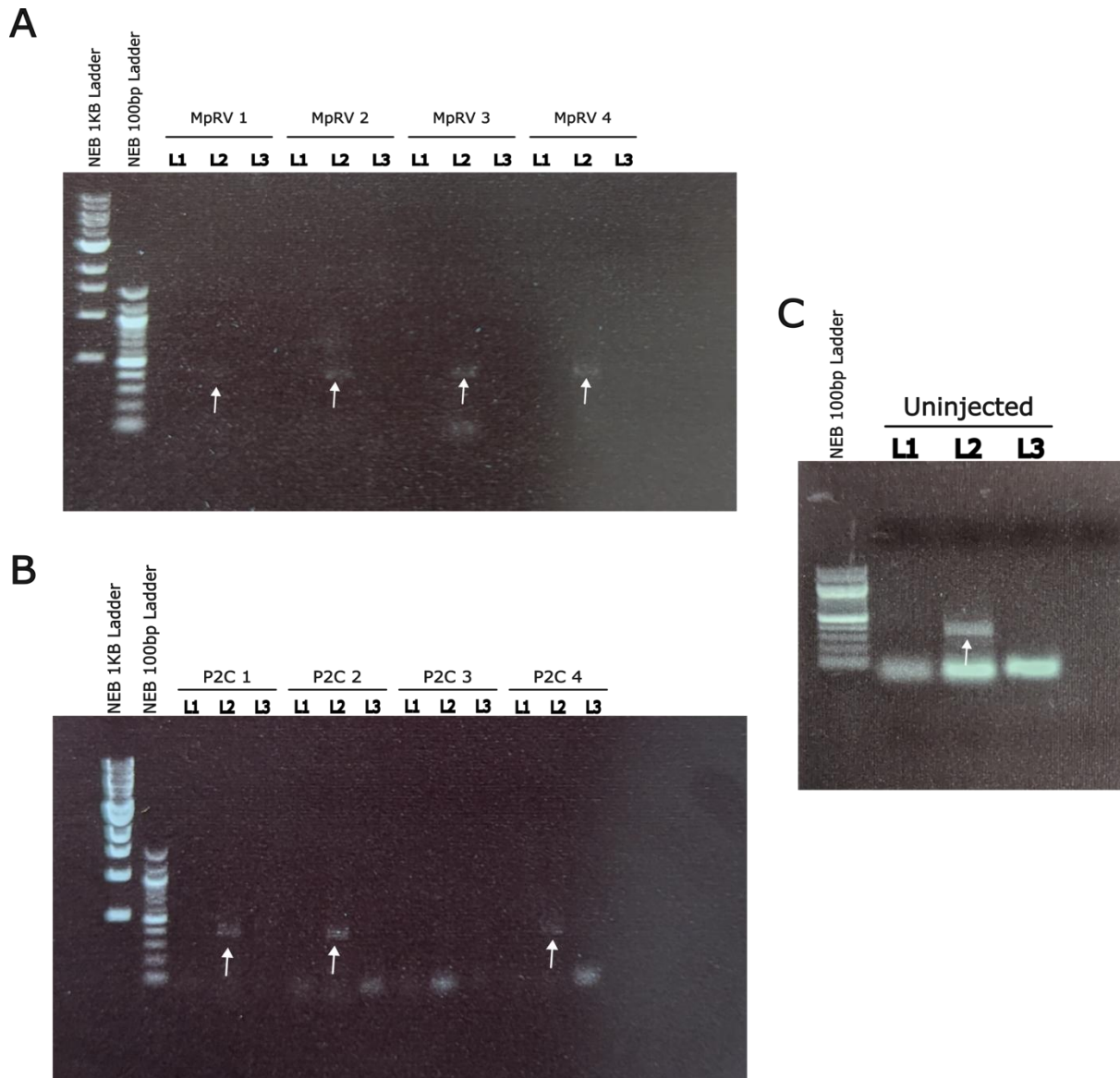


Figure 5.5: Electrophoresis gels showing detection of products when investigating large deletions caused by genomic editing by ReMOT Control. L1, L2, and L3 represent those reactions used to check if a large deletion has occurred between sgRNA1 and 2, 2 and 3, and 1 and 3 respectively. (A) Samples MpRV 1-4 were offspring of mothers injected on 12/04/2024, collected on 18, 21, 26 and 27/04/2024 respectively. Bands are observed at ~400bp in the 'L2' reaction of each sample, marked by white arrows. (B) Samples P2C 1 and 2 represent offspring of mothers injected on 15/03/2024, collected on 27/03 and 03/04/2024. Samples P2C 3 and 4 represent offspring of mothers injected on 03/05/2024, collected on 08 and 12/05/2024. Bands are observed in the 'L2' reaction of P2C 1, 2 and 4 as marked by white arrows. (C) A negative control group of ~50 nymphs from un-injected mothers. The ~400bp band in 'L2' is also present here.

5.2.7 There is a natural SNP in the sgRNA3 target site

Because Illumina sequence data from non-injected control samples indicated that there may be sequence variations among individuals in the sgRNA1 target area within the *white* gene, we analysed existing whole-genome sequence data from 193 *M. persicae* individuals collected worldwide for SNPs in the *white* gene. The Illumina short read sequence data from these 193 individuals were aligned to the high-quality reference assembly of *M. persicae* clone O (154). This analysis was done and provided by Thomas Heaven. The resulting VCF file was visualised against *white* in Geneious prime, facilitating the identification of possible SNPs. There were no obvious SNPs identified within or near the sgRNA1 target region. However, we found a T → C SNP in the base before the sgRNA2 target region, and a C → T SNP in the base immediately after the PAM sequence of this region. Both SNPs are homozygous in many *M. persicae* individuals. There was also a heterozygous C → T SNP in the third base of the sgRNA3 target region in some individuals. Thus, the nature of the sequence variation identified in the sgRNA1 region remains to be investigated.

5.3 Discussion

We have established a ReMOT control system and experimental procedure for use in *M. persicae*. Firstly, I optimized methods to inject adult *M. persicae* females with ReMOT control constructs. I also optimized the logistics for phenotyping aphids for white eyes and developed methods to genotype large numbers of aphids for identifying edits in target regions. Finally, I have evidence that one of the Cas9 fusion proteins resulted in gene edits and may be taken forward to further develop the ReMOT control technology for *M. persicae*.

We established an experimental procedure for ReMOT control in *M. persicae*. By using cold treatment to anaesthetise the aphids, we were able to inject adult females at a good pace, and injected females produced many offspring. Borosilicate micropipettes were suitable for our purposes, but it is possible that quartz micropipettes would further increase survival rate. Indeed, quartz needles were used to inject *B. tabaci* in a previous ReMOT control study (49). By using 7-day old adult females for injection, we were able to ensure a high reproduction rate immediately after injection. It has been shown that on *Brassica campestris*, preadult development lasts ~7.5 days, and adult longevity is ~17.8 days (203). We therefore injected 7-day old females, allowing the maximum possible time for post-injection reproduction.

Saponin has been used as an EER in multiple ReMOT control systems at low enough concentrations to avoid insecticidal effects (49, 89, 90, 94, 96, 101). In *B. tabaci* saponin was shown to have no effect on editing efficiency at 8 µg/mL, and higher concentrations

were found to be inhibitory (49). We found that the use of saponin did not have any observable effect on gene editing efficiency or survival post injection in *M. persicae*. It is possible that our findings show that Saponin has a similar effect on ReMOT control in *M. persicae* as in *B. tabaci*.

Throughout ReMOT control experiments, housing and rearing conditions of progeny were optimised to increase the chances of finding progeny with successful edits. Initially, progeny were removed from clip-cages where they were born approximately every 2 days. Any aphids displaying a phenotypic eye colour change, were moved to individual clip cages. 2 such progeny were identified (Figure 5.3). We believe that the nymphs were hurt in this process, as they did not survive after being moved. The nymphs dried out and this may have contributed to unsuccessful DNA extraction, meaning we could not verify if successful editing had occurred. Consequently, in remaining experiments, injected adults were moved to new clip cages every 2 days, and the offspring were allowed to mature for a further 2 days before phenotypic screening. This reduced the chances of damage during inspection. However, no further offspring displaying a phenotypic eye-colour change were identified. Artificial intelligence (AI) has been used to image and quantify phenotypes in a high-throughput manner. In plants, this has been used to quantify insect feeding damage (204). Further, these tools have been used to track two-spotted spider mite, *Tetranychus urticae*, infestation on *Arabidopsis thaliana* (205). Adaptation of AI based imaging tools for phenotypic screens on insects could enable high-throughput phenotypic screening to identify potentially-edited progeny in future.

By injecting adult female aphids with chimeric P2C-Cas9 RNPs in an injection mixture with NEB r3.1 cutting buffer, we were able to observe 2 offspring exhibiting a ‘patchy eye’ phenotype (Figure 5.3) and recover amplicon sequencing data which may be suggestive of low-level gene editing (Figure 5.5).

Phenotypic evidence of successful editing of the *white* gene was observed in two individuals (Figure 5.3). These individuals were offspring of mothers injected with P2C-Cas9 RNPs, without saponin. We observed a ‘patchy eye’ phenotype in one eye of both individuals. This may be indicative of mosaic/somatic editing, which would have occurred post-blastoderm formation. Unfortunately, the DNA extraction of these individuals was unsuccessful. To this point, no verification of loss-of-function mutations in *M. persicae white* have been characterised, so whether this is indeed a mosaic phenotype remains unclear. In *B. tabaci*, mosaic mutants have a full white eye phenotype (49). In *B. tabaci* eye development occurs at the 4th instar stage (206). When ReMOT control was successfully used to knock out *B. tabaci white*, whitefly females were injected <24h post emergence (49), prior to eye development, hence a mosaic edit could still cause a full white eye phenotype. Aphid eye development occurs from stage 17 of embryogenesis (28). Therefore, if editing occurred during these late stages, it is possible that part of the eye was already developed prior to the introduction of the loss-of-function *white* mutation, potentially causing the ‘patchy eye’ phenotype we observed.

Genotypic screening of *M. persicae* progeny is challenging. Due to the high number of progeny, screening individuals not presenting a visible phenotypic change is expensive. Therefore, we performed genotypic screens on pooled samples of progeny. PCR was

performed on pooled samples of progeny, producing amplicons which spanned each sgRNA target site. Products were first subject to Sanger sequencing, the results of which were analysed by ICE (Synthego) which infers the presence of CRISPR edits from Sanger sequencing traces. This provides a high-throughput, low cost genotyping for pooled samples. Where the tool inferred the presence of indels, samples were subjected to amplicon sequencing by Illumina for verification. This method allows for a wide capture, low resolution check to search for somatic or heterozygous edits in pooled samples of aphids not displaying a visible phenotypic change. Ideally, each individual aphid nymph would be screened in a high-throughput format, but this is expensive and time consuming.

We genotyped pools of offspring from injected mothers by sanger sequencing and analysis by ICE (Synthego). Samples for which ICE returned evidence of editing were subject to amplicon sequencing. This allowed us to check for any edits that were heterozygous, or somatic, and therefore would be unlikely to render a phenotype. ICE analysis did not find evidence of editing in any offspring of MpRV-Cas9 injected mothers. However, it found one pool of offspring of P2C-Cas9 injected mothers to have evidence of editing, but with an unreliable model fit (R^2) score of 45 (Figure 5.4). This was found in the reverse reaction of the target 1 (sgRNA1) region. No evidence of editing in the sgRNA 2 target region was found. ICE was unable to analyse the target 3 region in all cases as it was unable to find the sgRNA target region. We checked natural SNP data from 193 *M. persicae* individuals and identified a SNP in the sgRNA 3 recognition sequence which is heterozygous in some populations. This was likely present in our samples, and likely caused ICE to be unable to identify the sgRNA target sequence. The sample identified by

ICE to potentially contain edits (P2C group 1), one other sample (P2C group 2), and a negative control were subject to amplicon sequencing to check for edits (Figure 5.5). Both samples were derived from pools of offspring from mothers injected with P2C-Cas9 RNPs without saponin, while the negative control was derived from a pool of ~50 un-injected nymphs. A similar pattern was observed across all three, with deletions being detected at many base pairs. A small peak was observed at the cut site of sgRNA 1 (Figure 5.5) in all three samples but was 4 and 2 times more prevalent in P2C group 1 and 2 respectively. This may indicate very low efficiency gene editing, which would likely be somatic, and therefore would not give rise to an observable phenotype. However, we observe a pattern of deletions in the negative control, similar to that in the sample groups, which may suggest that the patterns observed are due to variation in sequencing.

We used 8 µg/mL saponin in some of our injection mixtures. In samples of offspring injected with P2C- or MpRV-Cas9 RNPs and saponin, we detected no evidence of editing at individual sgRNA target sites. We also checked whether large DNA deletions had occurred between multiple cut sites. We observed gel electrophoresis bands indicative of a large deletion between sgRNAs 2 and 3, but this was shown to be a result of off-target amplification by the negative control reaction (Figure 5.6). We were unable to investigate these bands further due to poor quality sanger sequencing. The sequences we did receive did not align to the expected part of the *white* gene, further showing that these bands are a product of off-target amplification in the PCR. These results further show that Saponin is unlikely to be an effective EER for ReMOT control in *M. persicae*.

We have observed little evidence of low efficiency gene editing by ReMOT control in *M. persicae*. Therefore, optimisation of the system is necessary to achieve a reliable, high efficiency gene editing method in asexual *M. persicae*. As discussed, improvements of the injection procedure, such as the use of quartz micropipettes, may increase survival rate post injection, and allow for a more consistent injection volume. Further, exploration of the mechanisms regulating vitellogenesis in *M. persicae* may reveal a better injection timepoint, or pre-injection conditions, which may increase uptake of cargo into early developing embryos. The exploration of alternative EERs may also increase efficiency. Here we have explored the use of Saponin, but others such as chloroquine and monensin, which have been used in previous ReMOT control studies to increase gene editing efficiency (90). Also, an alternative embryo localising tag could be used. In the pursuit of enabling ReMOT control in aphids, it has been shown that cell penetrating peptides (CPPs) such as pentratin (PEN), may be used for this purpose (168). PEN is a 16-residue peptide from the homeodomain of *D. melanogaster* antennapedia. An endogenous *A. pisum* homologue was shown to enable cargo uptake of mVenus to mature embryos (168). Other CRISPR/Cas delivery methods could be explored. By establishing a reliable method of generating sexual *M. persicae* in the lab, and achieving egg hatching, the method of egg injection would become available, which has been successful in the pea aphid *Acyrtosiphon pisum* (26). Another alternative method is direct parental (DIPA) CRISPR (167). This method is like ReMOT control. Preformed RNPs consisting of commercially available SpCas9 and sgRNAs are injected into adult female insects. The RNPs are then taken up to embryos by random uptake via receptor mediated endocytosis (167). This method has been successful in the German cockroach *Blattella germanica*, and the model insect *Tribolium castaneum* (167). However, this method, like

ReMOT control, relies on active vitellogenesis, which increases the rate of receptor mediated endocytosis into embryos (167), so it is likely that similar challenges would be faced in aphids as seen in our work on ReMOT control.

Injection timing is important in ReMOT control studies. In *B. tabaci*, females were injected <24h post emergence to ensure no egg development had begun (49). Aphids have telescoping generations; the final embryonic stage (stage 20) contains developed ovaries before birth (28-30). Therefore, we could not inject *M. persicae* prior to initial embryogenesis. In other cases, the ReMOT control mixture is injected after vitellogenesis-triggering event; in mosquitos, P2C-Cas9 RNPs were injected 24h post blood-meal (90). In mosquitos, synchronous egg development is triggered by a blood meal, as is vitellogenesis (207). Nutrition (amino acids) from feeding activates vitellogenesis through target-of-rapamycin (TOR) signalling (107-112). Aphids constantly reproduce during their asexual reproductive cycle, so a nutritional trigger is lacking. Insulin signalling causes the phosphorylation of forkhead box O (FoxO) which reduces juvenile hormone production (107). This process has been shown to reduce vitellogenesis in *Rhodnius prolixus* (107). Inhibition of insulin like peptides (ILPs) has been implicated in activating the sexual reproductive cycle in aphids upon photoperiod changes (208, 209). In their sexual reproductive cycles, aphids lay eggs, so we can hypothesise that the ILP inhibition induced by photoperiod change leading to sexual reproduction, may also increase vitellogenesis in aphids. Therefore, it is likely that injection time, and the pre-injection conditions, could be further optimised to increase success of gene editing by ReMOT control in *M. persicae*. Finally, as parthenogenic *M. persicae* have telescoping generations (i.e. daughters are born with developed ovaries),

the granddaughters (F2 generation) as well as the daughters (F1) could reveal that ReMOT control-based methods skip a generation in *M. persicae*.

To conclude, we have laid the groundwork for a ReMOT control system for gene editing in *M. persicae*. However, we were unable to achieve a high efficiency of gene editing in *M. persicae* with this system. Without the use of an EER, we have observed some evidence of low efficiency gene editing. Saponin inclusion did not increase this efficiency. These results indicate that further optimisation of this method is necessary.

5.4 Materials and Methods

5.4.1 Establishment of age matched *M. persicae* colonies

For each round of ReMOT control experiments, an age matched colony of *M. persicae* was set up. To achieve this, ~150-200 adult females were taken from the stock colonies and transferred into clip cages of 15 adults on a fresh Chinese cabbage (*Brassica rapa*). These adult females were allowed to reproduce for 24 hours. Then, the adult females were removed. The offspring were left on the plant in clip cages and returned to a controlled environment room (CER). After a further 6 days these offspring would all be 7 days old and ready for injection. Throughout, aphids were kept in a CER under long day (LD) conditions (14h light at 24°C, 10h dark at 20°C).

5.4.2 Micropipette preparation

Micropipettes were prepared using a Sutter P-97 needle puller to pull 1.14mm diameter capillaries (WPI, #504949). The programme used was: heat=500; pull=150; velocity=90; delay=235. Pulled capillaries were kept in a square Petri dish on plasticine or blu-tac until use.

5.4.3 Aphid injection procedure

All aphid injections were carried out the same way on 7-day old adult females. Micropipettes were first backfilled with mineral oil (Sigma-Aldrich, #M5904-5 ML), using a syringe with a long stainless-steel tip. Then, after insertion into the injection apparatus of the World Precision Instruments NANOLITRE2020, 3 μ L of the oil was ejected. A 20 μ L drop of the desired injection mixture was pipetted onto a sheet of parafilm. The tip of the micropipette was inserted into the drop using a hand cranked micromanipulator, and 2.8 μ L of injection mixture was sucked into the front end of the micropipette. Aphids were pre-anesthetised in a Petri dish on ice. For injection, aphids were transferred onto a slide on a pre-chilled block of metal. The mechanical micromanipulator was then used to move the micropipette tip into the abdomen of the aphid, while the aphid was held in place using a small paint brush. At this point, ~60 nL of the injection mixture was injected into the aphid at a rate of 15 nL/s. Once completed, the aphid was carefully removed from the micropipette tip using the paintbrush and transferred to a clip cage. Each clip cage contained a maximum of 15 injected adults. Once full the clip cage was placed on a fresh Chinese cabbage (*B. rapa*) plant, with the aphids on the underside of a leaf. The plant was then housed in a CER under LD conditions.

5.4.4 Formation of MpRV- and P2C-Cas9-sgRNA RNPs

MpRV- and P2C-Cas9-sgRNA ribonucleoprotein complexes (RNPs) were formed in vitro prior to injection into 7-day old adult females. Each mixture contained all three sgRNAs (sgRNA1, 2 and 3), each at a molar ratio of >5:1 sgRNA to Cas9. For MpRV-Cas9 RNPs, 40 μ L of MpRV-Cas9 at 2.48 mg/mL (14.85nmol) was mixed with 5 μ L of each sgRNA

(75nmol each), and 25 μ L nuclease free water (Merck). This mixture was incubated at room temperature for 20 minutes to allow RNP complexes to form. Then the mixture was concentrated to a volume of $\sim$ 40 μ L in a 5kDa gated vivispin column (Sartorius), returning the RNP concentration to $\sim$ 2.48 mg/mL. P2C-Cas9 RNPs were prepared in the same way but were concentrated to $\sim$ 20 μ L to give an RNP concentration of $\sim$ 3 mg/mL.

5.4.5 ReMOT control injection mixture preparation

The final injection mixture contained the preformed RNPs as well as 1x NEB r3.1 cutting buffer. In later experiments, 8 μ g/mL saponin (Sigma Aldrich [Merck], #SAE0073-10G) was also included.

5.4.6 ReMOT control experimental procedure

For ReMOT control experiments the same procedure was followed throughout. In each case an age matched colony of 7-day old adult females was used. For each experiment, 10 7-day old adult females were injected with a control mixture consisting of IMAC buffer A4 and 1x NEB r3.1 cutting buffer. In later experiments, 8 μ g/mL saponin was also included. The remaining 7-day old adult females were injected with ReMOT control injection mixture containing either MpRV- or P2C-Cas9 RNPs with sgRNA 1, 2 and 3. Each aphid was injected with $\sim$ 60 nL of mixture at a rate of 15 nL/s. After injection, the control group and the ReMOT injected groups were housed in clip cages on fresh *B. rapa* in long day conditions. Offspring were collected every 2-4 days for at least the first 7 days, and

then again on day 14 post injection. In later experiments, adults were moved to a new clip cage at each offspring collection time point, and offspring were left in the original clip cage to grow for a further 2 days. This was done to minimise damage to potentially edited offspring when handling them. Offspring were checked for a phenotypic eye colour change under a light microscope. Those displaying any eye colour change were transferred as individuals to new clip cages on fresh *B. rapa* to see if that aphid could produce offspring and if that offspring had the eye colour change trait. They were later collected individually and stored at -20°C for genotyping. The remaining offspring of injected females were pooled for each time-point and stored at -20°C for genotyping. The number of aphids injected with each RNP, dates of injection, and inclusion or exclusion of saponin are presented in Table 1. Summary tables of each individual experiment, including adult and offspring counts, are available in the appendix (MpRV-Cas9 ReMOT control experiment tables and P2C-Cas9 ReMOT control experiment tables)

Table 5.1: The number of aphids injected with P2C- or MpRV-Cas9 in each experiment, with the date of experiment and whether saponin was included in the injection mixture.

Injected RNP	Saponin inclusion	Aphids injected	Date
P2C-Cas9	No	60	15/12/23
P2C-Cas9	No	60	15/02/24
P2C-Cas9	No	60	23/02/24
P2C-Cas9	No	60	08/03/24
P2C-Cas9	Yes	50	15/03/24
P2C-Cas9	Yes	30	03/05/24
MpRV-Cas9	No	60	14/02/24
MpRV-Cas9	No	40	21/02/24
MpRV-Cas9	Yes	60	12/04/24
MpRV-Cas9	Yes	60	01/05/24
MpRV-Cas9	Yes	35	13/06/24

5.4.7 DNA extraction from pooled offspring of *M. persicae* injected with MpRV- or P2C-Cas9

For genotyping, DNA was extracted from pooled samples of offspring from mothers injected with MpRV- or P2C-Cas9. Offspring were pooled by time of collection and kept at -20°C until DNA extraction. For pools of >50 aphids, DNA extraction was performed using the DNA extraction kit PHYTOPURE for plant DNA extraction (Cytiva, #RPN8510). Samples were taken from -20°C and kept on dry ice, before grinding with a steel ball

bearing in a TissueLyser II (QIAGEN) for a total of 1min at a frequency 30 revolutions per second. After this, the manufacturers protocol was followed including the protease step. Samples of <50 aphids DNA were extracted using a DNEasy blood and tissue kit (QIAGEN, #69504). Samples were taken from -20°C and kept on dry ice before grinding on dry ice with a plastic pestle. After this, the manufacturers protocol was followed, including the protease step. Final DNA concentrations were measured by nanodrop (ThermoFischer).

5.4.8 Genotyping of offspring from injected *M. persicae*

Genomic DNA extracted from pooled samples of offspring from injected mothers were subject to genotyping via PCR and DNA sequencing. Initially, PCR products of pooled samples were subject to Sanger sequencing (Azenta) followed by analysis using the ICE analysis tool from Synthego (210). Any samples that gave interesting results through ICE were then subject to amplicon sequencing by Illumina, performed by Azenta. Therefore, amplicons produced needed to be suitable for both ICE analysis, and short-read sequencing by Illumina. To achieve this, primers were designed against the *white* genomic DNA region from the *M. persicae* clone O v2.1 annotation (154) (MYZPE13164_O_Elv2.1_0184430 CDS=266-2269 loc:scaffold_2|63534977-63556269)) to give rise to amplicons of ~400bp with the cut site approximately in the centre. The primers designed were ordered from Merck and are presented in table 2. A total of 6 amplicons were obtained from PCR using these primers. Three amplicons spanned ~400bp over individual sgRNA targets and were called short (S)1, 2 and 3, for which primers ICET1_F and R, ICET2_F and R, and ICET3_F and R were used respectively. A

further three amplicons were obtained to check for large deletions of genomic DNA between two target sites. These were long(L)1, 2 and 3, for which primer pairs ICET1_F and ICET2_R, ICET2_F and ICET3_R, and ICET1_F and ICET3_R were used respectively. PCR for all amplicons was carried out using either Q5 (NEB, #M0491S) or Go Taq G2 (Promega, #M7822) according to the manufacturer's protocols. Gradient PCR was first carried out to find suitable annealing temperatures for PCR with Q5 and Go Taq G2, which were 55°C and 52°C respectively. Final genotyping PCRs were carried out on 96-well plates with a 25 µL reaction in each well, including a no-template negative control for each primer pair. Cycling conditions were those of the manufacturers protocol with the annealing temperatures of 55°C for Q5, and 52°C for Go Taq G2, and extension at 72°C for 1min 30s. Products were purified either by QIAquick PCR purification kit (QIAGEN, #28104) or by Azenta as part of the 'Predefined' Vario Plate Sanger sequencing service. Products were sent for sanger sequencing by Azenta. The results were then subject to analysis by ICE to detect potential gene editing occurrences. Any product samples which showed potential gene edits in the ICE results were then sent to Azenta for amplicon sequencing by short-read Illumina sequencing. Analysis of results for gene editing in our target locations was also done by Azenta as part of this service.

Table 5.2: Primers used for genotyping offspring of adult female *M. persicae* injected with MpRV- or P2C-Cas9. Primers were designed to amplify regions of *white* genomic DNA containing each sgRNA target site.

Primer ID	Sequence
ICET1_F	TCTCTCTAATAGATACTACAACCACG
ICET1_R	CTTTGACCCATATTCCTATTTATTCTCA
ICET2_F	CACATACCGTACGAAAAACGAATG
ICET2_R	TTTGACTAGCTTTATATAATAACAGTA
ICET3_F	GTCACCGTACAAAGCAACATGGAT
ICET3_R	CATCACCTCCGATATGCACCAC

5.4.9 Analysis of SNPs in sgRNA target sites

SNP data from 193 individual samples from different *M. persicae* populations was provided by Thomas Heaven (Hogenhout lab). Most of these samples were originally collected and sequenced by Roland Wouters in the Hogenhout lab, while others were collected and sequenced by the Bass lab (University of Exeter) (211). This data was mapped to the *white* gene in Geneious Prime. The sgRNA target sites were then checked for SNPs.

6 Investigating Cas ϕ as an alternative to Cas9 in ReMOT control in *Bemisia tabaci*

6.1 Introduction

Gene editing in insects by CRISPR systems is commonly achieved through the use Cas9 (91, 212, 213). However, many other Cas proteins have been discovered and characterised. Cas12s are a group of nucleases which are part of the Class 2 Type II Cas enzymes (65). Cas12a (also sometimes known as Cpf1) is the best characterised of these, and acts through a single RuvC-like domain which cleaves both DNA strands in a staggered cut (70). Further, the guide RNA Cas12a requires is shorter, and consists of a single RNA, rather than the dual RNA system consisting of crRNA and tracrRNA required by Cas9 (58, 70). Finally, the PAM sequences recognised by Cas12a is a 5' T-rich PAM, rather than a 3' G-rich PAM for Cas9, opening options for a wide range of targets in AT rich organisms, such as aphids and whiteflies (58, 70). Beyond Cas12a, there are many other Cas12 enzymes. Recently, 'mini' Cas proteins in the Cas12 family have been discovered (71, 72). Cas ϕ (also known as Cas12j) is a hypercompact genome editor discovered in 'Biggiephages', which is about half the size of Cas9 and Cas12a (71). Like Cas12a, Cas ϕ cuts DNA via a single RuvC domain, and relies on a short single guide RNAs (71). The PAM sequence required is 5'-TBN-3' where B is either T, C or G (71). Cas12f is another hypercompact genome editor discovered in *Acidibacillus sulfuroxidans*, which is approximately one quarter the size of Cas9 (72). This Cas effector makes use of a 5'-NTTR-3' PAM sequence in its associated sgRNA, with R being C or G, however the

associated sgRNA is much longer than that associated with Cas ϕ (72). Both Cas ϕ and Cas12f have been shown to have similar editing efficiencies to Cas9 (71, 72). These Cas proteins are of particular interest, as smaller molecule size can be very helpful for ribonucleoprotein (RNP) delivery.

All ReMOT control studies to date use chimeric Cas9 to enable germline editing in insects (49, 89, 90, 94-101, 166). This includes when deployed in the silverleaf whitefly *Bemisia tabaci* (Hemiptera) (49). *B. tabaci* is a global polyphagous pest, responsible for crop losses due to feeding damage and viral infection. Moreover, *B. tabaci* infest over 1000 plant species and transmit more than 300 plant pathogenic viruses (38-41). The *B. tabaci* life cycle consists of 6 stages: egg, 4 immature instar stages, and adult (46, 47). When gene editing by ReMOT control was performed on the *white* gene in *B. tabaci*, eye colour change in successfully edited individuals was visible from 4th instar onwards (49). *B. tabaci* have a haplodiploid sex determination system, where males have only one copy of each chromosome (48), making proof of principle gene editing simple to screen for, as a single edited allele can give rise to an observable phenotype (49).

ReMOT control in *B. tabaci* makes use of BtKV, an embryo-localising peptide derived from Vitellogenin (49). By creating chimeric BtKV-Cas ϕ proteins, it may be possible to deliver more gene editing molecules to *B. tabaci* embryos when using ReMOT control, than with chimeric BtKV-Cas9, owing to the small size of Cas ϕ . Further, use of Cas ϕ in *B. tabaci* would further expand the currently limited insect CRISPR toolbox. Therefore, the goal herein is to analyse whether BtKV-Cas ϕ could be a viable embryo-targeting genome editor for use in *B. tabaci*. This is achieved through expression and purification of BtKV-

Cas ϕ , followed by validation of its nuclease activity in vitro using sgRNAs designed against the *B. tabaci white* gene. Finally, injection of BtKV-Cas ϕ into *B. tabaci* adult females to test its efficacy as a ReMOT control suitable genome editor.

6.1.1 Contributions to this chapter

Work discussed in this chapter was part of a year in industry placement programme. I designed the project, its aims and the experiments performed herein. Experimental work and analysis was carried out by Amber Hall with my supervision.

6.2 Results

6.2.1 Construction of pET28a-BtKV-Cas ϕ -Cys

For production of chimeric BtKV-Cas ϕ , the plasmid pET28a-BtKV-Cas ϕ -Cys was produced. This was achieved by Gibson assembly to replace Cas9 in pET28a-BtKV-Cas9-Cys with Cas ϕ (Figure 6.1).

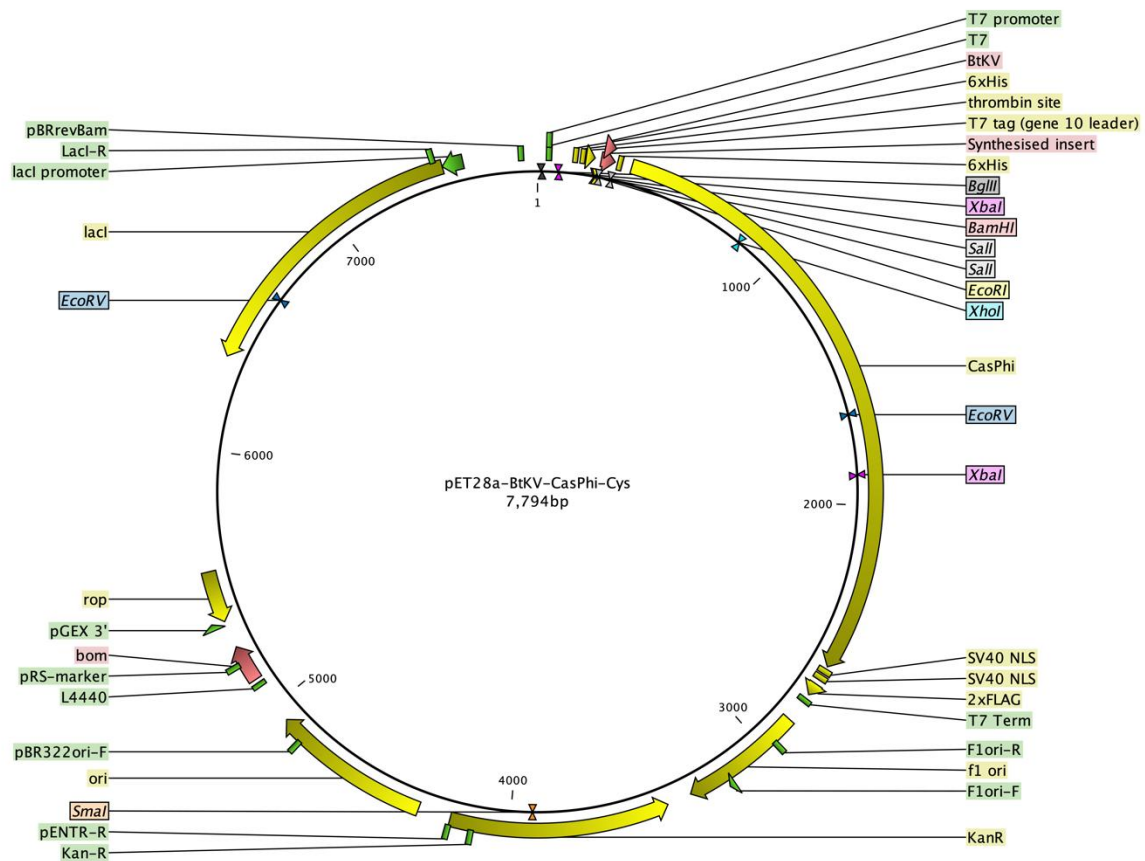


Figure 6.1: A plasmid map of pET28a-BtKV-Cas ϕ -Cys, created by swapping the Cas9-NLS-HA portion of pET28a-BtKV-Cas9-Cys with the Cas ϕ -2xSV40 NLS-2xFLAG portion of pPP441 by Gibson assembly.

6.2.2 BtKV-Cas ϕ heterologous production in *E. coli*

We expressed BtKV-Cas9 from pET28a-BtKV-Cas ϕ -Cys in *E. coli* and purified the protein on a large scale. Following standard pET28a expression and IMAC purification protocols, BtKV-Cas ϕ was produced to a concentration of 1.25 mg/mL, with high purity as shown by SDS-PAGE (Figure 6.2). Pure sample elution wells show strong bands larger than the expected ~90kDa product size. Therefore, further analysis of protein function through an in vitro cutting assay was conducted to confirm the identity and functionality of the protein.

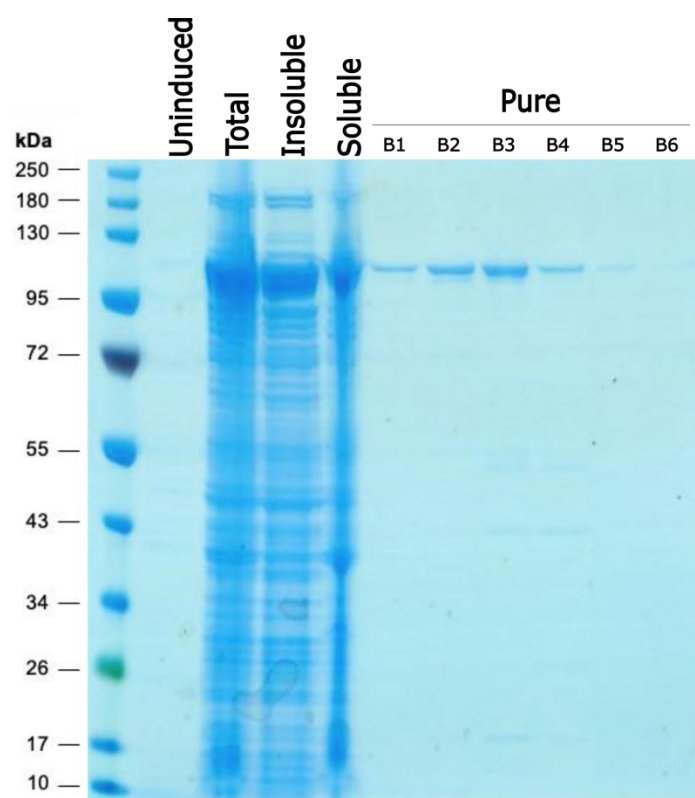


Figure 6.2: SDS-PAGE shows a pure sample of BtKV-Cas ϕ was produced. Uninduced, total, insoluble, and soluble fractions show that the protein is expressed and much of it remains in the insoluble fraction. Pure sample elution wells B1-6 all show strong bands larger than the expected product size of ~90kDa.

6.2.3 Design and synthesis of Cas ϕ sgRNAs targeting *B. tabaci white*

We designed 5 Cas ϕ sgRNAs targeting exons 2, 3 and 5 of *B. tabaci white*. These targets were chosen for their similarity to the successful Cas9 sgRNAs used in the previous ReMOT control study (49). Each candidate sgRNA target sequence was subjected to BLASTn to identify and off target effects. Any sgRNA designs displaying 100% identity or a single mismatch with off-target sequences were rejected. The final five sgRNAs are listed in table 6.1 and were synthesised by Merck. sgRNA1 and 3 were also synthesised by T7 transcription, and quality of the results was analysed by Tapestation (Agilent). This showed that both sgRNAs 1 and 3 had the expected sizes.

Table 6.1: Cas ϕ sgRNA target sequences, the white gene exons they target, and their 5' PAM sequences.

sgRNA ID	Target sequence	Exon targeted	PAM (5'- 3')
sgRNA1	GCCTCCGGACAGGACCAAAG	2	TTC
sgRNA2	GCGATAATGGGCTCCAGTGG	3	TTA
sgRNA3	CATTGACTGCTCGCTGGCCG	3	TTC
sgRNA4	AGTAAATGTCAGCACACTGT	5	TTA
sgRNA5	TTCTACCAGGAGTTCCTATC	5	TTA

6.2.4 BtKV-Cas ϕ cleaves *B. tabaci* white amplicons when complexed with sgRNAs

We performed *in vitro* cutting assays to assess whether BtKV-Cas ϕ cleaves genomic *B. tabaci* DNA when complexed with sgRNAs. Amplicons containing the target exons 2, 3 and 5 (targeted by sgRNA1, sgRNAs 2 and 3, and sgRNAs 4 and 5 respectively), were used as substrate DNA for a nuclease reaction by BtKV-Cas ϕ -sgRNA RNPs, and expected to degrade substrate DNA, and generate ~400bp fragments corresponding to the cleavage products. All 5 sgRNAs synthesised by Merck, and sgRNA 1 and 3 synthesised by T7 transcription were tested in *in vitro* cutting assay with BtKV-Cas ϕ .

DNA cleavage was not observed in assays using the 5 sgRNAs designed by Merck. Of the 5 sgRNAs synthesised by Merck, sgRNAs1, 2 and 3 complexed with BtKV-Cas ϕ initially cleaved the substrate DNA, but generated DNA fragments at unexpected sizes, as

evidenced by a low molecular weight smear between 100 and 200bp, which is absent in the negative control (untreated substrate DNA), while sgRNAs 4 and 5 did not generate DNA fragments at all. In assays with sgRNAs 1, 2 and 3, much of the substrate DNA remained, suggesting low cleavage efficiencies (Figure 6.3A). Investigations of the nature of the low molecular weight DNA fragments via including additional negative controls revealed that the low molecular weight smear is derived from excess sgRNAs. Finally, band intensities corresponding to DNA fragments in controls without sgRNA and in the presence of BtKV-Cas ϕ were weaker or absent compared to the DNA only negative control (Figure 6.3B). We therefore concluded that these sgRNAs had failed. In contrast, the sgRNAs synthesised by T7 transcription (sgRNAs1 and 3) both enabled cutting of substrate DNA in the presence of BtKV-Cas ϕ . We observed excess sgRNA, demonstrated by low molecular weight smears between 100 and 200bp that were absent from the negative controls. In these cases, no uncut substrate DNA was identified, suggesting T7 transcribed sgRNAs enable nuclease activities at high efficiencies (Figure 6.4).

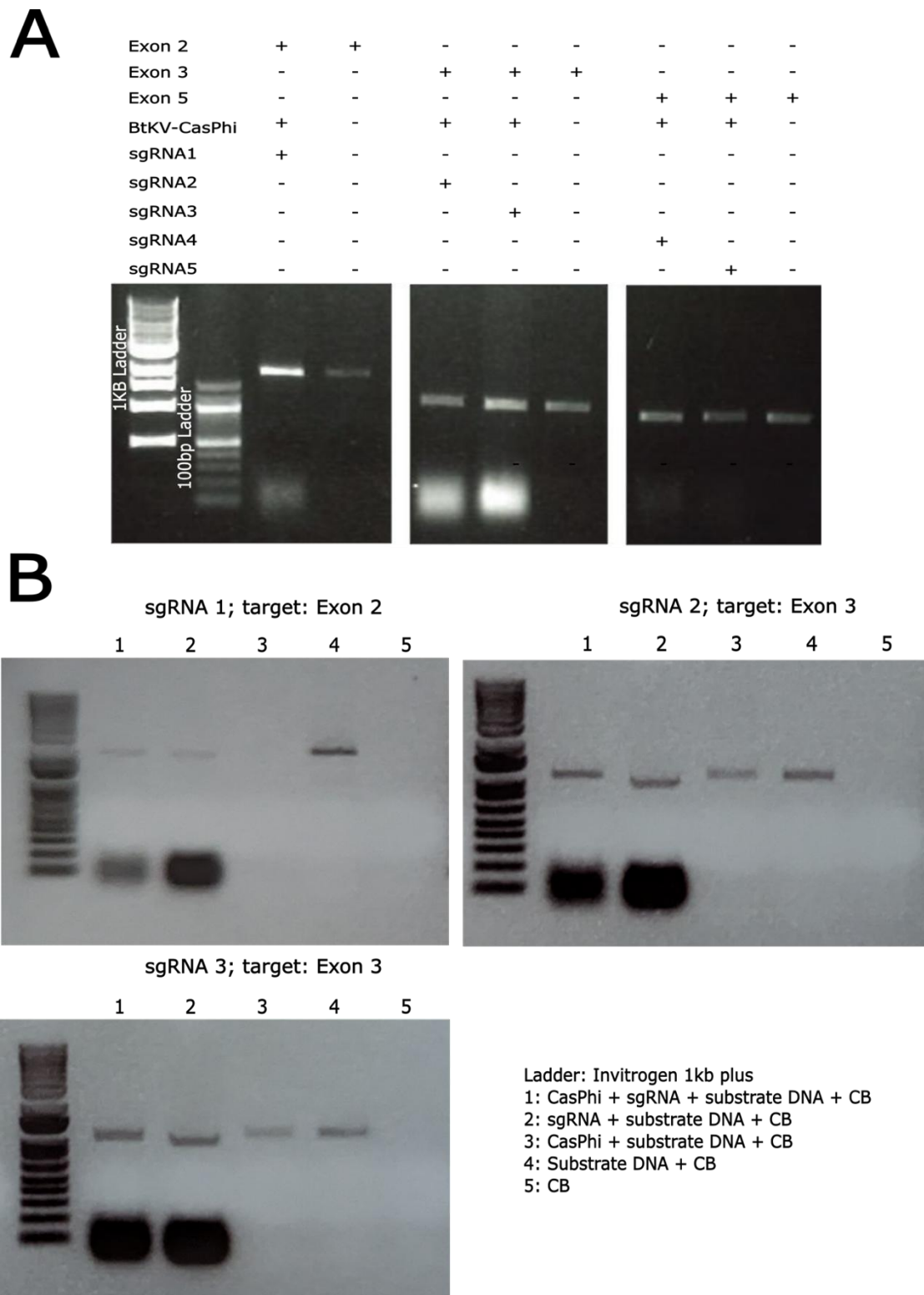


Figure 6.3: In vitro nuclease assays using BtKV-Cas ϕ with sgRNAs synthesised by Merck. (A) Initial assays using sgRNAs 1, 2 and 3 showed a low molecular weight smear at 1-200bp, absent from the negative control. No evidence of cleavage was observed for sgRNAs 4 and 5. (B) Assays for sgRNA1, 2 and 3 were re-run with more negative controls, revealing the same low molecular weight smear at 1-200bp in the control with no BtKV-Cas ϕ (lane 2), showing the smear is sgRNA. There is no observable difference in remaining substrate DNA treated with sgRNAs in the presence (lane 1) or absence of BtKV-Cas ϕ (lane 2). Some degradation of DNA appears to have occurred in reactions containing substrate DNA and BtKV-Cas ϕ in the absence of an sgRNA (lane 3).

6.2.6 Evidence of editing of *white* was observed

In preliminary ReMOT control injections, we ultimately found no evidence of successful gene editing. We observed eye colour change in offspring of whitefly injected with BtKV-Cas ϕ -sgRNA RNPs targeting the *white* gene. Loss of function mutations in *white* cause a reddening of eyes in whitefly visible in 4th instar nymphs (49). Given that males are haploid, an observable phenotype should be seen upon editing their single allele (49). Wild type *B. tabaci* eyes are dark, appearing black under a light microscope (Figure 6.5A). The offspring of females injected with BtKV-Cas ϕ complexed with T7 transcribed sgRNA1 and 3 included two nymphs that displayed red-eye phenotypes. We believed the *white* gene of these individuals may be successfully edited via ReMOT control using BtKV-Cas ϕ (Figure 6.5B and C). However, Sanger sequencing of these nymphs and WT male nymphs was inconclusive. In the progeny displaying reddened eyes, double peaks were observed in the target 3 region of the gene (Figure 6.6); however, there is a high level of noise in all samples, so further sequencing is required to verify if this is successful gene editing.

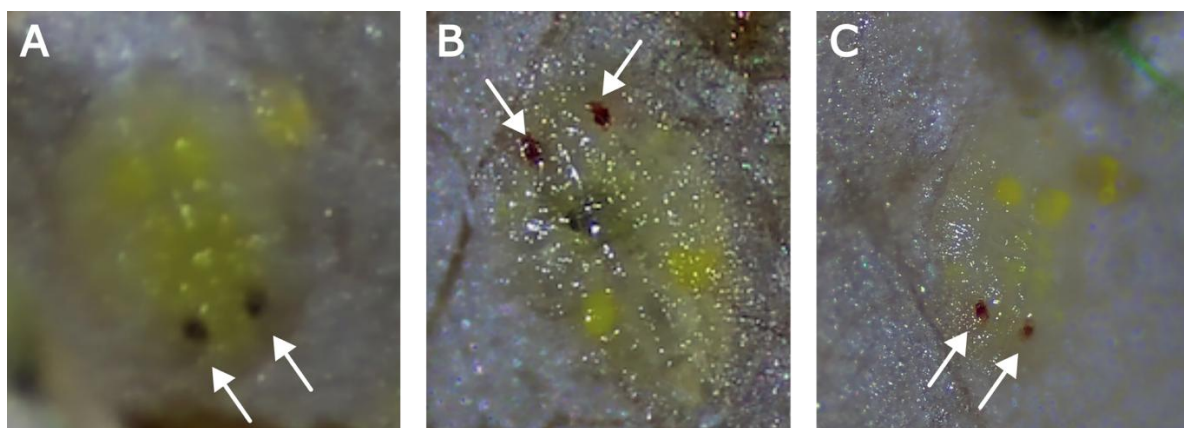


Figure 6.5: Two 4th instar nymphs showed phenotypic evidence of successful gene editing by BtKV-Cas ϕ -sgRNA RNPs. (A) A wildtype nymph with black eyes developing marked by white arrows. (B and C) Two nymphs displayed a red eye phenotype, indicating a loss of function mutation in *white* caused by gene editing.

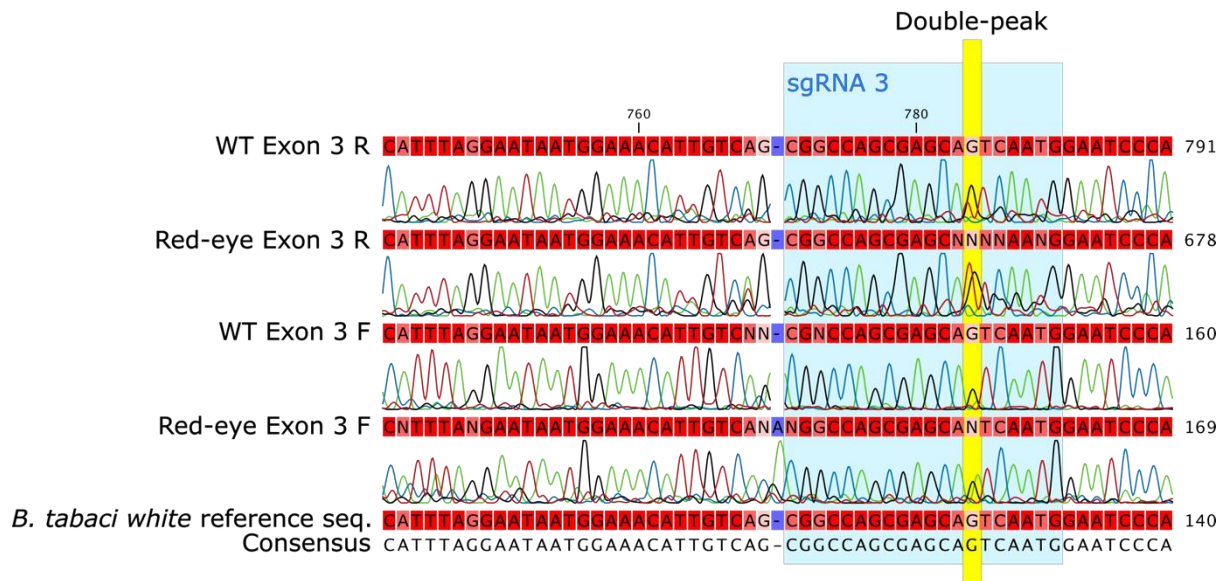


Figure 6.6: An alignment with Sanger sequencing trace data of the sgRNA 3 target region of exon 3. The WT sample consisted of 2 WT male nymphs. The Red-eye sample consisted of the 2 red-eye nymphs shown in Figure 6.5. A double peak is observed in the target region in the red-eye sample, causing it to read as 'N', while the WT and reference sequence read as G. High background 'noise' is present in all samples.

6.3 Discussion

Using *B. tabaci* as a study system, we have laid the groundwork for testing the efficacy of using Cas ϕ as an alternative to Cas9 in ReMOT Control. We produced BtKV-Cas ϕ in *E. coli* and shown that in complex with Cas ϕ sgRNAs synthesised via T7 transcription, BtKV-Cas ϕ cleaves amplicons derived from the *B. tabaci white* gene. Further, we have observed phenotypic evidence that, when deployed in ReMOT control experiments, BtKV-Cas ϕ is capable of editing *B. tabaci* genes in the offspring of injected mothers.

Cas ϕ is a genome editor approximately half the size of Cas9 (71), which makes it a good alternative for methods such as ReMOT control where delivery of preformed RNPs is essential. ReMOT control relies on receptor mediated endocytosis in the oocytes, initiated by the binding of yolk precursor protein (YPP) derived peptides fused to Cas9, with the peptide's receptor (49, 89, 90, 94-101, 166). It follows that if more Cas ϕ molecules can fit into the endosome than Cas9, then more Cas ϕ will be delivered leading to increased chances of successful gene editing. ReMOT control in *B. tabaci* has already been achieved via injecting females with a fusion of the *B. tabaci* Vg derived peptide 'BtKV' and Cas9 (49). Here, we have adapted this system using arguably a more efficient Cas ϕ .

We successfully purified BtKV-Cas ϕ from *E. coli*. Construct pET28a-BtKV-Cas9-Cys was used to produce BtKV-Cas9 (49). Here, we replaced the Cas9 part of this construct with Cas ϕ using Gibson assembly, followed by IPTG-mediated induction of expression, and purification of BtKV-Cas ϕ using IMAC and gel filtration methods. This was proven to be

sufficient for obtaining BtKV-Cas ϕ (-2NLS-2xFLAG) at a high purity. This may be expected given that Cas ϕ , as well as mutant variants nCas ϕ and vCas ϕ , have all been purified in similar ways previously (71, 214). Nonetheless, the BtKV addition could have altered the properties of Cas ϕ making it harder to purify. This work showed that BtKV did not obviously alter the expression characteristics of Cas ϕ production. BtKV-Cas ϕ was ultimately purified to a concentration of 1.25 mg/mL at a high purity.

We demonstrated in *in vitro* cutting assays that BtKV-Cas ϕ -sgRNA RNPs targeting the *white* gene catalysed the cleavage of PCR-derived substrate DNA. However, in complex with commercially synthesised sgRNAs, BtKV-Cas ϕ was unable to cleave a proportion of the substrate DNA. However, when complexed with sgRNA1 and 3, synthesised by T7 transcription, substrate DNA was entirely cleaved by BtKV-Cas ϕ . The reason for this difference in cleavage efficiency remains unclear. It is possible that a lower number of successfully formed RNPs were generated prior to the addition of substrate DNA in these experiments, which could also account for the excess sgRNA observed at 100-200 bp. Interestingly, some degradation was noted in the negative control samples containing substrate DNA and BtKV-Cas ϕ without any sgRNA, suggesting the possibility of non-specific cleavage. Cas ϕ , like other Cas12 proteins, exhibits non-specific trans-cleavage activity on single-stranded (ss)DNA (7, 28); however, this activity should not affect PCR-amplified double-stranded DNA under these conditions.

We performed preliminary ReMOT control experiments via injecting whitefly females with BtKV-Cas ϕ complexed with sgRNAs targeting *white*. After injection, low survival rates of injected females were observed. This is most likely due to human error in the injection

procedure, as these were among the individuals in first injection experiments. However, it could also be due to damage from the needle or too much exposure to cold prior to injection. These issues may be solved through further optimisation of the injection protocol, using sharper needles, or using quartz needles that are less prone to breaking. Quartz needles were used in the injection procedure for ReMOT control using BtKV-Cas9 (49). Further, CO₂ treatment to anaesthetise insects rather than cold exposure may improve survival rates. *B. tabaci* handling post injection may also require optimisation. We caged injected *B. tabaci* on *B. rapa* leaves in clip cages, while in published ReMOT control system methods, injected *B. tabaci* were kept on a soybean leaflets wrapped in moist paper towels in petri dishes (49). We observed that during the two weeks required for injected *B. tabaci* to lay sufficient eggs, *B. rapa* plants often outgrew their pots. As a result, larger leaves—frequently supporting clip cages—would yellow and detach from the plant, leading to increased *B. tabaci* mortality. Despite these issues, surviving injected *B. tabaci* laid many eggs which hatched. These nymphs were observed under a light microscope and two of them appeared to show a red eye phenotype, which is caused by a loss of function mutation in *white* (49). Sanger sequencing across the target regions of *white* revealed a double peak in the red-eye nymphs in the sgRNA 3 target region possibly indicative of a gene editing event; however, high background noise renders this result inconclusive as the double peaks could be a result of a more pronounced moiety in the background of the red-eye sample. Further whitefly injections are required to gather more evidence for editing by BtKV-Cas ϕ .

Here we have laid the groundwork for a ReMOT control system making use of chimeric Cas ϕ as opposed to Cas9. Cas ϕ has been used for editing in bacteria, mammalian cells,

and plants (71, 214, 215). Successful use of Cas ϕ in ReMOT control for *B. tabaci* would be the first use of Cas ϕ in insects, therefore expanding the CRISPR toolbox for genetic studies in insects. Most ReMOT control studies made use of the peptide P2C, derived from *Drosophila melanogaster* yolk precursor protein 1 (DmYPP1) (89, 90, 94-97, 99, 101). Future work producing chimeric P2C-Cas ϕ could enable Cas ϕ based ReMOT control editing of many other insects, potentially with higher efficiencies.

6.4 Materials and methods

6.4.1 Construction of pET28a-BtKV-Cas9-Cys

BtKV was synthesised directly into pET28a-Cas9-Cys by NBS biologicals. CyspET28a/Cas9-Cys was a gift from Hyongbum Kim (Addgene plasmid # 53261 ; <http://n2t.net/addgene:53261> ; RRID:Addgene_53261 (193). The sequence ‘GTCGACGGAAACCGTATGGCGTGTATAAAACCATGGAAGATAGCGTGGGTCGAC’ was synthesised and cloned into the Sall cut site of pET28a-Cas9-Cys by NBS biologicals. To enable this, the BtKV sequence is flanked by Sall cut sites. Two random base pairs were added to the 3’ end to maintain framing.

6.4.2 PCR to generate a Cas ϕ fragment for Gibson cloning

PCR was used to generate a fragment to be used in Gibson assembly to create pET28a-BtKV-Cas ϕ -Cys. The plasmid pPP441, containing the Cas ϕ coding sequence, was used

as a template. pPP441 was a gift from Jennifer Doudna (Addgene plasmid # 158801 ; <http://n2t.net/addgene:158801> ; RRID:Addgene_158801) (71). Primers were designed to amplify the entire coding sequence of Cas ϕ , whilst also generating a fragment to be used in Gibson assembly. Therefore, the forward primer includes a 25bp ‘overlap’ sequence upstream of the annealing section, complementary to the sequence of the vector upstream of the chosen cut site in the vector pET28a-BtKV-Cas9-Cys. The annealing section of the primer is complementary to the first 23bp of the Cas ϕ coding sequence. To maintain framing, 4 extra bp were added upstream of the annealing section. In the reverse primer, the overlap sequence is complementary to the 23bp downstream of the chosen cut site, and the annealing section is complementary to the final 24bp (on the reverse strand) of the Cas ϕ coding sequence. To maintain framing, 2 extra bp were added. These primers were ordered from Merck and are presented in table 6.2. PCR was carried out using Platinum SuperFi master mix (Invitrogen, #12351010). Gradient PCR was first carried out to find an appropriate annealing temperature. Final PCR was carried out as 25 μ L reactions, containing 25 μ L 2x Platinum SuperFi master mix, 10 μ L SuperFi GC enhancer, 2.5 μ L each primer at 10 μ M, 1 μ L pPP441 (template) at 100 ng/ μ L, and 9 μ L nuclease-free water (Merck). Cycling conditions were initial denaturation at 98°C for 30s, 30 cycles of denaturation at 98°C for 10s, annealing at 67°C for 10s and extension 72°C for 1min followed by final extension at 72°C for 5 min then a 4°C hold. 10 μ L of the product was analysed by gel electrophoresis on a 1% agarose gel in 1xTAE. The remaining 40 μ L of the products was purified using the QIAquick PCR Purification Kit (QIAGEN, #28104). Concentration and purity were measured by Nanodrop (ThermoFischer). Finally, products were verified by sanger sequencing (Eurofins).

Table 6.2: Primers used to amplify a fragment suitable for Gibson assembly containing the Cas ϕ coding sequence from pPP441. ‘Overlap’ sections required for Gibson assembly are in lowercase. Extra base pairs, added to maintain framing, are in lowercase and underlined. Annealing sections of the primers are in uppercase.

Primer ID	Sequence
BtKV-Cas ϕ -F1	ccatggaagatagcgtgggtcgacaagctATGCCAAAGCCAGCCGTGGAGTC
BtKV-Cas ϕ -R1	ctcagtggtggtggtggtggtgctcCTTATCATCATCCTTGTAGTC

6.4.3 Gibson cloning to construct pET28a-BtKV-Cas ϕ -Cys

Gibson cloning was used to replace Cas9 with Cas ϕ in pET28a-BtKV-Cas9-Cys, creating pET28a-BtKV-Cas ϕ -Cys. Initially, pET28a-BtKV-Cas9-Cys was linearised by digestion with XhoI (ThermoFischer, #FD0694) and HindIII (ThermoFischer, #FD0505), removing the Cas9 part of the plasmid. Digested plasmid was and the Cas9 part were separated by gel electrophoresis on a 1% agarose gel in 1xTAE. The band corresponding to the digested plasmid was excised and DNA was extracted using the QIAEX II Gel Extraction Kit (QIAGEN, #20021). The concentration and purity of the extracted DNA was measured by Nanodrop (ThermoFischer). Gibson assembly was carried out using the NEBuilder HiFi DNA Assembly Kit (NEB, #E5520S) according to the manufacturers protocol. An insert to vector molar excess of 2:1 was desired, with 50-100 ng of vector for optimum efficiency. Therefore, in the cloning reaction 68 ng (8 μ L of 8.5 ng/ μ L stock) of linearised pET28a-BtKV-Cas9-Cys (with Cas9 removed in digestion) was mixed with 59 ng (2 μ L of 25.9 ng/ μ L stock) of amplified Cas ϕ , giving a molar excess of insert to vector of ~2:1 as calculated with NEBioCalculator. After the cloning reaction, 2 μ L of the product was transformed into NEB DH5 α cells (NEB, #C2987) by heat shock. The transformed cells were then spread onto LB agar plates containing carbenicillin at a working concentration

of 100 µg/mL and incubated standing at 37°C overnight. The following day, colonies were subject to colony PCR using the same primers and PCR method used to amplify the Casϕ Gibson fragment. Positive colonies were then grown overnight in 10 mL LB containing 100 µg/mL carbenicillin, shaking at 37°C. The following day, 500 mL of each culture was mixed with 1 mL 40% glycerol and stored at -80°C as a glycerol stock. Then 5 mL of each culture was subject to plasmid extraction by miniprep using a QIAprep Spin Miniprep Kit (QIAGEN, # 27106). Concentration and purity were analysed by Nanodrop (ThermoFischer) and cloned constructs were verified by sequencing (Eurofins) using the universal T7 promoter primer as a forward primer, and the T7 terminator primer as a reverse primer.

6.4.4 Small scale production of BtKV-Casϕ

Small scale BtKV-Casϕ production was carried out using standard pET28a expression and purification systems for purification using IMAC. pET28a-BtKV-Casϕ-Cys was transformed into BL21(DE3) *E. coli* (Merck, #2572) by heat shock according to the manufacturers protocol. The transformed cells were then spread onto LB agar plates containing 50 µg/mL kanamycin and left standing at 37°C overnight. The next day, starter cultures of 10 mL LB with 50µg/mL kanamycin were each inoculated with a single colony and allowed to grow overnight shaking at 37°C. The following morning, 4 mL of starter culture was used to inoculate 100 mL TB with 9.4 g/L K₂HPO₄, 2.2 g/L KH₂PO₄. These cultures were incubated shaking at 37°C until an OD600 of 0.6-0.8 was reached. At this point, 2 µL ‘uninduced’ sample was taken. The cultures were then stored at 4°C for at

least 20 minutes. Expression was induced by addition of IPTG to a working concentration of 1 mM. The induced cultures were incubated shaking at 16°C overnight for expression to occur. The following morning, 2 µL was taken as a 'total protein' sample. Then cultures were centrifuged at 24,000 g for 30 minutes to collect the cells in a pellet. The supernatant was discarded, and samples were resuspended in 2 mL IMAC buffer A1 with 2 µL 100 mM PMSF protease inhibitor. Cells were lysed by sonication on ice with 20 cycles of 5s on 5s off. After this, all steps were carried out in a 4°C cold room. The samples were centrifuged at 20,8127 g (max speed). 2 µL of the supernatant was taken as a 'soluble' sample, and 2 µL resuspended pellet was taken as a 'insoluble' sample. The supernatant was transferred to a new 2 mL Eppendorf and 100µL Ni⁺ agarose beads (Ni-NTA His-Bind Resin, Merck, #70666-10 ML) were added. This mixture was centrifuged at 300 g for 1min. Then 2 µL supernatant was taken as an 'unbound' sample. The supernatant was removed and discarded, and the beads were washed with 1 mL IMAC buffer A1. This mixture was then centrifuged at 300 g for 1min, and the supernatant was discarded. Washing was repeated two more times. Finally, 200 µL IMAC buffer B1 was added to the beads. The mixture was centrifuged at 300 g for 1min, and the supernatant was transferred to a clean 1.5 mL Eppendorf tube as the 'pure' sample.

The samples taken were subject to analysis by SDS-PAGE using precast gels (NuPAGE 4-12% Bis-Tris, Invitrogen, #NP0321BOX). The samples were mixed with 4x NuPAGE LDS sample buffer (Invitrogen, NP0007), buffer A1 and IDT, boiled for 5 minutes for denaturing, and loaded onto the gels, in tanks containing NuPAGE MOPS SDS running buffer (Invitrogen, #NP0001). Also loaded was colour protein standard size marker (NEB, #P7719S). The proteins were separated at 100 volts for 10 minutes followed by 200 volts

for ~20 minutes. Gels were then moved to square plates and stained with ReadyBlue Protein Gel Stain (Sigma-Aldrich, #RSB-1 L) via shaking for ~1 hour. This was then washed off with distilled water. Gels could then be imaged by scanning.

6.4.5 Large scale production of BtKV-Cas ϕ

Large scale expression of *BtKV-Cas9* was carried out in the same way as for small scale production but scaled up to produce 2 x 1 L expression cultures. Samples were again taken for both ‘uninduced’ and ‘total protein’ samples. After expression, these cultures were subject to centrifugation at 24,000 g for 30 mins. The supernatant was discarded, and the pellets were resuspended in 50 mL IMAC buffer A1 with 1 c0mplete, EDTA-free protease inhibitor cocktail tablet (Roche, #11873580001). Each 50 mL sample was subject to cell lysis by sonication with 1s on, 3s off cycles for a total of 20mins. Then, the samples were centrifuged at 24,000 g for 30 mins. Samples were taken from the supernatant and the pellet as ‘soluble’ and ‘insoluble’ samples respectively. The Lysate was then combined and subject to purification by IMAC and gel filtration using an AKTA express system with a 5 mL HisTrap High Performance column (Merck) and an S200 16/60 gel filtration column with no detergent. Pure protein was eluted in IMAC buffer A4 into a 24 well plate. The uninduced, total protein, soluble and insoluble samples, as well as samples from wells corresponding to elution peaks, were subject to analysis by SDS-PAGE as described for small scale production. The pure samples from wells found to be containing the desired product were combined and concentration was measured by Qubit (ThermoFischer) using the Protein Assay Kit (ThermoFischer, #Q33211). Finally, the

protein sample was aliquoted into 50 μ L aliquots which were snap frozen in liquid nitrogen and stored at -80°C for later use.

6.4.6 sgRNA design against *B. tabaci white*

The *white* gene sequence (NW_017550151, region 408016-472564) was chosen as a target gene as it had been used for ReMOT Control previously in *B. tabaci* (49). Cas ϕ sgRNAs were designed against this sequence by searching the exons for 5'-TBN-3' PAM sequences, where B is either G, C or T. The 18bp downstream of these sites were selected as complementary target sequences in the 'protospacer' part of the sgRNA. 5 candidate target sequences were subject to BLASTn (NCBI) against the *B. tabaci* MEAM1 genome to check for off target complementary sequences. The required structural part of the sgRNA sequence (CAACGATTGCCCCTCACGAGGGGAC) was then added upstream of the protospacer of candidates unlikely to cause off target effects.

6.4.7 sgRNA synthesis

sgRNAs were either synthesised as RNA molecules by Merck or via T7 transcription. sgRNA 1-5, targeting exon 2 (sgRNAs 1), exon 3 (sgRNA2 and 3), and exon 5 (sgRNA4 and 5) were synthesised by Merck. sgRNAs 1 and 3 were also synthesised by T7 transcription, based on a protocol developed by Richardson et al., 2016 (216). Forward and reverse template oligos which encode important parts of the sgRNA sequence are used as templates for PCR to generate a T7 transcription template. These oligo templates were

adapted to generate Cas ϕ sgRNAs. The forward (T7FwdLong) oligo remains constant, containing the structural part of the sgRNA, while the reverse oligo (T7RevVar) varies as it contains the protospacer part of the sgRNA. Primers were designed against these oligos. As such the forward primer (T7FwdAmp) was the same throughout, while the reverse primer (T7RevAmp1 and 2, applying to sgRNA1 and 3) varied according to the desired sgRNA transcription template. All oligos and primers were ordered from Merck and are presented in Table 6.3 and 6.4. PCR to generate sgRNA synthesis templates was carried out using Phusion DNA polymerase (ThermoFischer, #F530S). The reaction consisted of 10.6 μ L nuclease free water (Merck), 4 μ L 5x Phusion HF buffer, 0.4 μ L each T7FwdLong and T7RevVar1 or 2, 2 μ L T7FwdAmp primer, 2 μ L T7RevAmp1 or 2 primers, and 0.2 μ L Phusion (2 U/ μ L). Cycling conditions were denaturing at 98°C for 30s, then 30 cycles of denaturing at 98°C for 10s, annealing at 51°C for 10s, and extension at 72°C for 10s, followed by a final extension at 72°C for 2 mins. Products were directly added to the T7 transcription reaction. For T7 transcription the HiScribe T7 High Yield RNA Synthesis Kit (NEB, #E2040S) was used according to the manufacturers protocol, with an overnight incubation at 37°C. Template DNA was then removed by a DNase. 5'-triphosphate groups were removed using shrimp alkaline phosphatase (rSAP) enzyme (NEB, #M0371S) with incubation at 37°C for 3 hours. Finally, sgRNAs were purified using the RNEasy Mini Kit (QIAGEN, #74104) according to the manufacturers protocol. The sgRNAs were analysed for concentration and purity by both Nanodrop (ThermoFischer) and TapeStation (Agilent).

Table 6.3: Oligos ordered for PCR DNA template for generation of sgRNA T7 transcription templates.

Template ID	oligo Sequence
T7FwdLong	TAATACGACTCACTATAGCAACGATTGCCCTCACGAGGGGAC
T7RevVar1	CTTTGGTCCTGTCCGGAGGCGTCCCCTCGTGAGGGGCAATCGTTG
T7RevVar2	CCACTGGAGCCCATTATCGCGTCCCCTCGTGAGGGGCAATCGTTG

Table 6.4: Primers used to generate sgRNA T7 transcription templates by PCR.

Primer ID	Sequence
T7FwdAmp	TAATACGACTCACTATAG
T7RevAmp1	CTTTGGTCCTGTCCGGAGGCGT
T7RevAmp2	CCACTGGAGCCCATTATCGCGT

6.4.8 *B. tabaci* genomic DNA extraction

B. tabaci genomic DNA was extracted using the GenElute -E Single Spin Tissue DNA High Yield Kit (Merck, #EC300) according to the manufacturers protocol with the addendum that 10 mg whitefly was ground in a 1.5 mL Eppendorf using a plastic micro-pestle in the supplied lysis buffer, rather than manual tissue cutting/crushing prior to lysis buffer addition. Extracted DNA was analysed using Nanodrop (ThermoFischer).

6.4.9 PCR for fragments of *B. tabaci* white

Amplicons containing target sites for sgRNAs were generated by PCR using VeriFi DNA polymerase (PCR Biosystems, #PB10.43-01). As the sgRNAs targeted exons 2, 3 and 5, one fragment for each exon was generated. Primers were designed to amplify each required exon in CLC Main Workbench (QIAGEN) using default primer design settings. These primers were ordered from Merck and are presented in Table 6.5. Each PCR reaction consisted of 19 μ L nuclease free water (Merck), 25 μ L VeriFi master mix, 2 μ L each of forward and reverse primer from a 10 μ M stock, and 2 μ L *B. tabaci* template gDNA (46.8 ng/ μ L). PCR was carried out according to the VeriFi manufacturers protocol. First, gradient PCRs were conducted to find appropriate annealing temperatures. For exon 2, the required annealing temperature required was 65°C, and for exons 3 and 5 the annealing temperature required was 58°C. Cycling conditions were denaturation at 98°C for 1 min, followed by 35 cycles of denaturation at 95°C for 15s, annealing at 65°C or 58°C for exon 2 and exons 3 and 5 respectively for 15s, and extension at 72°C for 1min 30s, followed by a final extension at 72°C for 2 mins. PCR products were purified using the QIAquick PCR purification kit (QIAGEN, #28104). Products were then visualised by gel electrophoresis on a 1% gel in 1xTAE. Finally, concentration and purity of products were measured by Nanodrop (ThermoFischer).

Table 6.5: Primers used to amplify exon 2, 3 and 5 amplicons.

Primer ID	Sequence
Exon2-F1	GAATCACGTAGAGCATTG
Exon2-R1	CAGAGAAAACCAGCGGTA
Exon3-F1	TGTTAAGGAAGAACGTCG
Exon3-R1	TTAGCACAGAACACATTGG
Exon5-F1	TTGGGATGAGATCTACTG
Exon5-R1	TTAGGTACAAAGTCGGGG

6.4.10 In vitro cutting assays

In vitro cutting assays were based on the protocols described by Pausch et al., 2020 (71) Li et al., 2023 (215).

For in vitro cutting assays using sgRNAs synthesised by Merck, sgRNAs were first diluted to 50 μ M. R-loop formation was achieved by heating to 65°C for 3 mins, before allowing gradual cooling to room temperature. 3 μ L of each sgRNA was then added to 17.72 μ L 2x cleavage buffer (20 mM HEPES Free Acid pH7.5, 300 mM KCl, 10 mM MgCl₂, 20% glycerol, 1 mM TCEP). Then, BtKV-Cas ϕ was added at a working concentration of 4 μ M (9.28 μ L of 1.25 mg/mL BtKV-Cas ϕ stock). The solution was incubated at room temperature for 30 mins to allow RNP complexes to form. The final reaction mixture was made up of 7.5 μ L formed RNP complexes, 10 nM substrate DNA, and 1x cleavage buffer

to make the final volume up to 30 μ L. Nuclease reactions were then allowed to occur by incubation at 37°C for 30 mins. Reaction products were analysed by gel electrophoresis on a 1% agarose gel in 1xTAE.

In vitro cutting assays using sgRNAs generated by T7 transcription were done in a similar way. However, before R-loop formation sgRNAs were made up to a working concentration of 1.25 μ M in 2x cutting buffer. BtKV-Cas ϕ was added to a working concentration of 1 μ M. Final reaction mixtures were made up of 24 μ L RNPs, and substrate DNA to a working concentration of 10 nM. If the resulting mixture was less than 30 μ L, 1xCB was added to make up the volume to 30 μ L. The remainder of the protocol was carried out the same way as for sgRNAs synthesised by Merck.

6.4.11 Micropipette preparation

Micropipettes were prepared using a Sutter P-97 needle puller to pull 1.14 mm diameter capillaries (WPI, #504949). The programme used was heat=500; pull=150; velocity=90; delay=235. Pulled capillaries were kept in a square Petri dish on plasticine or blu-tac until use.

6.4.12 ReMOT Control experimental procedure

For ReMOT control experiments, *B. tabaci* were injected with an injection mixture of either 2x CB (negative control), or pre-complexed BtKV-Cas ϕ (2.153 mg/mL) in CB.

All injections were performed under a dissection microscope. For injection, micropipettes were backfilled with mineral oil before insertion into a World Precision Instruments Nanoliter 2020 injector. Then, 3 μL of mineral oil was ejected. 20 μL of injection mixture was pipetted onto parafilm and placed under the micropipette tip. The micropipette was then inserted into the droplet using a mechanical micromanipulator, and 2.8 μL of injection mixture was sucked into the micropipette. *B. tabaci* of unknown age were anaesthetised on ice in a petri dish for 10 mins prior to injection. Once docile, *B. tabaci* were moved onto double sided tape on a slide, placed on a pre-cooled metal block. At this stage, any males were removed. Females were injected with ~10 nL of injection mixture. Injected *B. tabaci* were then moved into clip cages on Chinese Cabbage (*B. rapa*) and kept in long day (LD) conditions (14h light at 24°C, 10h dark at 20°C). Plants were checked periodically for eggs and nymphs. Once nymphs were observed, their eye colour was checked under a light microscope. Any nymphs with discoloured eyes were imaged and collected.

7 Discussion

7.1 Summary and discussion of research findings

To achieve the aims of this thesis, I first investigated *M. persicae* vitellogenin to find a suitable peptide for use in ReMOT control (Chapter 2). I found that aphids have lost the canonical Vg, instead retaining a Vg-like protein homologous to Vg-like proteins of other insects. This protein retains the structural domains of canonical Vg. Through multiple sequence alignment, alphafold modelling, and structural alignment, the peptide MpRV was derived from MpVg to be tested for use in ReMOT control.

To test MpRV's efficacy as an embryo-targeting peptide, MpRV-mCherry chimeras were produced from *E. coli* and injected into adult female aphids (Chapter 3). The ovaries of these females were dissected out 24 hours later and inspected for embryo-specific fluorescence. The peptide P2C (used in previous ReMOT control studies) was also tested in the same way, while untagged mCherry and buffer A4 served as negative controls. This experiment involved optimising injection procedures for *M. persicae*. I found that both P2C and MpRV were able to transport mCherry to embryos. Therefore, these peptides were both brought forward to test for ReMOT control.

Firstly, functional MpRV- and P2C-Cas9 chimeras were needed (Chapter 4). MpRV- and P2C-Cas9 production was optimised at a large scale, providing sufficient protein for ReMOT control. The target marker gene *white* was selected, and sgRNAs were designed

and synthesised targeting three different sites on the gene. To test both the functionality of the proteins and the sgRNAs, in vitro nuclease assays against PCR generated fragments of *white* were carried out. These experiments revealed that both MpRV- and P2C-Cas9 were able to cleave *white* at the expected sgRNA target sites. Therefore, these RNPs were ready for use in ReMOT Control (Chapter 5).

MpRV- and P2C-Cas9 complexed with each sgRNA were injected into 7-day old female aphids with cutting buffer, and with or without saponin (Figure 7.1). No evidence was observed suggesting that saponin had an effect on efficiency of gene editing, nor aphid survival post-injection. Offspring were screened for up to 2 weeks post injection; every 2 days offspring were examined for a 'white eye' phenotype which would be expected from successful homozygous gene editing of the gene *white*. Those not presenting with a phenotype were pooled and genotypically screened by PCR, Sanger and amplicon sequencing. Though these screens, two individual aphids were found to have a 'patchy eye' phenotype, thought to have arisen from somatic gene editing. Further, PCR and amplicon sequencing of pooled offspring revealed increased deletions around one sgRNA target cutsite. These findings suggest that the ReMOT control method developed in this thesis gives rise to low level, somatic gene editing. However, much optimisation is needed to make the method reliable and efficient. One possible optimisation is explored in the whitefly *B. tabaci* (Chapter 6).

Through production of BtKV-Cas Φ , we tested whether a smaller Cas effector could be used in ReMOT control. Cas Φ is approximately half the size of Cas9. Smaller size of gene editing complexes in the context of ReMOT control could lead to increased uptake

into oocytes, due to a greater number of molecules being taken up in the same space. This could lead to more efficient gene editing. In vitro nuclease assays showed that we produced a functional BtKV-Cas Φ protein, able to cut the *B. tabaci white* gene at sgRNA targets. Preliminary ReMOT control experiments on *B. tabaci* using this protein resulted in two nymphs displaying a 'red eye' phenotype. Upon Sanger sequencing, the *white* gene of these nymphs appeared to have a deletion at the sgRNA cutsite. This suggests that Cas Φ can be used in ReMOT control. However, more ReMOT experiments are needed to validate this.

Taken together, these findings represent advancements in ReMOT Control gene editing in *M. persicae* and *B. tabaci*. Given enough optimisation, these methods could be further developed to enable genetic studies on these pest insects, which could uncover new pest management strategies, with the possibility of finding an "Achilles' Heel" to target. Further, the characterisation of MpVg (to find an embryo-targeting peptide) has advanced knowledge about aphid Vgs. Canonical Vg has been lost in aphids, but Vg-like has been retained as the sole Vg protein. This protein has conserved Vg domains and is likely to carry out reproductive functions in sexually reproducing aphids. However, Vg's role in asexually reproducing aphids, which don't require its reproductive function, remains unclear. Further research on aphid Vg functions as an immune protein could uncover more information about the need, or lack thereof, for Vg in asexually reproducing aphids.

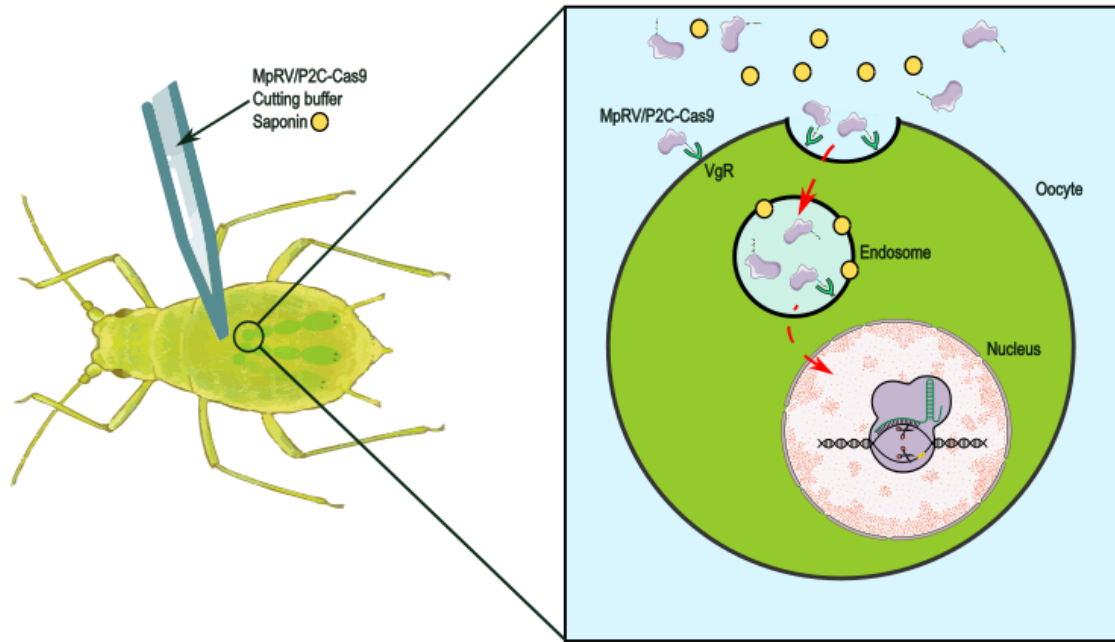


Figure 7.1: A schematic of ReMOT control in *M. persicae*. Chimeric MpRV/P2C-Cas9-sgRNA RNPs are injected into 7-day old adult females with cutting buffer and saponin. This mixture enables effective uptake of the chimeras via MpRV/P2C-VgR binding, followed by endosomal escape mediated by saponin. Finally, the Cas9 portion of the chimeras can cleave the target DNA.

7.2 Future directions

In this thesis, we have characterized key features of MpVg to inform the design and testing of a ReMOT control system in *M. persicae* and developed a ReMOT control system using CasΦ as an alternative to Cas9 in *B. tabaci*. While these systems show promise, further optimization is required to establish their viability as gene-editing strategies. These optimizations could be guided by fundamental studies on vitellogenesis and development, as well as methodological advances in CRISPR/Cas delivery.

A deeper understanding of aphid vitellogenesis and its regulation could refine the timing of injections in ReMOT control experiments in *M. persicae*. Firstly, determining whether any level of vitellogenesis occurs in asexually reproducing aphids would shed light on the likelihood of ReMOT control being effective. Essentially, the VgR must be expressed for

Vg derived peptides to enable ReMOT. By performing fluorescence in-situ hybridisation (FISH) to analyse VgR expression, one could ascertain whether the VgR is indeed expressed in asexual ovaries. Combined with effective DAPI staining, one could determine at which embryo stage (if any) the receptor is expressed.

If it is found that vitellogenesis does not occur at all in asexual reproducing aphids, one possible solution could be to activate it by altering rearing conditions. In shorter day conditions, sexual reproducing aphids lay yolk filled eggs (28), suggesting vitellogenesis (in some form) does take place. Placing aphids in short day conditions prior to injection could induce hormonal regulation to upregulate Vg and VgR expression. Vitellogenesis is regulated by the juvenile hormone (JH), ecdysone (20E), and amino acid/target-of-rapamycin (AA/TOR) pathways, which are well-studied in other insects (27). Homologues of vitellogenesis-regulating genes were identified in *M. persicae*. In mosquitoes, vitellogenesis is triggered by nutrition-activated TOR signalling, which initiates JH biosynthesis through FoxO dephosphorylation (57, 58). Similar pathways are implicated in *N. lugens* (59) and *Locusta migratoria*, where JH also promotes polyploidy by activating protein phosphatase 2A (PP2A) through LCMT1 (33, 60). JH acts via the Methoprene-tolerant receptor (Met) to upregulate vitellogenesis through massive polyploidy (33, 60) and represses JH degradation genes in *Bombyx mori* (61). Additionally, JH-mediated phosphorylation of VgR is required for Vg uptake in oocytes of *L. migratoria* (62).

Insulin-like peptides (ILPs), key regulators of JH signalling, prevent FoxO nuclear localization through phosphorylation by Akt/PKB, thereby suppressing JH biosynthesis and VgR activation (27, 57, 58, 62, 63). In aphids, ILPs have been linked to reproductive

morph switching in response to photoperiod (64, 65). In *A. pisum*, ILP1 and ILP4 are expressed in neurosecretory group 1 cells (NSC1) during long-day conditions, suppressing sexual morph development (65). Micro-cauterization of NSC1 cells in long-day conditions led to sexual morph development, suggesting ILP expression is reduced under short-day conditions (65, 66). This reduction likely allows FoxO dephosphorylation, upregulation of JH biosynthesis, and upregulation of Vg and VgR. Based on this knowledge, we propose a model for vitellogenesis activation through ILP inactivation under short-day conditions (Figure 7.2).

To test this hypothesis, qRT-PCR could measure the expression of *ilp1*, *ilp4*, the insulin receptor (*InR*), *MpVg*, and JH biosynthesis genes in *M. persicae* reared under long- or short-day conditions. For ReMOT control, rearing aphids under short-day conditions before injection could enhance vitellogenesis, increasing the likelihood of successful cargo uptake by oocytes. Post-injection, females could be returned to long-day conditions to prevent full reproductive switching. To identify the optimal injection timing, *M. persicae* nymphs would be reared under long- and short-day conditions, with aphid samples collected at regular intervals for qRT-PCR analysis of *ilp1*, *ilp4*, *InR*, *MpVg*, and *MpVgR* expression. This approach would allow dynamic tracking of expression over time, enabling precise timing for injections and improving ReMOT control outcomes.

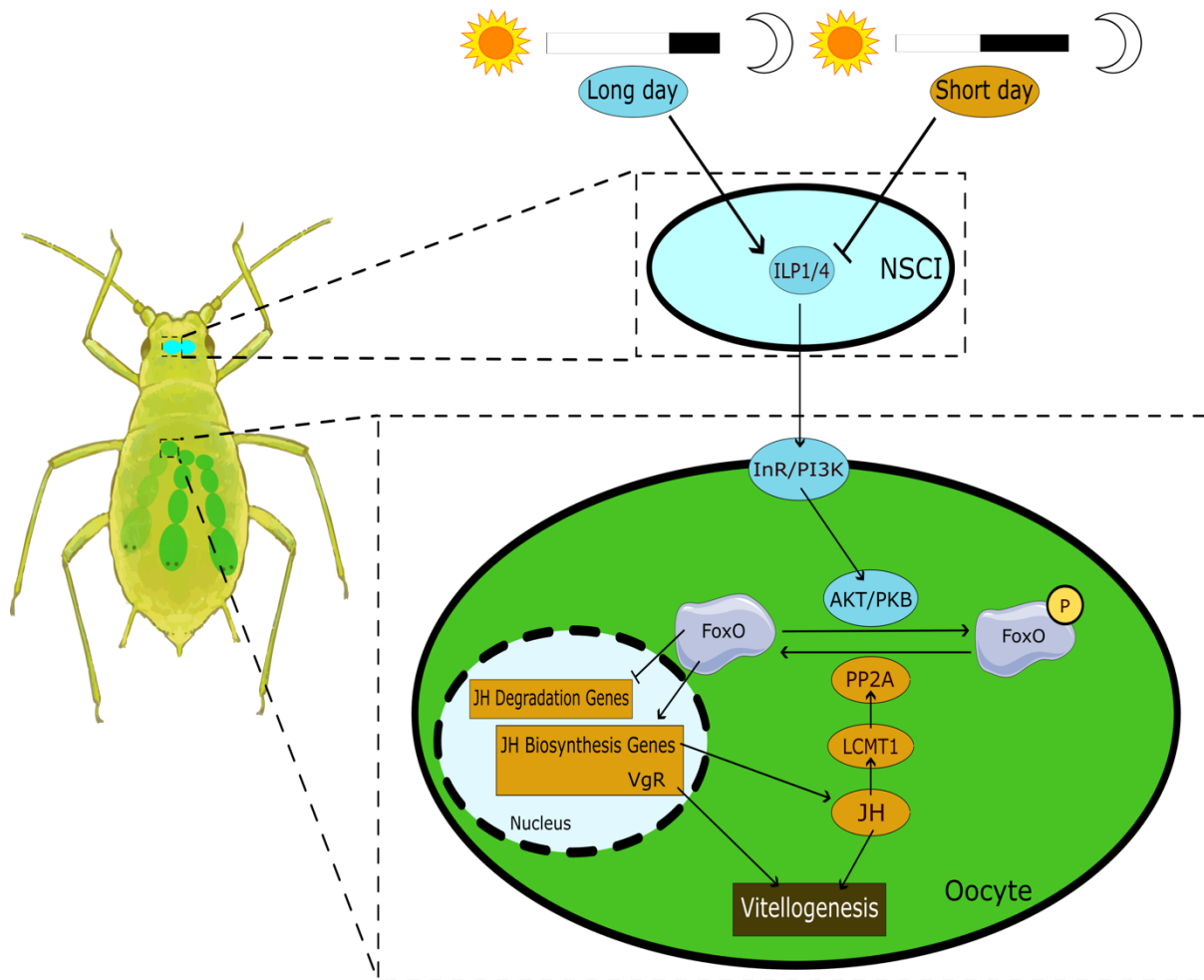


Figure 7.2: A model of the hypothesised signalling pathways regulating vitellogenesis in aphids through day-length sensing and its impact on ILP expression. Under short day conditions, *ilp* expression is inhibited, allowing FoxO dephosphorylation, which upregulates JH biosynthesis, and therefore vitellogenesis. The opposite is true under long day conditions.

Alternative CRISPR methods for *M. persicae* could be explored. Gene editing in *A. pisum* has been achieved by embryo injection after inducing sexual morphs (26). Work by Dr. Mar Marzo, with discussion from Dr. Sam Mugford, and myself has started in our lab to establish a reliable method of inducing sexual reproduction and recovering successfully hatching eggs in *M. persicae*. This work is guided by the experience of our lab and reports by Blackman (1975) on photoperiodic determination of male and female morphs of *M. persicae* (217). If this can be achieved, an embryo (egg) injection based CRISPR/Cas gene editing system like that in *A. pisum* (26) may be used for *M. persicae*. DIPA-CRISPR could

present another option for delivery of Cas9-sgRNA RNPs to developing oocytes via maternal injection (167). However, like ReMOT control, this method relies on timing injections with active vitellogenesis (90, 167). Therefore, a better understanding of aphid vitellogenesis activation will be required. Other alternative delivery methods, such as use of Branched Amphiphilic Peptide Capsules (BAPC), which have been used to deliver gene editing in *Nasonia vitripennis* (96, 144), face the same requirements.

It is possible that vitellogenin derived peptides will not be sufficiently effective in *M. persicae* for reliable gene editing. In this case, other peptides may be considered. For example, cell penetrating peptides (CPPs) such as penetratin (PEN), may be used for ReMOT control in *A. pisum* (168). Localization of PEN was observed in mature embryos; however, no localization to pre-blastoderm embryos was detected (168). Other suitable peptides may be found by studying activities of symbionts or (viral/bacterial) pathogens capable of maternal transmission via transovarial endocytosis. In *A. pisum*, pathogenic cultured strains of *Serratia symbiotica* are endocytosed into early embryos (218). Understanding the mechanisms behind this could uncover new peptides or pathways to use for ReMOT control. An approach may be to inject multiple *Serratia* strains expressing GFP into *A. pisum* and analyse vertical transmission patterns. If some strains are more readily endocytosed, a genome wide association study (GWAS) could reveal the identity of candidate genes involved in the vertical transmission mechanism. Subsequently, CRISPR/Cas systems may enable the mutation of these genes in *Serratia*. CRISPR/Cas-mediated editing works in *Serratia* species (219), and most bacterial species possess endogenous CRISPR/Cas systems that may be utilised (219, 220). If the loss of function in a gene results in the loss of vertical transmission, it suggests that the gene product

plays a critical role. Functional analysis could then determine whether it is possible to identify a peptide that enables ReMOT control.

Alternative Cas editors may also enable better target selection in *M. persicae*. The *M. persicae* genome is A-T rich (140). Given this Cas12 enzymes may be more optimal for generating edits in aphid genomes, because Cas9 utilises a G-rich protospacer adjacent motif (PAM) for recognition and activation (58), while Cas12 enzymes use a T-rich PAM (71, 72). Cas12a has been well characterised to a point of commercial availability, including sgRNA synthesis services. Unlike Cas9, Cas12a induces double stranded staggered cuts. There are many members of the Cas12 family, including newly discovered mini-Cas enzymes such as Cas ϕ and AsCas12f (71, 72). These enzymes are significantly smaller than Cas9 and Cas12a, and therefore may be more readily deliverable to oocytes by ReMOT control and/or DIPA-CRISPR. Indeed, in this thesis, we have generated a BtKV-Cas ϕ chimera for use in *B. tabaci*, but further testing is required to establish its ability as a ReMOT-Control gene editor. The use of these enzymes in *M. persicae* may increase rate of uptake. We established that MpRV- and P2C-mCherry were both localized in early *M. persicae* embryos. However, we did not observe efficient gene editing, possibly suggesting that MpRV- and P2C-Cas9 were not internalized into the embryos via endocytosis. Smaller gene editors may be more efficiently internalized into embryos, thereby increasing the likelihood of inducing successful gene editing of germline cells.

Neither ReMOT nor DIPA-CRISPR systems have been employed to achieve targeted insertions through homology-directed repair (HDR) or other methods. In CRISPR-based

approaches, successful DNA insertions require the co-delivery of an HDR template containing the desired sequence alongside the CRISPR/Cas machinery. Following a Cas9-induced double-stranded break, the cell utilizes its endogenous HDR pathways, incorporating the provided template DNA into the repair process to precisely insert the desired sequence at the target location (221). This has been performed in insects using embryo injection methods, involving injection of DNA elements to deliver sgRNAs and Cas9 encoding sequences, as well as the HDR template (81). ReMOT control and DIPA-CRISPR rely on injection of preformed RNPs (90, 167). Co-injection of an HDR template with these systems is unlikely to yield any result, as the DNA is likely to be degraded, or not likely to colocalise with the Cas9. Prime editors could present a viable option for a ReMOT control based CRISPR knock-in system (222, 223). Prime editing utilises a fusion protein of a nickase (n)Cas9 and a reverse transcriptase. The prime editing guide (peg)RNA consists of the standard sgRNA with a 3' extension containing a template for DNA insertion. The nCas9 is mutated to remove one cutting domain, so upon activation it 'nicks' one strand of the DNA. The reverse transcriptase then uses the template on the pegRNA to repair the DNA (222). This approach has been successfully implemented in *D. melanogaster* (79) and could potentially be adapted for use as a preformed RNP in a ReMOT control or DIPA-CRISPR-like system. However, the increased size of the protein may reduce delivery efficiency. Further, a fusion consisting of a Vg-peptide, Cas9 and a reverse transcriptase could become unstable. As smaller Cas editors become more ubiquitous, a prime editing approach using Cas ϕ or AsCas12f may be feasible.

7.3 Conclusions

In this thesis, I have made significant progress in optimizing a ReMOT control-based gene editing system for *M. persicae*. Through the characterization of *M. persicae* vitellogenin (MpVg), I discovered that the canonical insect vitellogenin is absent in Aphidomorpha and Coccoidae species but present in psyllids. I identified MpVg as a vitellogenin-like protein that likely has retained some or all functions of the canonical vitellogenin. From MpVg, I derived the peptide MpRV, which I demonstrated to effectively localize mCherry to early embryos of *M. persicae*, comparable to the DmYPP1-derived peptide P2C (1).

I successfully expressed and purified recombinant MpRV- and P2C-Cas9 chimeras, validated their activity *in vitro*, and deployed them in ReMOT control injection experiments in *M. persicae*. While early results show promise, the gene-editing efficiency was low. I propose further optimizations, particularly in injection timing and aphid housing conditions, to improve efficiency.

Additionally, I have laid the groundwork for testing CasΦ as a potential alternative to Cas9 for ReMOT control in *B. tabaci*. These findings advance the development of targeted gene editing strategies in aphids and other insect species.

Appendix

Full BUSCO scores of 19 Aphid genomes

Table S1: BUSCO scores of the Aphid genomes used.

ID	BUSCO
<i>Myzus ligustri</i> Vg t1	C:97.4%[S:94.7%,D:2.7%],F:0.8%,M:1.8%,n:1066
<i>Phorodon humuli</i> Vg	C:98.1%[S:95.8%,D:2.3%],F:0.1%,M:1.8%,n:1066
<i>Myzus varians</i> Vg	C:97.4%[S:94.7%,D:2.7%],F:0.8%,M:1.8%,n:1066
<i>Myzus lythri</i> Vg	C:97.4%[S:94.6%,D:2.8%],F:0.7%,M:1.9%,n:1066
<i>Myzus cerasi</i> Vg	C:97.2%[S:93.9%,D:3.3%],F:0.8%,M:2.0%,n:1066
<i>Brachycaudus cardui</i> Vg	C:97.9%[S:95.6%,D:2.3%],F:0.7%,M:1.4%,n:1066
<i>Brachycaudus helichrysi</i> Vg	C:97.5%[S:94.7%,D:2.8%],F:0.9%,M:1.6%,n:1066
<i>Brachycaudus klugkisti</i> Vg t1	C:97.4%[S:95.0%,D:2.4%],F:0.9%,M:1.7%,n:1066
<i>Sitobion avenae</i> Vg t1	C:96.2%[S:93.2%,D:3.0%],F:1.1%,M:2.7%,n:1066
<i>Brevicoryne brassicae</i> Vg t1	C:97.3%[S:94.6%,D:2.7%],F:0.7%,M:2.0%,n:1066
<i>Sitobion miscanthi</i> Vg	C:96.0%[S:92.1%,D:3.9%],F:1.3%,M:2.7%,n:1066
<i>Metapolophium dirhodum</i> Vg	C:97.1%[S:93.6%,D:3.5%],F:0.7%,M:2.2%,n:1066
<i>Aphis thalictri</i> Vg	C:98.0%[S:94.4%,D:3.6%],F:0.8%,M:1.2%,n:1066
<i>Rhodalosiphum padi</i> Vg	C:97.6%[S:94.6%,D:3.0%],F:0.5%,M:1.9%,n:1066
<i>Aphis glycines</i> Vg	C:95.9%[S:91.4%,D:4.5%],F:1.0%,M:3.1%,n:1066
<i>Aphis fabae</i> Vg	C:97.4%[S:94.3%,D:3.1%],F:0.6%,M:2.0%,n:1066
<i>Aphis gossypii</i> Vg	C:97.1%[S:94.8%,D:2.3%],F:0.7%,M:2.2%,n:1066
<i>Myzus persicae</i> Vg	C:91.4%[S:88%,D:3.4%],F:0.7%,M:3.2%,n:1658
<i>Acyrosiphon pisum</i> Vg	C:97.6%[S:94.7%,D:2.9%],F:0.4%,M:2.0%,n:1066

Sequences used in the vitellogenin and vitellogenin-like phylogeny

Table S2: Vitellogenin and Vitellogenin-like sequences sourced from NCBI for use in a phylogeny (Figure 3.1).

Sequence ID	Sequence Name	Species	Common Name	Source
AAB72001.1	vitellogenin [Riptortus clavatus]	Riptortus clavatus	Bean Bug	NCBI
ADU04392.1	vitellogenin [Bemisia tabaci]	Bemisia tabaci	Silverleaf whitefly	NCBI
AGJ26478.1	vitellogenin [Laodelphax striatellus]	Laodelphax striatellus	Small brown planthopper	NCBI
AGV05363.1	vitellogenin [Nesidiocoris tenuis]	Nesidiocoris tenuis	Mirid bug	NCBI
AIA09041.1	vitellogenin 2, partial [Triatoma infestans]	Triatoma infestans	Kissing bug	NCBI
ALN70475.1	vitellogenin [Geocoris pallidipennis]	Geocoris pallidipennis	Bigeyed bug	NCBI
AOY34570.1	vitellogenin [Nephotettix virescens]	Nephotettix virescens	Green Paddy Leafhopper	NCBI
AQM52239.1	vitellogenin 1-like protein [Bactericera cockerelli]	Bactericera cockerelli	Potato psyllid	NCBI
ATY35166.1	Vg6-like protein [Bactericera cockerelli]	Bactericera cockerelli	Potato psyllid	NCBI
BAA85987.1	vitellogenin [Graptopsaltria nigrofuscata]	Graptopsaltria nigrofuscata	Large Brown Cicada	NCBI
BAA88075.1	vitellogenin-1 [Plautia stali]	Plautia stali	Oriental stinkbug	NCBI

BAG12118.1	vitellogenin [Lethocerus deyrollei]	Lethocerus deyrollei	Giant Water Bug	NCBI
BAJ33507.1	vitellogenin [Trigonotylus caelestialium]	Trigonotylus caelestialium	Rice Leaf Bug	NCBI
BAP87098.1	vitellogenin [Nilaparvata lugens]	Nilaparvata lugens	Brown Planthopper	NCBI
BAU68162.1	vitellogenin1 [Glaucias subpunctatus]	Glaucias subpunctatus	Stinkbug	NCBI
BtabMEAM1_Bta14071/1-1330	Vitellogenin-like 1 [Bemisia tabaci]	Bemisia tabaci	Silverleaf whitefly	Whitefly Genome Database
CAH1388709.1	unnamed protein product [Nezara viridula]	Nezara viridula	Southern green stink bug	NCBI
CAH1401941.1	unnamed protein product [Nezara viridula]	Nezara viridula	Southern green stink bug	NCBI
KAF6211930.1	hypothetical protein GE061_012447 [Apolygus lucorum]	Apolygus lucorum	small green plant bug	NCBI
KAF6216357.1	hypothetical protein GE061_000698 [Apolygus lucorum]	Apolygus lucorum	small green plant bug	NCBI
KAG8258928.1	hypothetical protein J6590_021722 [Homalodisca vitripennis]	Homalodisca vitripennis	Glassy-winged sharpshooter	NCBI
KAI5699898.1	hypothetical protein M8J75_010797 [Diaphorina citri]	Diaphorina citri	Asian citrus psyllid	NCBI
KAI5749250.1	hypothetical protein M8J76_005907 [Diaphorina citri]	Diaphorina citri	Asian citrus psyllid	NCBI
KAK9500937.1	hypothetical protein O3 M35_002098 [Rhynocoris fuscipes]	Rhynocoris fuscipes	Assassin bug	NCBI
KAL1455371.1	hypothetical protein WDU94_009470 [Cyamophila willieti]	Cyamophila willieti		NCBI

QDD67294.1	vitellogenin [Sogatella furcifera]	Sogatella furcifera	Whitebacked planthopper	NCBI
QFQ33313.1	vitellogenin-like protein 1 [Nilaparvata lugens]	Nilaparvata lugens	Brown Planthopper	NCBI
QXD38625.1	vitellogenin-1 [Eurygaster maura]	Eurygaster maura	Tortoise bug	NCBI
RZF40029.1	hypothetical protein LSTR_LSTR002432 [Laodelphax striatellus]	Laodelphax striatellus	Small brown planthopper	NCBI
UOL49140.1	vitellogenin1-like protein [Bactericera trigonica]	Bactericera trigonica		NCBI
WTM25993.1	vitellogenin [Empoasca flavescens]	Empoasca flavescens	Castor leafhopper	NCBI
XCA47670.1	vitellogenin [Recilia dorsalis]	Recilia dorsalis	Zigzag leafhopper	NCBI
XP_014239787.1	uncharacterized protein LOC106661112 isoform X1 [Cimex lectularius]	Cimex lectularius	Bedbug	NCBI
XP_014239788.1	uncharacterized protein LOC106661112 isoform X2 [Cimex lectularius]	Cimex lectularius	Bedbug	NCBI
XP_014261592.1	vitellogenin-1-like [Cimex lectularius]	Cimex lectularius	Bedbug	NCBI
XP_014270480.1	vitellogenin [Halyomorpha halys]	Halyomorpha halys	Brown marmorated stink bug	NCBI
XP_014270696.1	vitellogenin-4 [Halyomorpha halys]	Halyomorpha halys	Brown marmorated stink bug	NCBI
XP_050534020.1	uncharacterized protein LOC126901548 isoform X1 [Daktulosphaira vitifoliae]	Daktulosphaira vitifoliae	Grape phylloxera	NCBI
XP_050534022.1	uncharacterized protein LOC126901548 isoform X2 [Daktulosphaira vitifoliae]	Daktulosphaira vitifoliae	Grape phylloxera	NCBI

Appendix

XP_054269378.1	vitellogenin-1-like [Macrosteles quadrilineatus]	Macrosteles quadrilineatus	Aster leafhopper	NCBI
XP_054270572.1	uncharacterized protein LOC128991582 [Macrosteles quadrilineatus]	Macrosteles quadrilineatus	Aster leafhopper	NCBI
XP_066907447.1	vitellogenin-2 isoform X2 [Halyomorpha halys]	Halyomorpha halys	Brown marmorated stink bug	NCBI

Table S3: Aphid vitellogenin sequences used in a phylogeny (Figure 3.1) and the genome annotations to which they belong. All were downloaded internally on the Norwich Bioscience Institutes (NBI) High Performance Computing (HPC) cluster. Most are published in Mathers et al., 2022 (151). *M. persicae* genome clone O v2.1 is published in Liu et al., 2024 (154). *A. pisum* v1.0 is published in Mathers et al., 2021 (140). *M. cerasi* genome assembly is published in Thorpe et al., 2019 (152) and the annotation was published in Mathers et al., 2022 (151).

Gene ID	ID	Clone	Annotation	Genome and annotation source
g13794.t1	<i>Myzus ligustri</i> Vg t1	JIC	v1.1	Mathers et al., 2022
g13794.t2	<i>Myzus ligustri</i> Vg t2	JIC	v1.1	Mathers et al., 2022
g16166.t1	<i>Phorodon humuli</i> Vg	JIC	v2	Mathers et al., 2022
g6903.t1	<i>Myzus varians</i> Vg	JIC	v1.1	Mathers et al., 2022
g3879.t1	<i>Myzus lythri</i> Vg	JIC	v1.1	Mathers et al., 2022
g1419.t1	<i>Myzus cerasi</i> Vg	Thorpe	v1.2	Thorpe et al., 2018; Mathers et al., 2022
g700.t1	<i>Brachycaudus cardui</i> Vg	JIC	v1.1	Mathers et al., 2022
g791.t1	<i>Brachycaudus helichrysi</i> Vg	JIC	v1.1	Mathers et al., 2022
g2450.t1	<i>Brachycaudus klugkisti</i> Vg t1	JIC	v1.1	Mathers et al., 2022
g2450.t2	<i>Brachycaudus klugkisti</i> Vg t2	JIC	v1.1	Mathers et al., 2022
g10513.t1	<i>Sitobion avenae</i> Vg t1	JIC1	v2.1	Mathers et al., 2022; Mathers et al., 2023
g10513.t2	<i>Sitobion avenae</i> Vg t2	JIC1	v2.1	Mathers et al., 2022; Mathers et al., 2023
g2994.t1	<i>Brevicoryne brassicae</i> Vg t1	JIC	v2	Mathers et al., 2022
g2994.t2	<i>Brevicoryne brassicae</i> Vg t2	JIC	v2	Mathers et al., 2022
g599.t1	<i>Sitobion miscanthi</i> Vg	JIC	v2	Mathers et al., 2022
g21201.t1	<i>Metapolophium dirhodum</i> Vg	JIC1	v1.1	Mathers et al., 2022; Mathers et al., 2023
g12538.t1	<i>Aphis thalictri</i> Vg	JIC	v1	Mathers et al., 2022
g9247.t1	<i>Rhodalosiphum padi</i> Vg	JIC1	v1	Mathers et al., 2022; Mathers et al., 2023

Appendix

g1810.t1	<i>Aphis glycines</i> Vg	JIC1	v1.0	Mathers et al., 2022
g13804.t1	<i>Aphis fabae</i> Vg	JIC1	v2	Mathers et al., 2022
g22465.t1	<i>Aphis gosypii</i> Vg	1033E	v1	Mathers et al., 2022
MYZPE13164_O_Elv2.1_0213490.1	<i>Myzus persicae</i> Vg	O	v2.1	Liu et al., 2024
g21334.t1	<i>Acyrosiphon pisum</i> Vg	JIC1	v1.0	Mathers et al., 2021

General code for command line BLAST

Command line BLAST was used on multiple occasions, including to find MpVg in the clone O v2.1 annotation, and *Acyrtosiphon pisum* Vg in the JIC1 v1.0 annotation. The code below was used. This code was also used to identify the *M. persicae* *white* gene.

```
#source blast
```

```
source blast+-2.2.30
```

```
# Description: The following command creates a BLAST database using the input FASTA file.
```

```
makeblastdb -in genome annotation -dbtype nucl/prot
```

```
# Description: The following command performs a BLAST search using the specified query sequence against the previously created BLAST database. The output is saved to a file named "desired output file.txt".
```

```
Blast(n/p) -query 'query file' -db 'db file' -out 'desired output file.txt'
```

BLAST for Vg sequences in compiled proteomes of aphid genomes

Presented here is the code used to blast the compiled proteomes of aphids. The compiled proteomes are in the file 'Aphid_prot_db.fa' and were compiled as described in Chapter 2. The file "MpVg_O_2_1_pep.fa" contained the MpVg clone O v2.1 sequence MYZPE13164_O_Elv2.1_0213490.1.

#working directory, date

```
[jamesr@NBI-HPC interactive MpVg_vs_aphids_blast]$ pwd
/hpc-home/jamesr/Vg_phylogeny/MpVg_vs_aphids_blast
[jamesr@NBI-HPC interactive MpVg_vs_aphids_blast]$ date
Wed 31 May 13:23:51 BST 2023
```

#initial contents of working directory ##Aphid_prot_db.fa compiles the pep seqs from Tom's aphid genomes

```
[jamesr@NBI-HPC MpVg_vs_aphids_blast]$ ls
Aphid_prot_db.fa MpVg_O_2_1_pep.fa
```

#source blast

```
[jamesr@NBI-HPC interactive MpVg_vs_aphids_blast]$ source blast+-2.2.30
```

#make the blast db

```
[jamesr@NBI-HPC interactive MpVg_vs_aphids_blast]$ makeblastdb -in
Aphid_prot_db.fa -dbtype prot
```

Building a new DB, current time: 05/31/2023 13:23:02

New DB name: Aphid_prot_db.fa

New DB title: Aphid_prot_db.fa

Sequence type: Protein

Keep Linkouts: T

Keep MBits: T

Maximum file size: 1000000000B

Adding sequences from FASTA; added 462454 sequences in 13.8982 seconds.

#check contents of working directory for blast db

```
[jamesr@NBI-HPC interactive MpVg_vs_aphids_blast]$ ls
Aphid_prot_db.fa Aphid_prot_db.fa.pin MpVg_O_2_1_pep.fa
Aphid_prot_db.fa.phr Aphid_prot_db.fa.psq
```

#run blastp

```
[jamesr@NBI-HPC interactive MpVg_vs_aphids_blast]$ blastp -query
MpVg_O_2_1_pep.fa -db Aphid_prot_db.fa -out MpVg_vs_aphids_blast.txt
```

#Check contents of working directory for blast.txt file

```
[jamesr@NBI-HPC interactive MpVg_vs_aphids_blast]$ ls
Aphid_prot_db.fa Aphid_prot_db.fa.pin MpVg_O_2_1_pep.fa
Aphid_prot_db.fa.phr Aphid_prot_db.fa.psq MpVg_vs_aphids_blast.txt
```

```
#view blast.tct file
```

```
[jamesr@NBI-HPC interactive MpVg_vs_aphids_blast]$ less MpVg_vs_aphids_blast.txt
```

Alphafold script

The script below was used to perform Alphafold 2 analysis on the NBI HPC. This was provided by Dr. Rea Antoniou-Kourounioti.

```
#!/bin/bash
#SBATCH --job-name=
#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=32
#SBATCH --partition=jic-long
#SBATCH --mem 80G
#SBATCH --time=20-00:00
#SBATCH -o alphafold_%j_%N.out
#SBATCH -e alphafold_%j_%N.err
#SBATCH --mail-type=ALL
#SBATCH --mail-user=

pwd
hostname
date

filename=filename.fasta # INPUT FILENAME CAN BE GIVEN ON COMMAND LINE OR
INSTEAD REPLACE "$1" HERE
DATA_DIR="/jic/scratch/projects/AlphaFold" # DO NOT CHANGE
INPUT_DIR= # ADD DIRECTORY WHERE PROTEIN SEQUENCE IS STORED
OUTPUT_DIR= # ADD DIRECTORY WHERE YOU WANT ALPHAFOLD OUTPUT TO BE
MADE

source package 9a272fce-308f-46c4-892d-c80a3797dc0f

srun run_alphafold.py --fasta_paths=${filename} \
--data_dir=${DATA_DIR} \
--output_dir=${OUTPUT_DIR} \
--model_names=model_1,model_2,model_3,model_4,model_5 \
--max_template_date=2021-10-06 \
--preset=full_dbs \
--
bfd_database_path=${DATA_DIR}/bfd/bfd_metaclust_clu_complete_id30_c90_final_se
q.sorted_opt \
--
uniclust30_database_path=${DATA_DIR}/uniclust30/uniclust30_2018_08/uniclust30_2
018_08 \
--uniref90_database_path=${DATA_DIR}/uniref90/uniref90.fasta \
--mgnify_database_path=${DATA_DIR}/mgnify/mgy_clusters_2018_12.fa \
--pdb70_database_path=${DATA_DIR}/pdb70/pdb70 \
```

```
--template_mmcif_dir=${DATA_DIR}/pdb_mmcif/mmcif_files \  
--obsolete_pdbs_path=${DATA_DIR}/pdb_mmcif/obsolete.dat
```

date

R script for generation of pLDDT plots

This script was written primarily by Dr. Sam Mugford, and amended by myself.

```
# Install svDialogs package if not already installed
if (!requireNamespace("svDialogs", quietly = TRUE)) {
  install.packages("svDialogs")
}

# Load necessary libraries
library(ggplot2)
library(dplyr)
library(svDialogs)

# Function to read PDB file and extract data
read_pdb <- function(file) {
  cat("Reading file:", file, "\n")

  # Read the file into a character vector
  lines <- readLines(file)

  # Add a space between the letter and the number
  lines <- gsub("([A-Za-z])(\\d+)", "\\1 \\2", lines)

  # Convert the adjusted lines into a data frame
  pdb_data <- read.table(text = lines, header = FALSE, stringsAsFactors = FALSE, fill =
TRUE)

  # Check if the expected columns exist
  if (ncol(pdb_data) < 11) {
    stop("The file does not contain the expected number of columns: ", file)
  }

  # Check for numeric conversion issues and filter valid rows
  valid_rows <- !is.na(as.numeric(pdb_data$V6)) & !is.na(as.numeric(pdb_data$V11))

  if (sum(valid_rows) == 0) {
    stop("No valid numeric data found in the required columns: ", file)
  }

  # Extract the necessary columns (amino acid position and PDDL score)
  pdb_df <- data.frame(
    Position = as.numeric(pdb_data$V6[valid_rows]),
    PDDL_Score = as.numeric(pdb_data$V11[valid_rows]),
    Filename = basename(file)
  )
}
```

```

cat("File processed successfully:", file, "\n")
cat("Number of positions read:", nrow(pdb_df), "\n") # Debugging statement
cat("Unique positions:", length(unique(pdb_df$Position)), "\n") # Debugging statement

return(pdb_df)
}

# Function to read all PDB files in a folder and combine into one data frame
read_pdb_folder <- function(folder) {
  pdb_files <- list.files(folder, pattern = "\\..pdb$", full.names = TRUE)

  if (length(pdb_files) == 0) {
    stop("No PDB files found in the selected folder.")
  }

  pdb_data <- do.call(rbind, lapply(pdb_files, read_pdb))

  cat("Total number of positions read from all files:", nrow(pdb_data), "\n") # Debugging
statement
  cat("Total unique positions:", length(unique(pdb_data$Position)), "\n") # Debugging
statement

  return(pdb_data)
}

# Plotting function
plot_pddlt <- function(data, folder_name) {
  ggplot(data, aes(x = Position, y = PDDL_Score, color = Filename)) +
    geom_line() +
    labs(title = paste("pLDDT Score vs. Amino Acid Position (", folder_name, ")"),
         x = "Amino Acid Position",
         y = "pLDDT Score") +
    theme_minimal() +
    theme(legend.title = element_blank()) +
    scale_x_continuous(expand = c(0, 0)) + # Removes padding around the x-axis
    scale_y_continuous(expand = c(0, 0)) # Removes padding around the y-axis
}

# Main script
folder_path <- dlg_dir()$res

if (is.null(folder_path) || folder_path == "") {
  stop("No folder selected. Please select a folder and try again.")
}

folder_name <- basename(folder_path)

```

```
pdb_data <- read_pdb_folder(folder_path)

if (nrow(pdb_data) == 0) {
  stop("No data to plot. Please check the PDB files in the folder.")
}

cat("Number of unique positions in the data:", length(unique(pdb_data$Position)), "\n")
# Debugging statement
print(head(pdb_data)) # Print first few rows of the combined data for inspection

pddlt_plot <- plot_pddlt(pdb_data, folder_name)

ggsave(paste0(folder_name, "_PDDLTL_Score_vs_Position.png"), pddlt_plot, width = 15,
height = 10)

print(pddlt_plot)
```

Scripts for morph specific RNASeq analysis in R

The script to provide the JH pathway Box plot is below (Figure 2.8):

```
library(readxl)
library(ggplot2)
library(ggpubr)
library(dplyr)
library(broom)

#load in vertical tables
JH_Pathway <- read_excel(path
="Vg_Expression_Supplementary_Table_6_MperO_morphs_expression_TPM.xlsx",
sheet = 'JH_Pathway')

#Classify order for JH pathwaY
JH_Pathway <- within(JH_Pathway, Gene_Name <- factor(Gene_Name, levels =
c('MpVg', 'MpVgR', 'Met', 'SRC', 'FoxO_X1', 'FoxO_X2', 'LCMT1', 'PP2A', 'Cdc6', 'Cdc2',
'MCM3', 'MCM4', 'MCM7', 'Orc5')))

my_comparisons <- list( c('FA','FW'),c('FW', 'MA'), c('MA', 'NY'), c('FA', 'MA'), c('FW',
'NY'), c('FA', 'NY'))

#load packages for ANOVA + Tukey
library(tidyr)
library(purrr)
library(broom)

#Run ANOVA + Tukey
results <- JH_Pathway %>%
  group_by(Gene_Name) %>%
  nest() %>%
  mutate(
    model = map(data, ~ aov(Expression ~ Morph, data = .x)),
    anova = map(model, tidy),          # Tidy ANOVA results
    tukey = map(model, ~ tidy(TukeyHSD(.x))) # Tidy TukeyHSD results
  )

#Generate ANOVA results
anova_results <- results %>%
  select(Gene_Name, anova) %>%
  unnest(anova) %>%
  filter(term == "Morph")

#generate Tukey results
tukey_results <- results %>%
```

```

select(Gene_Name, tukey) %>%
unnest(tukey)

# Filter significant comparisons
sig_tukey <- tukey_results %>%
  filter(adj.p.value < 0.05)

# Create flat tidy table of only significant comparisons
sig_tukey <- tukey_results %>%
  filter(adj.p.value < 0.05) %>%
  separate(contrast, into = c("group1", "group2"), sep = "-") %>%
  mutate(
    p.adj = signif(adj.p.value, 3),
    label = paste0("p = ", signif(adj.p.value, 3)),
    y.position = max(estimate, na.rm = TRUE) + 2 # Adjust height above points
  )

#Generate JH-Pathway plot
ggboxplot(JH_Pathway, x = "Morph", y = "Expression", outlier.shape = NA, alpha = 0.5,
facet.by = 'Gene_Name') +
  theme(text = element_text(size=20)) +
  geom_jitter(aes(color = Morph),width = 0.25, height = 0.1, cex=3)+
  theme(text = element_text(size=20)) +
  scale_y_continuous(name="Expression (TPM)") +
  xlab("Aphid Morph") +
  scale_color_manual(values=c("#E69F00", "#56B4E9", "#009E73", "#F0E442",
"#D55E00")) +
  #facet_wrap(nrow = 3, ncol = 5, Gene_Name~.)+
  theme(text=element_text(size=15))

```

The script to provide the ECR pathway Box plot (Figure 2.10) is below:

```

library(readxl)
library(ggplot2)
library(ggpubr)
library(dplyr)

#load in vertical tables
ECR_Pathway <- read_excel(path
="Vg_Expression_Supplementary_Table_6_MperO_morphs_expression_TPM.xlsx",
sheet = 'ECR_Pathway')

#Classify order for ecr pathwaY

```

```
ECR_Pathway <- within(ECR_Pathway, Gene_Name <- factor(Gene_Name, levels =
c('MpVg', 'MpVgR', 'EcR', 'Ultraspiracle', 'BC core protein-like', 'E75', 'E78C', 'HR3')))
```

```
my_comparisons <- list( c('FA','FW'),c('FW', 'MA'), c('MA', 'NY'), c('FA', 'MA'), c('FW',
'NY'), c('FA', 'NY'))
```

```
#load packages for ANOVA + Tukey
library(tidyr)
library(purrr)
library(broom)
```

```
#Run ANOVA + Tukey
results <- ECR_Pathway %>%
  group_by(Gene_Name) %>%
  nest() %>%
  mutate(
    model = map(data, ~ aov(Expression ~ Morph, data = .x)),
    anova = map(model, tidy),          # Tidy ANOVA results
    tukey = map(model, ~ tidy(TukeyHSD(.x))) # Tidy TukeyHSD results
  )
```

```
#Generate ANOVA results
anova_results <- results %>%
  select(Gene_Name, anova) %>%
  unnest(anova) %>%
  filter(term == "Morph")
```

```
#generate Tukey results
tukey_results <- results %>%
  select(Gene_Name, tukey) %>%
  unnest(tukey)
```

```
# Filter significant comparisons
sig_tukey <- tukey_results %>%
  filter(adj.p.value < 0.05)
```

```
# Create flat tidy table of only significant comparisons
sig_tukey <- tukey_results %>%
  filter(adj.p.value < 0.05) %>%
  separate(contrast, into = c("group1", "group2"), sep = "-") %>%
  mutate(
    p.adj = signif(adj.p.value, 3),
    label = paste0("p = ", signif(adj.p.value, 3)),
    y.position = max(estimate, na.rm = TRUE) + 2 # Adjust height above points
```

)

```
#Generate ecr-Pathway plot
ggboxplot(ECR_Pathway, x = "Morph", y = "Expression", outlier.shape = NA, alpha = 0.5,
facet.by = 'Gene_Name') +
  theme(text = element_text(size=20)) +
  geom_jitter(aes(color = Morph),width = 0.25, height = 0.1, cex=3)+
  theme(text = element_text(size=20)) +
  scale_y_continuous(name="Expression (TPM)") +
  xlab("Aphid Morph") +
  scale_color_manual(values=c("#E69F00", "#56B4E9", "#009E73", "#F0E442")) +
  #stat_compare_means(method = 't.test', comparisons = my_comparisons, label =
'p.signif') +
  #facet_wrap(nrow = 2, ncol = 4, Gene_Name~.)+
  theme(text=element_text(size=15))
```

The script used to generate both the JH and 20E pathway heatmaps (Figure 2.9 and 2.11 respectively) is below:

```
#load your packages#
```

```
library("readxl")
library("dplyr")
library("pheatmap")
library("tidyverse")
```

```
##JH
```

```
#import and reformat your excel sheet#
```

```
JH_df <-
```

```
read_excel("Vg_Expression_Supplementary_Table_6_MperO_morphs_expression_TPM.
xlsx", sheet = "JH_Pathway") #you'll need your own file path
```

```
JH_df <- JH_df %>% select(-Morph)
```

```
JH_df <- JH_df %>%
```

```
  pivot_wider(
```

```
    names_from = Individ, # Column names come from Individ
```

```
    values_from = Expression, # Values come from Expression
```

```
    id_cols = Gene_Name
```

```
  ) %>%
```

```
  column_to_rownames("Gene_Name")
```

```
MORPH_META <- data.frame(row.names = colnames(JH_df),
```

```
  Morph = c("female apterous", "female apterous", "female apterous", "female
apterous", "female apterous", "female apterous",
```

```
            "female winged", "female winged", "female winged", "female
winged", "female winged", "female winged",
```

```
            "male", "male", "male", "male", "male", "male",
```

```
"nymph","nymph","nymph","nymph","nymph","nymph")) #not sure if
these are right but just for display
```

```
my_colour = list(
  Morph = c('female apterous' = "#E69F00", 'female winged' = "#56B4E9", 'male' =
"#009E73" , 'nymph' = "#F0E442")) #pick your own colours, maybe make them match
your box plots?
```

```
pheatmap(JH_df,
  #color = brewer.pal(n = 7, name = "PiYG"),
  #filename = "~/Desktop/Top_100_CathB_DEG_labelled.pdf", #uncomment this if
you want to save it rather than just output it to the plots pane
  fontsize = 10,
  fontsize_row = 8,
  fontsize_col = 8,
  annotation_col = MORPH_META,
  #annotation_row = gene_meta,
  annotation_colors = my_colour,
  show_colnames = T,
  show_rownames = T,
  border_color = "grey",
  height = 10, width = 20,
  legend = T,
  annotation_legend = T,
  cluster_cols = F,
  clustering_distance_rows = "correlation",
  treeheight_row = 20,
  treeheight_col = 20,
  cellheight = 20, cellwidth = 15) #make your heatmap
```

```
##ECR
ECR_df <-
read_excel("Vg_Expression_Supplementary_Table_6_MperO_morphs_expression_TPM.
xlsx", sheet = "ECR_Pathway")
ECR_df <- ECR_df %>% select(-Morph)
ECR_df <- ECR_df %>%
  pivot_wider(
    names_from = Individ, # Column names come from Individ
    values_from = Expression, # Values come from Expression
    id_cols = Gene_Name
  ) %>%
  column_to_rownames("Gene_Name")
```

```
MORPH_META <- data.frame(row.names = colnames(ECR_df),
```



```

Morph = c("female apterous", "female apterous", "female apterous", "female
apterous", "female apterous", "female apterous",
        "female winged", "female winged", "female winged", "female
winged", "female winged", "female winged",
        "male", "male", "male", "male", "male", "male",
        "nymph", "nymph", "nymph", "nymph", "nymph", "nymph"))

```

```

my_colour = list(
  Morph = c('female apterous' = "#E69F00", 'female winged' = "#56B4E9", 'male' =
"#009E73" , 'nymph' = "#F0E442")) #pick your own colours, maybe make them match
your box plots?

```

```

pheatmap(ECR_df,
  #color = brewer.pal(n = 7, name = "PiYG"),
  #filename = "~/Desktop/Top_100_CathB_DEG_labelled.pdf", #uncomment this if
you want to save it rather than just output it to the plots pane
  fontsize = 10,
  fontsize_row = 8,
  fontsize_col = 8,
  annotation_col = MORPH_META,
  #annotation_row = gene_meta,
  annotation_colors = my_colour,
  show_colnames = T,
  show_rownames = T,
  border_color = "grey",
  height = 10, width = 20,
  legend = T,
  annotation_legend = T,
  cluster_cols = F,
  clustering_distance_rows = "correlation",
  treeheight_row = 20,
  treeheight_col = 20,
  cellheight = 20, cellwidth = 15)

```

Tukey statistics for morph specific RNASeq analysis

Table S 4: Tukey results for JH Pathway associated genes.

Gene Name	Term	Contrast	null.value	estimate	conf.low	conf.high	adj.p.value
<i>MpVg</i>	Morph	FW-FA	0	1.277199	-3.61021	6.164603	8.83E-01
<i>MpVg</i>	Morph	MA-FA	0	15.09738	10.20998	19.98478	1.96E-07
<i>MpVg</i>	Morph	NY-FA	0	-9.01797	-13.9054	-4.13057	2.56E-04
<i>MpVg</i>	Morph	MA-FW	0	13.82018	8.932777	18.70759	7.79E-07
<i>MpVg</i>	Morph	NY-FW	0	-10.2952	-15.1826	-5.40777	5.03E-05
<i>MpVg</i>	Morph	NY-MA	0	-24.1154	-29.0028	-19.2279	6.33E-11
<i>MpVgR</i>	Morph	FW-FA	0	-0.02046	-0.13736	0.096432	9.60E-01
<i>MpVgR</i>	Morph	MA-FA	0	0.239744	0.12285	0.356637	7.07E-05
<i>MpVgR</i>	Morph	NY-FA	0	0.023942	-0.09295	0.140835	9.39E-01
<i>MpVgR</i>	Morph	MA-FW	0	0.260205	0.143312	0.377099	2.43E-05
<i>MpVgR</i>	Morph	NY-FW	0	0.044403	-0.07249	0.161297	7.15E-01
<i>MpVgR</i>	Morph	NY-MA	0	-0.2158	-0.3327	-0.09891	2.54E-04
<i>Met</i>	Morph	FW-FA	0	30.94698	22.42191	39.47206	1.39E-08
<i>Met</i>	Morph	MA-FA	0	55.62206	47.09699	64.14713	3.47E-13
<i>Met</i>	Morph	NY-FA	0	-14.8148	-23.3399	-6.28977	5.05E-04
<i>Met</i>	Morph	MA-FW	0	24.67507	16.15	33.20015	5.44E-07
<i>Met</i>	Morph	NY-FW	0	-45.7618	-54.2869	-37.2367	1.38E-11
<i>Met</i>	Morph	NY-MA	0	-70.4369	-78.962	-61.9118	2.38E-14
<i>SRC</i>	Morph	FW-FA	0	3.817035	-0.50727	8.141338	9.57E-02
<i>SRC</i>	Morph	MA-FA	0	10.03669	5.71239	14.36099	1.38E-05
<i>SRC</i>	Morph	NY-FA	0	-1.71111	-6.03541	2.613193	6.89E-01
<i>SRC</i>	Morph	MA-FW	0	6.219657	1.895355	10.54396	3.41E-03
<i>SRC</i>	Morph	NY-FW	0	-5.52815	-9.85245	-1.20384	9.36E-03
<i>SRC</i>	Morph	NY-MA	0	-11.7478	-16.0721	-7.4235	1.43E-06
<i>MCM3</i>	Morph	FW-FA	0	-7.67812	-11.6774	-3.67885	1.60E-04
<i>MCM3</i>	Morph	MA-FA	0	-18.7135	-22.7128	-14.7142	1.64E-10
<i>MCM3</i>	Morph	NY-FA	0	-10.0957	-14.0949	-6.09641	4.22E-06
<i>MCM3</i>	Morph	MA-FW	0	-11.0354	-15.0347	-7.03612	1.13E-06
<i>MCM3</i>	Morph	NY-FW	0	-2.41756	-6.41683	1.581712	3.54E-01
<i>MCM3</i>	Morph	NY-MA	0	8.617827	4.618558	12.6171	3.74E-05
<i>MCM4</i>	Morph	FW-FA	0	-5.02113	-8.27099	-1.77128	1.73E-03
<i>MCM4</i>	Morph	MA-FA	0	-15.8823	-19.1322	-12.6325	7.52E-11
<i>MCM4</i>	Morph	NY-FA	0	-6.79041	-10.0403	-3.54056	5.58E-05
<i>MCM4</i>	Morph	MA-FW	0	-10.8612	-14.1111	-7.61135	5.49E-08
<i>MCM4</i>	Morph	NY-FW	0	-1.76928	-5.01913	1.480578	4.43E-01
<i>MCM4</i>	Morph	NY-MA	0	9.091925	5.842071	12.34178	9.17E-07
<i>MCM7</i>	Morph	FW-FA	0	-11.0519	-16.4315	-5.67236	6.92E-05
<i>MCM7</i>	Morph	MA-FA	0	-29.2558	-34.6354	-23.8762	1.09E-11
<i>MCM7</i>	Morph	NY-FA	0	-15.3172	-20.6968	-9.9376	7.01E-07

Appendix

MCM7	Morph	MA-FW	0	-18.2039	-23.5834	-12.8243	4.47E-08
MCM7	Morph	NY-FW	0	-4.26523	-9.64482	1.11435	1.52E-01
MCM7	Morph	NY-MA	0	13.93863	8.559046	19.31821	2.89E-06
Cdc6	Morph	FW-FA	0	-2.48244	-4.1661	-0.79878	2.71E-03
Cdc6	Morph	MA-FA	0	-4.82725	-6.51091	-3.1436	6.30E-07
Cdc6	Morph	NY-FA	0	-2.92656	-4.61022	-1.2429	5.03E-04
Cdc6	Morph	MA-FW	0	-2.34481	-4.02847	-0.66116	4.56E-03
Cdc6	Morph	NY-FW	0	-0.44412	-2.12778	1.239537	8.80E-01
Cdc6	Morph	NY-MA	0	1.900693	0.217035	3.584352	2.34E-02
Orc5	Morph	FW-FA	0	-1.79692	-2.49413	-1.09972	3.12E-06
Orc5	Morph	MA-FA	0	-3.87466	-4.57187	-3.17745	7.34E-12
Orc5	Morph	NY-FA	0	-1.56214	-2.25935	-0.86494	2.23E-05
Orc5	Morph	MA-FW	0	-2.07774	-2.77494	-1.38053	3.45E-07
Orc5	Morph	NY-FW	0	0.234782	-0.46243	0.931989	7.83E-01
Orc5	Morph	NY-MA	0	2.312517	1.61531	3.009724	6.21E-08
Cdc2	Morph	FW-FA	0	-12.4437	-18.9086	-5.9789	1.55E-04
Cdc2	Morph	MA-FA	0	-33.9233	-40.3881	-27.4584	2.09E-11
Cdc2	Morph	NY-FA	0	-13.3725	-19.8373	-6.90765	6.35E-05
Cdc2	Morph	MA-FW	0	-21.4795	-27.9444	-15.0147	6.04E-08
Cdc2	Morph	NY-FW	0	-0.92875	-7.39359	5.536097	9.77E-01
Cdc2	Morph	NY-MA	0	20.55076	14.08592	27.01561	1.24E-07
FoxO_X1	Morph	FW-FA	0	17.12515	2.895771	31.35452	1.49E-02
FoxO_X1	Morph	MA-FA	0	68.46571	54.23634	82.69509	9.95E-11
FoxO_X1	Morph	NY-FA	0	-14.9205	-29.1499	-0.69113	3.78E-02
FoxO_X1	Morph	MA-FW	0	51.34057	37.11119	65.56994	1.54E-08
FoxO_X1	Morph	NY-FW	0	-32.0457	-46.275	-17.8163	2.08E-05
FoxO_X1	Morph	NY-MA	0	-83.3862	-97.6156	-69.1568	2.73E-12
FoxO_X2	Morph	FW-FA	0	23.29152	11.45545	35.12759	1.18E-04
FoxO_X2	Morph	MA-FA	0	76.6558	64.81973	88.49187	4.00E-13
FoxO_X2	Morph	NY-FA	0	5.787992	-6.04808	17.62406	5.32E-01
FoxO_X2	Morph	MA-FW	0	53.36428	41.52821	65.20035	3.19E-10
FoxO_X2	Morph	NY-FW	0	-17.5035	-29.3396	-5.66746	2.63E-03
FoxO_X2	Morph	NY-MA	0	-70.8678	-82.7039	-59.0317	1.82E-12
LCMT1	Morph	FW-FA	0	-3.1031	-4.83633	-1.36987	3.61E-04
LCMT1	Morph	MA-FA	0	-3.19143	-4.92466	-1.4582	2.62E-04
LCMT1	Morph	NY-FA	0	-4.49075	-6.22398	-2.75752	2.89E-06
LCMT1	Morph	MA-FW	0	-0.08832	-1.82155	1.644907	9.99E-01
LCMT1	Morph	NY-FW	0	-1.38764	-3.12087	0.345586	1.46E-01
LCMT1	Morph	NY-MA	0	-1.29932	-3.03255	0.433909	1.88E-01
PP2A	Morph	FW-FA	0	6.121524	0.366383	11.87667	3.46E-02
PP2A	Morph	MA-FA	0	37.18738	31.43224	42.94252	4.18E-13
PP2A	Morph	NY-FA	0	-4.40241	-10.1576	1.352727	1.74E-01
PP2A	Morph	MA-FW	0	31.06586	25.31071	36.821	1.25E-11

Appendix

PP2A	Morph	NY-FW	0	-10.5239	-16.2791	-4.7688	2.84E-04
PP2A	Morph	NY-MA	0	-41.5898	-47.3449	-35.8347	6.35E-14

Table S 5: Tukey results for 20E pathway associated genes.

Gene Name	Term	Contrast	null.value	estimate	conf.low	conf.high	adj.p.value
<i>MpVg</i>	Morph	FW-FA	0	1.277199	-3.61021	6.164603	8.83E-01
<i>MpVg</i>	Morph	MA-FA	0	15.09738	10.20998	19.98478	1.96E-07
<i>MpVg</i>	Morph	NY-FA	0	-9.01797	-13.9054	-4.13057	2.56E-04
<i>MpVg</i>	Morph	MA-FW	0	13.82018	8.932777	18.70759	7.79E-07
<i>MpVg</i>	Morph	NY-FW	0	-10.2952	-15.1826	-5.40777	5.03E-05
<i>MpVg</i>	Morph	NY-MA	0	-24.1154	-29.0028	-19.2279	6.33E-11
<i>MpVgR</i>	Morph	FW-FA	0	-0.02046	-0.13736	0.096432	9.60E-01
<i>MpVgR</i>	Morph	MA-FA	0	0.239744	0.12285	0.356637	7.07E-05
<i>MpVgR</i>	Morph	NY-FA	0	0.023942	-0.09295	0.140835	9.39E-01
<i>MpVgR</i>	Morph	MA-FW	0	0.260205	0.143312	0.377099	2.43E-05
<i>MpVgR</i>	Morph	NY-FW	0	0.044403	-0.07249	0.161297	7.15E-01
<i>MpVgR</i>	Morph	NY-MA	0	-0.2158	-0.3327	-0.09891	2.54E-04
<i>EcR</i>	Morph	FW-FA	0	6.756302	2.948053	10.56455	4.01E-04
<i>EcR</i>	Morph	MA-FA	0	13.59782	9.789571	17.40607	1.83E-08
<i>EcR</i>	Morph	NY-FA	0	-0.59908	-4.40733	3.209166	9.71E-01
<i>EcR</i>	Morph	MA-FW	0	6.841518	3.033268	10.64977	3.48E-04
<i>EcR</i>	Morph	NY-FW	0	-7.35539	-11.1636	-3.54714	1.49E-04
<i>EcR</i>	Morph	NY-MA	0	-14.1969	-18.0052	-10.3887	8.84E-09
<i>Ultraspiracle</i>	Morph	FW-FA	0	7.842499	3.977552	11.70745	8.09E-05
<i>Ultraspiracle</i>	Morph	MA-FA	0	15.72888	11.86393	19.59383	1.95E-09
<i>Ultraspiracle</i>	Morph	NY-FA	0	-6.55266	-10.4176	-2.68771	6.61E-04
<i>Ultraspiracle</i>	Morph	MA-FW	0	7.886381	4.021434	11.75133	7.55E-05
<i>Ultraspiracle</i>	Morph	NY-FW	0	-14.3952	-18.2601	-10.5302	8.98E-09
<i>Ultraspiracle</i>	Morph	NY-MA	0	-22.2815	-26.1465	-18.4166	3.72E-12
<i>BC core protein-like</i>	Morph	FW-FA	0	0.498996	-1.44961	2.447603	8.89E-01
<i>BC core protein-like</i>	Morph	MA-FA	0	-3.3718	-5.32041	-1.4232	5.29E-04
<i>BC core protein-like</i>	Morph	NY-FA	0	5.907207	3.9586	7.855814	2.64E-07
<i>BC core protein-like</i>	Morph	MA-FW	0	-3.8708	-5.81941	-1.92219	1.05E-04
<i>BC core protein-like</i>	Morph	NY-FW	0	5.408211	3.459604	7.356818	1.04E-06
<i>BC core protein-like</i>	Morph	NY-MA	0	9.279011	7.330404	11.22762	1.20E-10
<i>E75</i>	Morph	FW-FA	0	15.64157	3.208116	28.07503	1.06E-02
<i>E75</i>	Morph	MA-FA	0	69.55892	57.12546	81.99238	6.49E-12
<i>E75</i>	Morph	NY-FA	0	11.98368	-0.44978	24.41713	6.14E-02
<i>E75</i>	Morph	MA-FW	0	53.91735	41.48389	66.3508	6.39E-10
<i>E75</i>	Morph	NY-FW	0	-3.6579	-16.0914	8.77556	8.43E-01

Appendix

<i>E75</i>	Morph	NY-MA	0	-57.5752	-70.0087	-45.1418	1.98E-10
<i>E78C</i>	Morph	FW-FA	0	0.241605	-2.08575	2.568962	9.91E-01
<i>E78C</i>	Morph	MA-FA	0	-4.20742	-6.53478	-1.88006	3.24E-04
<i>E78C</i>	Morph	NY-FA	0	0.925711	-1.40165	3.253069	6.86E-01
<i>E78C</i>	Morph	MA-FW	0	-4.44902	-6.77638	-2.12167	1.68E-04
<i>E78C</i>	Morph	NY-FW	0	0.684107	-1.64325	3.011465	8.43E-01
<i>E78C</i>	Morph	NY-MA	0	5.13313	2.805772	7.460488	2.75E-05
<i>HR3</i>	Morph	FW-FA	0	-0.75555	-5.32715	3.816044	9.66E-01
<i>HR3</i>	Morph	MA-FA	0	-5.16392	-9.73552	-0.59233	2.33E-02
<i>HR3</i>	Morph	NY-FA	0	13.8476	9.276007	18.4192	2.67E-07
<i>HR3</i>	Morph	MA-FW	0	-4.40837	-8.97997	0.163224	6.12E-02
<i>HR3</i>	Morph	NY-FW	0	14.60315	10.03156	19.17475	1.14E-07
<i>HR3</i>	Morph	NY-MA	0	19.01153	14.43993	23.58312	1.34E-09

Raw data for fluorescence intensity across early-stage embryos

Table S6: Raw fluorescence intensity data from Image J analysis of early stage embryos from MpRV-, P2C- and untagged mCherry, and buffer A4 injected aphids.

Aphid_ID	Image_ID	Injected	Area	Mean	Min	Max	Date
mCherry1	mCherry1a	mCherry	4021.962	4392.688	0	15360	10923
mCherry1	mCherry1b	mCherry	1555.154	5037.981	0	17699	10923
P2C-mCherry1	P2C-mCherry1	P2C-mCherry	292.625	10445.931	194	55225	10923
MpRV-mCherry1	MpRV-mCherry1	MpRV-mCherry	1124.143	7043.276	41	28759	10923
Buffer1	Buffer1a	Buffer	4955.158	352.409	20	2519	40823
Buffer1	Buffer1b	Buffer	4145.182	405.355	8	2991	40823
MpRV-mCherry2	MpRV-mCherry2a	MpRV-mCherry	1059.345	34490.047	16754	65535	40823
MpRV-mCherry2	MpRV-mCherry2b	MpRV-mCherry	892.008	33887.588	16238	65535	40823
Buffer2	Buffer2a	Buffer	830.455	204.959	0	2510	221123
Buffer2	Buffer2b	Buffer	950.208	179.427	0	4707	221123
MpRV-mCherry3	MpRV-mCherry3a	MpRV-mCherry	2161.625	30768.965	769	65535	221123
MpRV-mCherry3	MpRV-mCherry3b	MpRV-mCherry	2224.884	36301.205	1735	65535	221123
MpRV-mCherry3	MpRV-mCherry3c	MpRV-mCherry	2023.245	30129.975	1515	65535	221123
MpRV-mCherry3	MpRV-mCherry3d	MpRV-mCherry	2882.079	24907.893	785	65535	221123
P2C-mCherry2	P2C-mCherry2a	P2C-mCherry	1262.024	18098.83	0	65535	221123
P2C-mCherry2	P2C-mCherry2b	P2C-mCherry	938.347	20160.388	63	65535	221123
P2C-mCherry3	P2C-mCherry3	P2C-mCherry	726.868	22620.596	713	65535	221123
Buffer3	Buffer3	Buffer	1795.045	2113.836	72	13193	190724
Buffer4	Buffer4a	Buffer	1040.561	6623.406	116	42847	190724
Buffer4	Buffer4b	Buffer	1400.914	4375.48	68	23430	190724
Buffer5	Buffer5	Buffer	3408.448	8440.806	72	32926	190724
mCherry2	mCherry2	mCherry	2379.779	6828.536	85	27529	190724
mCherry3	mCherry3	mCherry	754.139	5393.514	147	34440	190724
mCherry4	mCherry4a	mCherry	2868.349	18217.66	1073	65535	190724
mCherry4	mCherry4b	mCherry	1681.131	15092.757	746	65535	190724
MpRV-mCherry4	MpRV-mCherry4a	MpRV-mCherry	1872.768	4351.694	70	31494	190724
MpRV-mCherry4	MpRV-mCherry4b	MpRV-mCherry	2630.01	10408.192	148	36782	190724
MpRV-mCherry5	MpRV-mCherry5a	MpRV-mCherry	416.707	14450.593	87	55215	190724

MpRV-mCherry5	MpRV-mCherry5b	MpRV-mCherry	1759.371	6386.62	96	64794	190724
P2C-mCherry4	P2C-mCherry4a	P2C-mCherry	3760.152	4442.929	0	31691	190724
P2C-mCherry4	P2C-mCherry4b	P2C-mCherry	1499.18	4272.935	7	31921	190724
P2C-mCherry5	P2C-mCherry5a	P2C-mCherry	1256.652	18384.292	189	65535	190724
P2C-mCherry5	P2C-mCherry5b	P2C-mCherry	1414.398	28979.978	184	65535	190724
P2C-mCherry6	P2C-mCherry6a	P2C-mCherry	835.243	5627.873	18	65025	190724
P2C-mCherry6	P2C-mCherry6b	P2C-mCherry	760.264	5296.975	43	31796	190724
P2C-mCherry6	P2C-mCherry6c	P2C-mCherry	1300.173	6557.25	54	38628	190724
P2C-mCherry6	P2C-mCherry6d	P2C-mCherry	3072.222	6012.753	40	60194	190724
P2C-mCherry7	P2C-mCherry7a	P2C-mCherry	2062.165	5981.821	58	25716	190724
P2C-mCherry7	P2C-mCherry7b	P2C-mCherry	537.859	8206.774	105	27188	190724
P2C-mCherry7	P2C-mCherry7c	P2C-mCherry	1031.255	6448.548	74	27390	190724

Script to generate fluorescence intensity Box plot in R

```

library(readxl)
library(ggplot2)
library(ggpubr)
library(dplyr)
library(stringr)
#library(rstatix) # For using `pairwise_t_test` and significance letters

##read in intensity file
Intensity_final <- read_excel("/Users/jamesr/Library/CloudStorage/OneDrive-
NorwichBioScienceInstitutes/PhD/Microscopy/mCherry injections/Myzus persicae/880/High
Conc/Intensity_final.xlsx")

##set date as factor
Intensity_final$Date <- as.factor(Intensity_final$Date)

##classify order
Intensity_final <- within(Intensity_final, Injected <- factor(Injected, levels = c('P2C-mCherry',
'MpRV-mCherry', 'mCherry', 'Buffer')))

##perform mixed anova to account for multiple datapoints per aphid and date
#install.packages("lme4", type = "source")
library(lme4)
mixed_anova <- lmer(Mean ~ Date + Injected + (1 | Aphid_ID), data = Intensity_final)
summary(mixed_anova)
pvals_named

##create plot
bxp <- ggboxplot(Intensity_final, x = 'Injected', y = 'Mean') +
  theme(text = element_text(size=20)) +
  geom_jitter(aes(color = Injected), width = 0.25, height = 0.1, cex=3) +
  scale_y_continuous(name = "Mean Intensity") +
  xlab("Injection") +
  scale_color_manual(values = c("#E69F00", "#56B4E9", "#FF0000", "Yellow")) +
  theme(text = element_text(size=15))
bxp

##Save plot as SVG
#install.packages("svglite")
#library(svglite)

#ggsave("Intensity_mixed_anova_bxp.svg", plot = bxp)

```


MpRV-Cas9 ReMOT control experiment tables

14/02/2024

Table S7: Injected adult counts for the ReMOT control injections with MpRV-Cas9 performed on 14/02/2024.

Group	Day 0	Day 2	Day 5	Day 8	Day 14
A4 + CB	10	6	1	1	1
1	10	6	4	4	4
2	10	6	6	5	3
3	10	10	4	3	2
4	10	6	4	1	0
5	10	3	3	3	3
6	10	6	3	2	4

Table S8: Progeny counts and eye colour inspection for the ReMOT control injections with MpRV-Cas9 performed on 14/02/2024. 'Eyes' refers to number of aphids with an eye colour change.

	Day 2		Day 5		Day 8		Day 14	
Group	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes
A4 + CB	41	0	23	0	5	0	0	0
1	38	0	48	0	43	0	26	0
2	72	0	87	0	51	0	40	0
3	53	0	70	0	34	0	10	0
4	34	0	56	0	12	0	0	0
5	27	0	25	0	16	0	4	0
6	33	0	50	0	9	0	5	0
Total (1-6)	257	0	336	0	165	0	85	0

21/02/24

Table S9: Injected adult and progeny counts, and eye colour inspection for ReMOT control injections with MpRV-Cas9 on 21/02/2024. Only data for day 10 was recorded in this case.

	Day 0		Day 10	
Group	Adults	Progeny	Adults	Progeny
A4 + CB	5	0	3	5
1	10	0	5	30
2	10	0	5	29
3	10	0	6	42
4	10	0	0	0
5	10	0	7	52
6	10	0	0	0
Total (1-6)	60	0	23	153

12/04/24

Table S10: Injected adult counts for the ReMOT control injections with MpRV-Cas9 performed on 12/04/2024.

Group	Day 0	Day 2	Day 4	Day 6	Day 9	Day 12	Day 14	Day 17
A4 + CB +Sap	5	4	0	0	0	0	0	0
1	10	10	9	8	6	2	2	2
2	10	10	8	7	5	3	3	3
3	10	9	6	4	3	1	1	1
4	10	8	5	5	3	3	2	2
5	10	6	5	4	3	3	2	1
6	10	3	3	3	3	3	3	3

Table S11: Progeny counts and eye colour inspection for the ReMOT control injections with MpRV-Cas9 performed on 12/04/2024. 'Eyes' refers to number of aphids with an eye colour change.

	Day 2		Day 4		Day 6		Day 9		Day 12		Day 14		Day 17	
Group	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes
A4 + CB + Sap	27	0	18	0	-	-	-	-	-	-	-	-		
1	48	0	78	0	64	0	59	0	21	0	8	0	11	0
2	43	0	66	0	35	0	65	0	36	0	22	0	23	0
3	38	0	43	0	42	0	38	0	4	0	0	0	14	0
4	31	0	26	0	31	0	40	0	31	0	20	0	22	0
5	25	0	26	0	38	0	37	0	30	0	19	0	10	0
6	15	0	27	0	28	0	23	0	33	0	15	0	8	0
Toal (1-6)	200	0	266	0	238	0	262	0	155	0	84	0	88	0

01/05/24

Table S12: Injected adult counts for the ReMOT control injections with MpRV-Cas9 performed on 01/05/2024. In this experiment, lack of survival was likely due to improper watering of the plant. **Any escapees were collected here.

Group	Day 0	Day 4	Day 8	Day 10	Day 12	Day 14**
A4 + CB + Sap	10	0	0	0	0	0
1	15	3	0	0	0	0
2	15	6	2	2	2	0
3	15	0	0	0	0	0
4	15	1	1	1	0	0

Table S13: Progeny counts and eye colour inspection for the ReMOT control injections with MpRV-Cas9 performed on 01/05/2024. 'Eyes' refers to number of aphids with an eye colour change. **Any escapees were collected here.

	Day 4*		Day 8		Day 10		Day 12**		Day 14**	
Group	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes
A4 + CB + Sap	0	0	0	0	0	0	0	0	0	0
1	0	0	10	0	0	0	0	0	0	0
2	0	0	12	0	9	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0
Total (1-4)	0	0	23	0	9	0	0	0	0	0

13/06/24

Table S14: Injected adult counts for the ReMOT control injections with MpRV-Cas9 performed on 13/06/2024.

Group	Day 0	Day 2	Day 5	Day 7	Day 11	Day 13	Day 15	Day 18
A4 + CB +Sap	6	1	1	1	1	0	0	0
1	15	11	3	3	2	2	2	1
2	15	8	7	5	2	2	2	2
3	5	1	1	1	1	1	2	0

Table S15: Progeny counts and eye colour inspection for the ReMOT control injections with MpRV-Cas9 performed on 13/06/2024. 'Eyes' refers to number of aphids with an eye colour change.

	Day 0		Day 5		Day 7		Day 11		Day 13		Day 15		Day 18	
Group	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes
A4 + CB +Sap	0	0	0	0	0	0	6	0	1	0	0	0	0	0
1	0	0	13	0	38	0	19	0	8	0	0	0	2	0
2	0	0	13	0	38	0	20	0	14	0	10	0	16	0
3	0	0	0	0	0	0	0	0	12	0	0	0	0	0
Total (1-6)	0	0	26	0	76	0	39	0	34	0	10	0	18	0

P2C-Cas9 ReMOT control experiment tables

15/12/23

Table S16: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 15/12/2023.

Group	Day 0	Day 3	Day 4	Day 6
A4 + CB	5	4	4	1
1	10	10	7	5
2	10	4	3	3
3	10	2	1	1
4	10	1	1	1
5	10	5	5	4
6	10	0	0	0

Table S17: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 15/12/2023. 'Eyes' refers to number of aphids with an eye colour change.

	Day 3		Day 4		Day 6	
Group	Progeny	eyes	Progeny	eyes	Progeny	eyes
A4 + CB	10	0	5	0	4	0
1	48	0	8	0	20	0
2	16	0	5	0	13	0
3	8	0	4	0	9	0
4	4	0	3	0	3	0
5	34	0	6	0	22	0
6	0	0	0	0	0	0
Total (1-6)	110	0	26	0	67	0

15/02/2024

Table S18: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 15/02/2024.

Group	Day 0	Day 10	Day 13
A4+CB	10	0	0
1	10	2	2
2	10	1	0
3	10	0	0
4	10	0	0
5	10	0	0
6	10	0	0

Table S19: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 15/02/2024. 'Eyes' refers to number of aphids with an eye colour change.

	Day 10		Day 13	
Group	Progeny	eyes	Progeny	eyes
A4+CB	7	0	0	0
1	5	0	26	0
2	8	0	0	0
3	3	0	0	0
4	1	0	0	0
5	1	0	0	0
6	0	0	0	0
Total (1-6)	18	0	26	0

23/02/2024

Table S20: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 23/02/2024.

Group	Day 0	Day 2	Day 5	Day 7
A4 + CB	10	7	6	2
1	10	5	4	5
2	10	8	7	5
3	10	3	3	3
4	10	1	0	1
5	10	1	0	0
6	10	0	0	0

Table S21: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 23/02/2024. 'Eyes' refers to number of aphids with an eye colour change.

	Day 2		Day 5		Day 7	
Group	Progeny	Eyes	Progeny	Eyes	Progeny	Eyes
A4 + CB	13	0	65	0	2	0
1	11	0	41	0	5	0
2	21	0	64	0	5	0
3	2	0	18	0	3	0
4	0	0	0	0	1	0
5	0	0	0	0	0	0
6	0	0	0	0	0	0
Total (1-6)	34	0	123	0	14	0

08/03/24

Table S22: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 08/03/2024.

Group	Day 0	Day 2	Day 5	Day 12	Day 14
A4 + CB	5	4	4	2	2
1	10	9	9	3	3
2	10	7	7	4	2
3	10	4	4	2	1
4	10	4	4	6	5
5	10	4	3	2	3
6	10	2	2	2	2

Table S23: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 08/03/2024. 'Eyes' refers to number of aphids with an eye colour change.

	Day 2		Day 5		Day 12		Day 14	
Group	Progeny	Eyes	Progeny	Eyes	Progeny	Eyes	Progeny	Eyes
A4 + CB	31	0	21	0	15	0	2	3
1	27	1	45	1	7	0	3	9
2	25	0	5	0	4	0	2	3
3	9	0	15	0	9	0	1	5
4	10	0	47	0	25	0	5	31
5	0	0	10	0	25	0	3	7
6	3	0	10	0	15	0	2	13
Total (1-6)	74	1	132	1	85	0	16	68

15/03/24

Table S24: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 15/03/2024. On Day 3 adults were removed from the original clip cage to a separate clip cage demarked by the 'a' suffix.

Group	Day 0	Day 3	Day 5	Day 8
A4 + CB + Sap	5	1	1	1
1	10	4	0	0
1a	0	0	2	2
2	10	5	0	0
2a	0	0	4	3
3	10	4	0	0
3a	0	0	1	3
4	10	5	0	0
4a	0	0	5	3
5	10	0	0	0

Table S25: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 15/03/2024. 'Eyes' refers to number of aphids with an eye colour change. On Day 3 adults were removed from the original clip cage to a separate clip cage demarked by the 'a' suffix. Remaining progeny from both clip cages were counted and screened.

	Day 3		Day 5		Day 8	
Group	Progeny	Eyes	Progeny	Eyes	Progeny	Eyes
A4 + CB + Sap	38	0	8	0	9	0
1	19	0	24	0	19	0
1a	0	0	19	0	11	0
2	13	0	17	0	25	0
2a	0	0	20	0	42	0
3	16	0	14	0	0	0
3a	0	0	0	0	27	0
4	20	0	22	0	20	0
4a	0	0	25	0	31	0
5	0	0	0	0	0	0
Total (1-5)	68	0	141	0	175	0

03/05/24

Table S26: Injected adult counts for the ReMOT control injections with P2C-Cas9 performed on 03/05/2023.

Group	Day 0	Day 2	Day 5	Day 9	Day 13	Day 15
A4 + CB +Sap	10	8	3	3	3	0
1	15	10	8	2	0	0
2	15	6	2	2	2	2

Table S27: Progeny counts and eye colour inspection for the ReMOT control injections with P2C-Cas9 performed on 03/05/2024. 'Eyes' refers to number of aphids with an eye colour change.

	Day 0		Day 2		Day 5		Day 9		Day 13		Day 15	
Group	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes	Progeny	eyes
A4 + CB + Sap	0	0	50	0	6	0	15	0	40	0	6	0
1	0	0	35	0	65	0	48	0	5	0	18	0
2	0	0	13	0	28	0	17	0	47	0	22	0
Total (1-2)	0	0	98	0	99	0	80	0	92	0	46	0

Summary table of ICE analysis for MpRV-Cas9 experiments

Table S28: Summary table of all ICE analysis on amplicons from pools of progeny from MpRV-Cas9 injected aphids.

Saponin	Target	Date injected	Date collected	ICE	KO-Score	ICE d	R Squared	Mean Discord Before	Mean Discord After	Guide Sequences	Control Sample Quality Score	Edit Sample Quality Score	Notes	Indels
No	1	08/02/24	10/02/24	0	0	0	0.98	0.185651 193	0.053054 584	GGTCGTTTTGC CGGCGCCAC	59	54		{'0': 98.0, '-1': 0.0}
No	1	08/02/24	10/02/24	0	0	2	0.99	0.105899 698	0.071580 018	GGTCGTTTTGC CGGCGCCAC	56	62		{'0': 99.0, '-1': 0.0}
No	1	08/02/24	12/02/24	0	0	1	0.98	0.054993 637	0.079593 997	GGTCGTTTTGC CGGCGCCAC	59	61		{'0': 98.0, '-1': 0.0}
No	1	08/02/24	12/02/24	0	0	1	0.99	0.119974 251	0.044131 824	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}
No	1	08/02/24	14/02/24	0	0	0	0.98	0.042449 169	0.049721 108	GGTCGTTTTGC CGGCGCCAC	59	61		{'0': 98.0, '-1': 0.0}
No	1	08/02/24	14/02/24	0	0	1	0.99	0.035266 954	0.063910 74	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}
No	1	08/02/24	16/02/24	0	0	3	0.98	0.117432 392	0.165565 472	GGTCGTTTTGC CGGCGCCAC	59	59		{'0': 98.0, '-1': 0.0}
No	1	08/02/24	16/02/24	0	0	1	0.99	0.033087 589	0.061309 254	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}
No	1	08/02/24	22/02/24	0	0	1	0.99	0.060521 111	0.059533 565	GGTCGTTTTGC CGGCGCCAC	59	62		{'0': 99.0, '-1': 0.0}
No	1	08/02/24	22/02/24	0	0	1	0.99	0.104098 221	0.070621 605	GGTCGTTTTGC CGGCGCCAC	56	59		{'0': 99.0, '-1': 0.0}
No	1	14/02/24	16/02/24	0	0	0	0.98	0.061055 147	0.073066 952	GGTCGTTTTGC CGGCGCCAC	59	61		{'0': 98.0, '-1': 0.0}
No	1	14/02/24	16/02/24	0	0	1	0.99	0.098004 797	0.061810 182	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}
No	1	14/02/24	19/02/24	0	0	0	0.98	0.033986 613	0.081714 247	GGTCGTTTTGC CGGCGCCAC	59	61		{'0': 98.0, '-1': 0.0}
No	1	14/02/24	19/02/24	0	0	1	0.99	0.190449 648	0.054762 342	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}
No	1	14/02/24	22/02/24	0	0	2	0.98	0.179333 694	0.082501 5	GGTCGTTTTGC CGGCGCCAC	59	61		{'0': 98.0, '-1': 0.0}
No	1	14/02/24	22/02/24	0	0	1	0.99	0.069126 969	0.069559 785	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}
No	1	14/02/24	01/03/24	0	0	1	0.98	0.092602 813	0.054092 975	GGTCGTTTTGC CGGCGCCAC	59	61		{'0': 98.0, '-1': 0.0}
No	1	14/02/24	01/03/24	0	0	1	0.99	0.111711 665	0.063531 602	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}

Appendix

No	1	21/02/24	25/02/24	0	0	2	0.98	0.185497 993	0.091459 049	GGTCGTTTTGC CGGCGCCAC	59	59		{'0': 98.0, '-1': 0.0}
No	1	21/02/24	25/02/24							GGTCGTTTTGC CGGCGCCAC			ERROR - Sample ab1 quality scores too low	
No	1	21/02/24	27/02/24	0	0	0	0.98	0.088521 102	0.051067 645	GGTCGTTTTGC CGGCGCCAC	59	60		{'0': 98.0, '-1': 0.0}
No	1	21/02/24	27/02/24	0	0	1	0.99	0.056410 272	0.057075 3	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}
No	1	21/02/24	01/03/24	0	0	2	0.98	0.353653 091	0.101987 724	GGTCGTTTTGC CGGCGCCAC	59	59		{'0': 98.0, '-1': 0.0}
No	1	21/02/24	01/03/24	0	0	1	0.99	0.068086 798	0.061236 739	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}
No	1	21/02/24	01/03/24	0	0	1	0.98	0.087443 922	0.074984 976	GGTCGTTTTGC CGGCGCCAC	59	61		{'0': 98.0, '-1': 0.0}
No	1	21/02/24	01/03/24	0	0	1	0.99	0.081119 65	0.060405 75	GGTCGTTTTGC CGGCGCCAC	56	61		{'0': 99.0, '-1': 0.0}
No	1	21/02/24	01/03/24	9	9	21	0.9	0.261908 342	0.238637 664	GGTCGTTTTGC CGGCGCCAC	59	28		{'0': 81.0, '1': 9.0}
No	1	21/02/24	01/03/24	0	0	9	0.97	0.192954 785	0.130332 551	GGTCGTTTTGC CGGCGCCAC	56	43		{'0': 97.0, '-1': 0.0}
No	2	08/02/24	10/02/24							GGCAGAATTAA AGGTTTATC			ERROR - Sample ab1 quality scores too low	
No	2	08/02/24	10/02/24							GGCAGAATTAA AGGTTTATC			ERROR - Sample ab1 quality scores too low	
No	2	08/02/24	12/02/24	0	0	0	1	0.295206 99	0.036690 537	GGCAGAATTAA AGGTTTATC	62	59		{'0': 100.0, '-1': 0.0}
No	2	08/02/24	12/02/24	0	0	6	0.99	0.278334 764	0.173913 712	GGCAGAATTAA AGGTTTATC	50	43		{'0': 99.0, '-1': 0.0}
No	2	08/02/24	14/02/24	0	0	1	1	0.305994 759	0.042477 241	GGCAGAATTAA AGGTTTATC	62	61		{'0': 100.0, '-1': 0.0}
No	2	08/02/24	14/02/24	0	0	1	0.99	0.354082 022	0.114067 867	GGCAGAATTAA AGGTTTATC	50	59		{'0': 99.0, '-1': 0.0}
No	2	08/02/24	16/02/24	0	0	1	1	0.413074 873	0.054246 957	GGCAGAATTAA AGGTTTATC	62	59		{'0': 100.0, '-1': 0.0}
No	2	08/02/24	16/02/24							GGCAGAATTAA AGGTTTATC			ERROR - Sample ab1 quality scores too low	
No	2	08/02/24	22/02/24	0	0	0	1	0.297434 332	0.048208 838	GGCAGAATTAA AGGTTTATC	62	62		{'0': 100.0, '-1': 0.0}
No	2	08/02/24	22/02/24	0	0	2	0.98	0.263647 087	0.106999 832	GGCAGAATTAA AGGTTTATC	50	62		{'0': 98.0, '-1': 0.0}
No	2	14/02/24	16/02/24	0	0	7	0.98	0.433224 737	0.091212 979	GGCAGAATTAA AGGTTTATC	62	44		{'0': 98.0, '-1': 0.0}
No	2	14/02/24	16/02/24							GGCAGAATTAA AGGTTTATC			ERROR - No quality alignment found between edited and wildtype upstream of cut site	

Appendix

No	2	14/02/24	19/02/24	0	0	6	0.96	0.452970 375	0.076772 392	GGCAGAATTAA AGGTTTATC	62	24		{'0': 96.0, '-1': 0.0}
No	2	14/02/24	19/02/24							GGCAGAATTAA AGGTTTATC			ERROR - No quality alignment found between edited and wildtype upstream of cut site	
No	2	14/02/24	22/02/24							GGCAGAATTAA AGGTTTATC			ERROR - Sample ab1 quality scores too low	
No	2	14/02/24	22/02/24							GGCAGAATTAA AGGTTTATC			ERROR - No quality alignment found between edited and wildtype upstream of cut site	
No	2	14/02/24	01/03/24	0	0	2	1	0.245348 091	0.061361 359	GGCAGAATTAA AGGTTTATC	62	61		{'0': 100.0, '-1': 0.0}
No	2	14/02/24	01/03/24	0	0	7	0.98	0.207207 973	0.367374 391	GGCAGAATTAA AGGTTTATC	50	59		{'0': 98.0, '-1': 0.0}
No	2	21/02/24	25/02/24	0	0	15	0.97	0.203111 529	0.321872 814	GGCAGAATTAA AGGTTTATC	62	59		{'0': 97.0, '-1': 0.0}
No	2	21/02/24	25/02/24							GGCAGAATTAA AGGTTTATC			ERROR - Sample ab1 quality scores too low	
No	2	21/02/24	27/02/24	0	0	3	1	0.304716 162	0.064720 975	GGCAGAATTAA AGGTTTATC	62	61		{'0': 100.0, '-1': 0.0}
No	2	21/02/24	27/02/24	0	0	4	0.98	0.195313 135	0.456378 434	GGCAGAATTAA AGGTTTATC	50	52		{'0': 98.0, '-1': 0.0}
No	2	21/02/24	01/03/24	0	0	9	0.98	0.273997 545	0.140776 065	GGCAGAATTAA AGGTTTATC	62	59		{'0': 98.0, '-1': 0.0}
No	2	21/02/24	01/03/24	0	0	12	0.95	0.368657 29	0.191016 63	GGCAGAATTAA AGGTTTATC	50	49		{'0': 95.0, '-1': 0.0}
No	2	21/02/24	01/03/24	0	0	13	0.98	0.435378 278	0.156533 163	GGCAGAATTAA AGGTTTATC	62	59		{'0': 98.0, '-1': 0.0}
No	2	21/02/24	01/03/24							GGCAGAATTAA AGGTTTATC			ERROR - Sample ab1 quality scores too low	
No	2	21/02/24	01/03/24	0	0	6	0.99	0.327588 452	0.106823 769	GGCAGAATTAA AGGTTTATC	62	61		{'0': 99.0, '-1': 0.0}
No	2	21/02/24	01/03/24	0	0	12	0.96	0.254692 909	0.195101 744	GGCAGAATTAA AGGTTTATC	50	39		{'0': 96.0, '-1': 0.0}
No	3	08/02/24	10/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	08/02/24	10/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	08/02/24	12/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	

Appendix

No	3	08/02/24	12/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	08/02/24	14/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	08/02/24	14/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	08/02/24	16/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	08/02/24	16/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	08/02/24	22/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	08/02/24	22/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	14/02/24	16/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	14/02/24	16/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	14/02/24	19/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	14/02/24	19/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	14/02/24	22/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	

Appendix

No	3	14/02/24	22/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	14/02/24	01/03/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	14/02/24	01/03/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	21/02/24	25/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	21/02/24	25/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	21/02/24	27/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	21/02/24	27/02/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	21/02/24	01/03/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	21/02/24	01/03/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	21/02/24	01/03/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	21/02/24	01/03/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
No	3	21/02/24	01/03/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	

Appendix

No	3	21/02/24	01/03/24							GACGGTGTGAT GAACATCAA			ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file	
Yes	1	12/04/24	14/04/24	0	0	0	0.99	0.341587 26	0.085225 49	GGTCGTTTTGC CGGCGCCAC	59	32		{'0': 99.0, '-1': 0.0}
Yes	1	12/04/24	16/04/24	0	0	0	0.99	0.044082 7	0.085354 47	GGTCGTTTTGC CGGCGCCAC	59	62		{'0': 99.0, '-1': 0.0}
Yes	1	12/04/24	18/04/24	0	0	1	0.99	0.126074 87	0.086259 24	GGTCGTTTTGC CGGCGCCAC	59	59		{'0': 99.0, '-1': 0.0}
Yes	1	12/04/24	21/04/24	0	0	0	0.99	0.102282 81	0.080744 03	GGTCGTTTTGC CGGCGCCAC	59	59		{'0': 99.0, '-1': 0.0}
Yes	1	12/04/24	26/04/24	0	0	1	0.99	0.092307 8	0.084739 06	GGTCGTTTTGC CGGCGCCAC	59	59		{'0': 99.0, '-1': 0.0}
Yes	1	12/04/24	27/04/24	0	0	1	0.99	0.041980 25	0.072622 18	GGTCGTTTTGC CGGCGCCAC	59	62		{'0': 99.0, '-1': 0.0}
Yes	1	12/04/24	01/05/24	0	0	1	0.99	0.185213 19	0.087984 93	GGTCGTTTTGC CGGCGCCAC	59	62		{'0': 99.0, '-1': 0.0}
Yes	2	12/04/24	14/04/24	0	0	2	1	0.301509 46	0.091249 68	GGCAGAATTAA AGGTTTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	12/04/24	16/04/24	0	0	1	1	0.288439 03	0.077245 8	GGCAGAATTAA AGGTTTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	12/04/24	18/04/24	0	0	2	1	0.267378 44	0.089092 04	GGCAGAATTAA AGGTTTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	12/04/24	21/04/24	0	0	3	1	0.279952 58	0.097472 9	GGCAGAATTAA AGGTTTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	12/04/24	26/04/24	0	0	3	1	0.262369 04	0.073545 47	GGCAGAATTAA AGGTTTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	12/04/24	27/04/24	0	0	2	1	0.269008 52	0.063155 67	GGCAGAATTAA AGGTTTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	12/04/24	01/05/24	0	0	2	1	0.297403 3	0.089443 36	GGCAGAATTAA AGGTTTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	3	12/04/24	14/04/24										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file,'guide_alignments'	
Yes	3	12/04/24	16/04/24										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file,'guide_alignments'	
Yes	3	12/04/24	18/04/24										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file,'guide_alignments'	

Appendix

Yes	3	12/04/24	21/04/24										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file,'guide_alignments'	
Yes	3	12/04/24	26/04/24										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file,'guide_alignments'	
Yes	3	12/04/24	27/04/24										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file,'guide_alignments'	
Yes	3	12/04/24	01/05/24										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCA A in control file,'guide_alignments'	

Summary table of ICE analysis for P2C-Cas9 experiments

Table S29: Summary table of all ICE analysis on amplicons from pools of progeny from P2C-Cas9 injected aphids.

Saponin	Target	Date injected	Date collected	ICE	KO-Score	ICE d	R Squared	Mean Discord Before	Mean Discord After	Guide Sequences	Control Sample Quality Score	Edit Sample Quality Score	Notes	Indels
No	1	15/02/2024	25/02/2024	0	0	1	0.99	0.29400773	0.078249196	GGTCGTTT TGCCGGC GCCAC	56	59		
No	1	15/02/2024	25/02/2024	0	0	1	0.99	0.08825944	0.100042852	GGTCGTTT TGCCGGC GCCAC	59	59		
No	1	15/02/2024	17/02/2024	0	0	0	0.98	0.07470539	0.054463729	GGTCGTTT TGCCGGC GCCAC	59	62		
No	1	15/02/2024	17/02/2024	0	0	1	0.99	0.14430145	0.06408296	GGTCGTTT TGCCGGC GCCAC	56	62		{'0': 98.0, '-1': 0.0}
No	1	15/02/2024	28/02/2024	0	0	0	0.98	0.16153697	0.070680266	GGTCGTTT TGCCGGC GCCAC	59	62		{'0': 100.0, '-1': 0.0}
No	1	15/02/2024	28/02/2024	0	0	0	0.99	0.25889379	0.051132268	GGTCGTTT TGCCGGC GCCAC	56	62		
No	1	23/02/2024	25/02/2024	0	0	1	0.99	0.26553007	0.056210614	GGTCGTTT TGCCGGC GCCAC	56	59		{'0': 99.0, '-1': 0.0}
No	1	23/02/2024	25/02/2024	0	0	1	0.98	0.06085602	0.036412068	GGTCGTTT TGCCGGC GCCAC	59	62		{'0': 9.0, '1': 21.0, '2': 3.0, '3': 5.0, '-8': 2.0, '-9': 2.0, '10': 3.0, '15': 1.0, '-11': 3.0, '-13': 5.0, '-16': 1.0, '-26': 2.0}
No	1	23/02/2024	28/02/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low	{'0': 98.0, '-1': 0.0}
No	1	23/02/2024	28/02/2024	0	0	1	0.98	0.23795055	0.084814477	GGTCGTTT TGCCGGC GCCAC	59	59		{'0': 97.0, '1': 1.0}
No	1	08/03/2024	10/03/2024	0	0	2	0.98	0.25520412	0.141130382	GGTCGTTT TGCCGGC GCCAC	59	59		{'0': 99.0, '-1': 0.0}

Appendix

No	1	08/03/2024	10/03/2024	0	0	1	0.99	0.2824 551	0.10604 348	GGTCGTTT TGCCGGC GCCAC	56	59		
No	1	15/02/2024	20/02/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low, WARNING - No high-quality regions found, exiting early	
No	1	15/02/2024	20/02/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low, WARNING - No high-quality regions found, exiting early	
No	1	08/03/2024	18/03/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low, WARNING - No high-quality regions found, exiting early	
No	1	08/03/2024	18/03/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low, WARNING - No high-quality regions found, exiting early	{'1': 23.0, '-10': 6.0, '-13': 7.0, '-26': 2.0, '-28': 1.0, '-30': 5.0, '-32': 4.0, '-35': 1.0, '-36': 2.0, '-39': 2.0}
No	1	08/03/2024	20/03/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Could not find cut site 171 in the edited object	
No	1	08/03/2024	20/03/2024							GGTCGTTT TGCCGGC GCCAC	59	59	ERROR - <class 'ValueError'> Found array with 0 sample(s) (shape= (0, 461)) while a minimum of 1 is required. A: (0, 461) B: (0,), WARNING - Inf. window after cut site, -30, is less than 3x indel_max_size of 20	{'0': 99.0, '-1': 0.0}
No	1	08/03/2024	22/03/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low	
No	1	08/03/2024	22/03/2024	53	53	91	0.53	0.1913 8261	0.75762 9469	GGTCGTTT TGCCGGC GCCAC	56	39	WARNING - Inf. window after cut site, 35, is less than 3x indel_max_size of 20	
No	1	08/03/2024	23/03/2024	22	12	66	0.45	0.0933 2486	0.62645 799	GGTCGTTT TGCCGGC GCCAC	56	59		
No	1	08/03/2024	23/03/2024	48	40	83	0.57	0.1058 5008	0.71979 2587	GGTCGTTT TGCCGGC GCCAC	59	59		
No	1	08/03/2024	14/03/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	
No	1	08/03/2024	14/03/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	
No	1	08/03/2024	16/03/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	

Appendix

No	1	08/03/2024	16/03/2024							GGTCGTTT TGCCGGC GCCAC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	{'0': 98.0, '-1': 0.0}
No	2	15/02/2024	25/02/2024	0	0	6	0.98	0.1736 8752	0.10027 8698	GGCAGAA TTAAAGGT TTATC	62	62		Indels
No	2	15/02/2024	25/02/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	
No	2	15/02/2024	17/02/2024	1	1	2	0.98	0.1271 6385	0.10277 677	GGCAGAA TTAAAGGT TTATC	50	62		
No	2	15/02/2024	17/02/2024	0	0	1	0.99	0.2743 2274	0.04271 4851	GGCAGAA TTAAAGGT TTATC	62	62		{'0': 98.0, '-1': 0.0}
No	2	23/02/2024	28/02/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	
No	2	23/02/2024	28/02/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Could not find cut site 164 in the edited object	{'0': 99.0, '-1': 0.0}
No	2	08/03/2024	11/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low	{'0': 98.0, '-1': 0.0}
No	2	08/03/2024	11/03/2024	0	0	4	0.99	0.0919 2182	0.09113 3408	GGCAGAA TTAAAGGT TTATC	62	62	WARNING - Inf. window after cut site, 54, is less than 3x indel_max_size of 20	{'0': 23.0, '6': 1.0, '9': 3.0, '-1': 8.0, '-9': 4.0, '17': 1.0, '-12': 2.0, '-22': 3.0}
No	2	15/02/2024	20/02/2024	0	0	0	1	0.0695 2214	0.03075 1582	GGCAGAA TTAAAGGT TTATC	62	62		
No	2	15/02/2024	20/02/2024	0	0	1	0.98	0.1007 4248	0.08314 6802	GGCAGAA TTAAAGGT TTATC	50	62		{'0': 99.0, '-1': 0.0}
No	2	08/03/2024	16/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	{'0': 98.0, '-1': 0.0}
No	2	08/03/2024	16/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	
No	2	08/03/2024	20/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - No quality alignment found between edited and wildtype upstream of cut site	{'0': 98.0, '-1': 0.0}
No	2	08/03/2024	20/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - No quality alignment found between edited and wildtype upstream of cut site	

Appendix

No	2	08/03/2024	22/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low	{'0': 99.0, '-1': 0.0}
No	2	08/03/2024	22/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low	
No	2	08/03/2024	23/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low	{'0': 99.0, '-1': 0.0}
No	2	08/03/2024	23/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low	{'0': 98.0, '-1': 0.0}
No	2	08/03/2024	14/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	{'0': 99.0, '-1': 0.0}
No	2	08/03/2024	14/03/2024							GGCAGAA TTAAAGGT TTATC			ERROR - Sample ab1 quality scores too low, WARNING - No high quality regions found, exiting early	{'0': 99.0, '-1': 0.0}
No	3	15/02/2024	25/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	15/02/2024	25/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	15/02/2024	28/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	15/02/2024	28/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	23/02/2024	25/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	23/02/2024	25/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	23/02/2024	28/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	23/02/2024	28/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	15/02/2024	17/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	15/02/2024	17/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	

Appendix

No	3	08/03/2024	12/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	12/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	15/02/2024	20/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	15/02/2024	20/02/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	13/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	13/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	16/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	16/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	20/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	20/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	22/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	22/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	23/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	23/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	14/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
No	3	08/03/2024	14/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	

Appendix

Yes	1	15/03/2024	20/03/2024	0	0	1	0.99	0.0731 90465	0.07269 6957	GGTCGTTT TGCCGGC GCCAC	59	62		{'0': 99.0, '-1': 0.0}
Yes	1	15/03/2024	23/03/2024	0	0	1	0.99	0.1082 13078	0.08038 3418	GGTCGTTT TGCCGGC GCCAC	59	59		{'0': 99.0, '-1': 0.0}
Yes	1	15/03/2024	27/03/2024	0	0	1	0.98	0.0657 32334	0.08675 6583	GGTCGTTT TGCCGGC GCCAC	59	62		{'0': 98.0, '-1': 0.0}
Yes	1	15/03/2024	03/04/2024	0	0	2	0.98	0.1366 03648	0.08603 7054	GGTCGTTT TGCCGGC GCCAC	59	59		{'0': 98.0, '-1': 0.0}
Yes	1	03/05/2024	08/05/2024	0	0	1	0.99	0.0691 59686	0.07981 202	GGTCGTTT TGCCGGC GCCAC	59	62		{'0': 99.0, '-1': 0.0}
Yes	1	03/05/2024	12/05/2024	0	0	1	0.99	0.0719 34922	0.08707 57	GGTCGTTT TGCCGGC GCCAC	59	62		{'0': 99.0, '-1': 0.0}
Yes	1	03/05/2024	18/05/2025	0	0	1	0.98	0.2346 83827	0.08102 261	GGTCGTTT TGCCGGC GCCAC	59	60		{'0': 98.0, '-1': 0.0}
Yes	1	03/05/2024	22/05/2024	0	0	0	0.99	0.0553 86459	0.07641 3222	GGTCGTTT TGCCGGC GCCAC	59	62		{'0': 99.0, '-1': 0.0}
Yes	2	15/03/2024	20/03/2024	0	0	6	0.99	0.6072 09602	0.08114 6015	GGCAGAA TTAAAGGT TTATC	62	43		{'0': 99.0, '-1': 0.0}
Yes	2	15/03/2024	23/03/2024	0	0	4	0.99	0.3185 50247	0.07886 0491	GGCAGAA TTAAAGGT TTATC	62	51		{'0': 99.0, '-1': 0.0}
Yes	2	15/03/2024	27/03/2024	0	0	2	1	0.2812 13119	0.06938 8726	GGCAGAA TTAAAGGT TTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	15/03/2024	03/04/2024	0	0	2	1	0.2622 97946	0.08683 6999	GGCAGAA TTAAAGGT TTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	03/05/2024	08/05/2024	0	0	7	0.99	0.3045 40265	0.08922 1704	GGCAGAA TTAAAGGT TTATC	62	56		{'0': 99.0, '-1': 0.0}
Yes	2	03/05/2024	12/05/2024	0	0	2	1	0.3024 28504	0.09118 1629	GGCAGAA TTAAAGGT TTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	03/05/2024	18/05/2025	0	0	2	1	0.2961 7263	0.09522 9493	GGCAGAA TTAAAGGT TTATC	62	62		{'0': 100.0, '-1': 0.0}
Yes	2	03/05/2024	22/05/2024	0	0	3	1	0.3459 21933	0.07122 8299	GGCAGAA TTAAAGGT TTATC	62	62		{'0': 100.0, '-1': 0.0}

Appendix

Yes	3	15/03/2024	20/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
Yes	3	15/03/2024	23/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
Yes	3	15/03/2024	27/03/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
Yes	3	15/03/2024	03/04/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
Yes	3	03/05/2024	08/05/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
Yes	3	03/05/2024	12/05/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
Yes	3	03/05/2024	18/05/2025										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	
Yes	3	03/05/2024	22/05/2024										ERROR - Could not find cut site for guide GACGGTGTGATGAACATCAA in control file,'guide_alignments'	

References

1. Hogenhout SA, Bos JIB. Effector proteins that modulate plant-insect interactions. *Current Opinion in Plant Biology*. 2011;14(4):422-8.
2. Fereres A, Moreno A. Behavioural aspects influencing plant virus transmission by homopteran insects. *Virus Research*. 2009;141(2):158-68.
3. Misof B, Liu SL, Meusemann K, Peters RS, Donath A, Mayer C, et al. Phylogenomics resolves the timing and pattern of insect evolution. *Science*. 2014;346(6210):763-7.
4. Douglas AE. Phloem-sap feeding by animals: problems and solutions. *Journal of Experimental Botany*. 2006;57(4):747-54.
5. Panfilio KA, Angelini DR. By land, air, and sea: hemipteran diversity through the genomic lens. *Current Opinion in Insect Science*. 2018;25:106-15.
6. Gaire S, Schal C, Mick R, DeVries Z. The Role of Antennae in Heat Detection and Feeding Behavior in the Bed Bug (Hemiptera: Cimicidae). *Journal of Economic Entomology*. 2020;113(6):2858-63.
7. Rabinovich JE, Kitron UD, Obed Y, Yoshioka M, Gottdenker N, Chaves LF. Ecological patterns of blood-feeding by kissing-bugs (Hemiptera: Reduviidae: Triatominae). *Memorias Do Instituto Oswaldo Cruz*. 2011;106(4):479-94.
8. Dongiovanni C, Cavalieri V, Bodino N, Tauro D, Di Carolo M, Fumarola G, et al. Plant Selection and Population Trend of Spittlebug Immatures (Hemiptera: Aphrophoridae) in Olive Groves of the Apulia Region of Italy. *Journal of Economic Entomology*. 2019;112(1):67-74.
9. Chen DY, Chen FY, Chen CY, Chen XY, Mao YB. Transcriptome analysis of three cotton pests reveals features of gene expressions in the mesophyll feeder *Apolygus lucorum*. *Science China-Life Sciences*. 2017;60(8):826-38.
10. Hogenhout SA, Ammar ED, Whitfield AE, Redinbaugh MG. Insect vector interactions with persistently transmitted viruses. *Annual Review of Phytopathology*. 2008;46:327-59.
11. Whitfield AE, Falk BW, Rotenberg D. Insect vector-mediated transmission of plant viruses. *Virology*. 2015;479:278-89.
12. Chougule NP, Bonning BC. Toxins for Transgenic Resistance to Hemipteran Pests. *Toxins*. 2012;4(6):405-29.
13. Nalam V, Louis J, Shah J. Plant defense against aphids, the pest extraordinaire. *Plant Sci*. 2019;279:96-107.
14. Blackman RL. Life cycle variation of *Myzus persicae* (Sulz.) (Hom., Aphididae) in different parts of the world, in relation to genotype and environment. *Bulletin of Entomological Research*. 1974;63(4):595-607.
15. Koga R, Meng XY, Tsuchida T, Fukatsu T. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proc Natl Acad Sci U S A*. 2012;109(20):E1230-7.
16. Munson MA, Baumann P, Kinsey MG. *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a Taxon Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Aphids. *International Journal of Systematic and Evolutionary Microbiology*. 1991;41(4):566-8.

17. Powell G, Tosh CR, Hardie J. Host plant selection by aphids: Behavioral, evolutionary, and applied perspectives. *Annual Review of Entomology*. 2006;51:309-30.
18. Mugford ST, Barclay E, Druce C, Findlay KC, Hogenhout SA. An Immuno-Suppressive Aphid Saliva Protein Is Delivered into the Cytosol of Plant Mesophyll Cells During Feeding. *Mol Plant-Microbe Interact*. 2016;29(11):854-61.
19. Gravino M, Mugford ST, Kreuter N, Joyce J, Wilson C, Klot A, et al. The conserved aphid saliva chemosensory protein effector Mp10 targets plant AMSH deubiquitinases at cellular membranes to suppress pattern-triggered immunity. *bioRxiv*. 2024:2024.11.15.622802.
20. Liu Q, Neefjes ACM, Kobylinska R, Mugford ST, Marzo M, Canham J, et al. Aphid effectors suppress plant immunity via recruiting defence proteins to processing bodies. *bioRxiv*. 2024:2024.11.20.624400.
21. Mathers TC, Chen YZ, Kaithakottil G, Legeai F, Mugford ST, Baa-Puyoulet P, et al. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biol*. 2017;18:20.
22. Hardy NB, Peterson DA, von Dohlen CD. The evolution of life cycle complexity in aphids: Ecological optimization or historical constraint? *Evolution*. 2015;69(6):1423-32.
23. Fenton B, Woodford JAT, Malloch G. Analysis of clonal diversity of the peach-potato aphid, *Myzus persicae* (Sulzer), in Scotland, UK and evidence for the existence of a predominant clone. *Molecular Ecology*. 1998;7(11):1475-87.
24. Le Trionnaire G, Wucher V, Tagu D. Genome expression control during the photoperiodic response of aphids. *Physiological Entomology*. 2013;38(2):117-25.
25. Blackman RL. Variation in photoperiodic response within natural populations of *Myzus persicae* (Sulz). *Bulletin of Entomological Research*. 1971;60(4):533-&.
26. Le Trionnaire G, Tanguy S, Hudaverdian S, Gleonnec F, Richard G, Cayrol B, et al. An integrated protocol for targeted mutagenesis with CRISPR-Cas9 system in the pea aphid. *Insect Biochem Mol Biol*. 2019;110:34-44.
27. Yan S, Wang WX, Shen J. Reproductive polyphenism and its advantages in aphids: Switching between sexual and asexual reproduction. *Journal of Integrative Agriculture*. 2020;19(6):1447-57.
28. Miura T, Braendle C, Shingleton A, Sisk G, Kambhampati S, Stern DL. A comparison of parthenogenetic and sexual embryogenesis of the pea aphid *Acyrtosiphon pisum* (Hemiptera : Aphidoidea). *Journal of Experimental Zoology Part B- Molecular and Developmental Evolution*. 2003;295B(1):59-81.
29. Buning J. Morphology, ultrastructure, and germ-cell cluster formation in ovarioles of aphids. *Journal of Morphology*. 1985;186(2):209-21.
30. Srinivasan DG, Abdelhady A, Stern DL. Gene Expression Analysis of Parthenogenetic Embryonic Development of the Pea Aphid, *Acyrtosiphon pisum*, Suggests That Aphid Parthenogenesis Evolved from Meiotic Oogenesis. *Plos One*. 2014;9(12).
31. Chang C-c, Lin G-w, Cook CE, Horng S-b, Lee H-j, Huang T-y. Apvsa marks germ-cell migration in the parthenogenetic pea aphid *Acyrtosiphon pisum* (Hemiptera: Aphidoidea). *Development Genes and Evolution*. 2007;217(4):275-87.
32. Jaquiere J, Stoeckel S, Rispe C, Mieuze L, Legeai F, Simon JC. Accelerated Evolution of Sex Chromosomes in Aphids, an XO System. *Molecular Biology and Evolution*. 2012;29(2):837-47.

33. Li YY, Zhang B, Moran NA. The Aphid X Chromosome Is a Dangerous Place for Functionally Important Genes: Diverse Evolution of Hemipteran Genomes Based on Chromosome-Level Assemblies. *Molecular Biology and Evolution*. 2020;37(8):2357-68.
34. Jaquiere J, Rispe C, Roze D, Legeai F, Le Trionnaire G, Stoeckel S, et al. Masculinization of the X Chromosome in the Pea Aphid. *Plos Genetics*. 2013;9(8).
35. Bickel RD, Cleveland, H. C., Barkas, J., Jeschke, C. C., Raz, A. A., Stern, D. L., Davis, G. K. The pea aphid uses a version of the terminal system during oviparous, but not viviparous, development. *EvoDevo*. 2013;4:10.
36. De Barro PJ. The Bemisia tabaci Species Complex: Questions to Guide Future Research. *Journal of Integrative Agriculture*. 2012;11(2):187-96.
37. Boykin LM, Bell CD, Evans G, Small I, De Barro PJ. Is agriculture driving the diversification of the Bemisia tabaci species complex (Hemiptera: Sternorrhyncha: Aleyrodidae)? Dating, diversification and biogeographic evidence revealed. *BMC Evolutionary Biology*. 2013;13(1):228.
38. Azab AK, Megahed MM, El-Mirsawi HD. On the range of host-plants of Bemisia tabaci (Genn.) (Homoptera:Aleyrodidae). *Bulletin de la Societe Entomologique d’Egypte*. 1970;54:319–26.
39. Navas-Castillo J, Fiallo-Olivé E, Sánchez-Campos S. Emerging virus diseases transmitted by whiteflies. *Annu Rev Phytopathol*. 2011;49:219-48.
40. Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, et al. The draft genome of whitefly Bemisia tabaci MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biology*. 2016;14(1):110.
41. Shaaban A-R, Alvin MS. Survey of Reproductive Host Plants of Bemisia tabaci (Hemiptera: Aleyrodidae) in Egypt, Including New Host Records. *Entomological News*. 2010;121(5):456-65.
42. Wei J, He YZ, Guo Q, Guo T, Liu YQ, Zhou XP, et al. Vector development and vitellogenin determine the transovarial transmission of begomoviruses. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;114(26):6746-51.
43. Ban FX, Yin TY, Guo Q, Pan LL, Liu YQ, Wang XW. Localization and Quantification of Begomoviruses in Whitefly Tissues by Immunofluorescence and Quantitative PCR. *Jove-Journal of Visualized Experiments*. 2020(156).
44. Czosnek H, Hariton-Shalev A, Sobol I, Gorovits R, Ghanim M. The Incredible Journey of Begomoviruses in Their Whitefly Vector. *Viruses-Basel*. 2017;9(10).
45. Naalden D, van Kleeff PJM, Dangol S, Mastop M, Corkill R, Hogenhout SA, et al. Spotlight on the Roles of Whitefly Effectors in Insect-Plant Interactions. *Front Plant Sci*. 2021;12:661141.
46. Sani I, Ismail SI, Abdullah S, Jalinias J, Jamian S, Saad N. A Review of the Biology and Control of Whitefly, Bemisia tabaci (Hemiptera: Aleyrodidae), with Special Reference to Biological Control Using Entomopathogenic Fungi. *Insects*. 2020;11(9).
47. Wakil W, Brust GE, Perring T. Sustainable management of arthropod pests of tomato: Academic Press; 2017.
48. Liu S-S, De Barro PJ, Xu J, Luan J-B, Zang L-S, Ruan Y-M, et al. Asymmetric Mating Interactions Drive Widespread Invasion and Displacement in a Whitefly. *Science*. 2007;318(5857):1769-72.

49. Heu CC, McCullough FM, Luan J, Rasgon JL. CRISPR-Cas9-Based Genome Editing in the Silverleaf Whitefly (*Bemisia tabaci*). *Crispr j.* 2020;3(2):89-96.
50. Walker GP, Perring TM, Freeman TP. Life History, Functional Anatomy, Feeding and Mating Behavior. In: Stansly PA, Naranjo SE, editors. *Bemisia: Bionomics and Management of a Global Pest*. Dordrecht: Springer Netherlands; 2010. p. 109-60.
51. Mocellin S, Provenzano M. RNA interference: learning gene knock-down from cell physiology. *Journal of Translational Medicine.* 2004;2(1):39.
52. Li Q, Cheng Y, Fan J, Chen J. Metabolic relay gene of aphid and primary symbiont as RNAi target loci for aphid control. *Frontiers in Plant Science.* 2023;Volume 13 - 2022.
53. Mutti NS, Park Y, Reese JC, Reeck GR. RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrtosiphon pisum*. *J Insect Sci.* 2006;6:1-7.
54. Shahid S, Amir MB, Ding T-B, Liu T-X, Smagghe G, Shi Y. RNAi of Neuropeptide CCHamide-1 and Its Receptor Indicates Role in Feeding Behavior in the Pea Aphid, *Acyrtosiphon pisum*. *Insects* [Internet]. 2024; 15(12). Available from: https://mdpi-res.com/d_attachment/insects/insects-15-00939/article_deploy/insects-15-00939-v2.pdf?version=1732874481.
55. Xu L, Hou Q, Zhao Y, Lu L, Li B, Ni Z, et al. Silencing of a lipase maturation factor 2-like gene by wheat-mediated RNAi reduces the survivability and reproductive capacity of the grain aphid, *Sitobion avenae*. *Archives of Insect Biochemistry and Physiology.* 2017;95(3):e21392.
56. Mao J, Zeng F. Plant-mediated RNAi of a gap gene-enhanced tobacco tolerance against the *Myzus persicae*. *Transgenic Research.* 2014;23(1):145-52.
57. Chen Y, Singh A, Kaithakottil GG, Mathers TC, Gravino M, Mugford ST, et al. An aphid RNA transcript migrates systemically within plants and is a virulence factor. *Proceedings of the National Academy of Sciences.* 2020;117(23):12763-71.
58. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science.* 2012;337(6096):816-21.
59. Terns MP, Terns RM. CRISPR-based adaptive immune systems. *Curr Opin Microbiol.* 2011;14(3):321-7.
60. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science.* 2008;321(5891):960-4.
61. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol.* 2011;9(6):467-77.
62. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* 2007;315(5819):1709-12.
63. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct.* 2006;1:26.
64. Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, et al. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *Journal of Biological Chemistry.* 2008;283(29):20361-71.

65. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol.* 2020;18(2):67-83.
66. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol.* 2015;13(11):722-36.
67. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao YJ, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature.* 2011;471(7340):602-+.
68. Strecker J, Jones S, Koopal B, Schmid-Burgk J, Zetsche B, Gao LY, et al. Engineering of CRISPR-Cas12b for human genome editing. *Nature Communications.* 2019;10.
69. Swarts DC, Jinek M. Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a. *Molecular Cell.* 2019;73(3):589-+.
70. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell.* 2015;163(3):759-71.
71. Pausch P, Al-Shayeb B, Bisom-Rapp E, Tsuchida CA, Li Z, Cress BF, et al. CRISPR-Cas Phi from huge phages is a hypercompact genome editor. *Science.* 2020;369(6501):333-+.
72. Wu Z, Zhang Y, Yu H, Pan D, Wang Y, Wang Y, et al. Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nature Chemical Biology.* 2021.
73. Meeske AJ, Nakandakari-Higa S, Marraffini LA. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature.* 2019;570(7760):241-+.
74. Huynh N, Depner N, Larson R, King-Jones K. A versatile toolkit for CRISPR-Cas13-based RNA manipulation in *Drosophila*. *Genome Biol.* 2020;21(1):29.
75. Srivastava S, Upadhyay DJ, Srivastava A. Next-Generation Molecular Diagnostics Development by CRISPR/Cas Tool: Rapid Detection and Surveillance of Viral Disease Outbreaks. *Frontiers in Molecular Biosciences.* 2020;7.
76. Arizti-Sanz J, Freije CA, Stanton AC, Petros BA, Boehm CK, Siddiqui S, et al. Streamlined inactivation, amplification, and Cas13-based detection of SARS-CoV-2. *Nature Communications.* 2020;11(1).
77. Cong L, Ran FA, Cox D, Lin SL, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science.* 2013;339(6121):819-23.
78. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 2013;41(7):4336-43.
79. Dickinson DJ, Ward JD, Reiner DJ, Goldstein B. Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nature Methods.* 2013;10(10):1028-+.
80. Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, et al. Genome Engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics.* 2013;194(4):1029-+.
81. Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, et al. Highly Specific and Efficient CRISPR/Cas9-Catalyzed Homology-Directed Repair in *Drosophila*. *Genetics.* 2014;196(4):961-+.

82. Hwang WY, Fu YF, Reyon D, Maeder ML, Tsai SQ, Sander JD, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology*. 2013;31(3):227-9.
83. Jiang WY, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*. 2013;31(3):233-9.
84. Mali P, Yang LH, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-Guided Human Genome Engineering via Cas9. *Science*. 2013;339(6121):823-6.
85. Wang HY, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell*. 2013;153(4):910-8.
86. Dong YM, Simoes ML, Marois E, Dimopoulos G. CRISPR/Cas9-mediated gene knockout of *Anopheles gambiae* FREP1 suppresses malaria parasite infection. *Plos Pathogens*. 2018;14(3).
87. Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C, Katsanos D, et al. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. *Nat Biotechnol*. 2016;34(1):78-83.
88. Kyrou K, Hammond AM, Galizi R, Kranjc N, Burt A, Beaghton AK, et al. A CRISPR-Cas9 gene drive targeting doublesex causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nature Biotechnology*. 2018;36(11):1062-+.
89. Macias VM, McKeand S, Chaverra-Rodriguez D, Hughes GL, Fazekas A, Pujhari S, et al. Cas9-Mediated Gene-Editing in the Malaria Mosquito *Anopheles stephensi* by ReMOT Control. *G3 (Bethesda)*. 2020;10(4):1353-60.
90. Chaverra-Rodriguez D, Macias VM, Hughes GL, Pujhari S, Suzuki Y, Peterson DR, et al. Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. *Nat Commun*. 2018;9(1):3008.
91. Gantz VM, Akbari OS. Gene editing technologies and applications for insects. *Curr Opin Insect Sci*. 2018;28:66-72.
92. Kanca O, Zirin J, Garcia-Marques J, Knight SM, Yang-Zhou D, Amador G, et al. An efficient CRISPR-based strategy to insert small and large fragments of DNA using short homology arms. *Elife*. 2019;8.
93. Corkill R. Embryo development and sex determination in the Cassava whitefly (*Bemisia tabaci*). [Doctoral Thesis]: University of East Anglia; 2019.
94. Li X, Xu Y, Zhang H, Yin H, Zhou D, Sun Y, et al. ReMOT Control Delivery of CRISPR-Cas9 Ribonucleoprotein Complex to Induce Germline Mutagenesis in the Disease Vector Mosquitoes *Culex pipiens pallens* (Diptera: Culicidae). *Journal of medical entomology*. 2021.
95. Yang X-l, Ling X, Sun Q, Qiu P-p, Xiang K, Hong J-f, et al. High-efficiency gene editing in *Anopheles sinensis* using ReMOT control. *Insect Science*. 2024;31(1):307-12.
96. Chaverra-Rodriguez D, Dalla Benetta E, Heu CC, Rasgon JL, Ferree PM, Akbari OS. Germline mutagenesis of *Nasonia vitripennis* through ovarian delivery of CRISPR-Cas9 ribonucleoprotein. *Insect Molecular Biology*. 2020.
97. Shirai Y, Daimon T. Mutations in cardinal are responsible for the red-1 and peach eye color mutants of the red flour beetle *Tribolium castaneum*. *Biochemical and Biophysical Research Communications*. 2020;529(2):372-8.
98. Lima L, Berni M, Mota J, Bressan D, Julio A, Cavalcante R, et al. Gene Editing in the Chagas Disease Vector *Rhodnius prolixus* by Cas9-Mediated ReMOT Control. *The CRISPR Journal*. 2024.

99. Chaverra-Rodriguez D, Bui M, Gilleland CL, Rasgon JL, Akbari OS. CRISPR-Cas9-Mediated Mutagenesis of the Asian Citrus Psyllid, *Diaphorina citri*. GEN Biotechnology. 2023;2(4):317-29.
100. Yu B, Dong S, Jiang X, Qiao L, Chen J, Li T, et al. Cas9-Mediated Gene Editing Using Receptor-Mediated Ovary Transduction of Cargo (ReMOT) Control in *Bombyx mori*. Insects [Internet]. 2023; 14(12). Available from: https://mdpi-res.com/d_attachment/insects/insects-14-00932/article_deploy/insects-14-00932.pdf?version=1701936639.
101. Sharma A, Pham MN, Reyes JB, Chana R, Yim WC, Heu CC, et al. Cas9-mediated gene editing in the black-legged tick, *Ixodes scapularis*, by embryo injection and ReMOT Control. iScience. 2022.
102. Roth Z, Weil S, Aflalo ED, Manor R, Sagi A, Khalaila I. Identification of receptor-interacting regions of vitellogenin within evolutionarily conserved beta-sheet structures by using a peptide array. Chembiochem. 2013;14(9):1116-22.
103. Sappington TW, Raikhel AS. Molecular characteristics of insect vitellogenins and vitellogenin receptors. Insect Biochemistry and Molecular Biology. 1998;28(5-6):277-300.
104. Tufail M, Takeda M. Molecular characteristics of insect vitellogenins. Journal of Insect Physiology. 2008;54(12):1447-58.
105. Shang F, Niu JZ, Ding BY, Zhang Q, Ye C, Zhang W, et al. Vitellogenin and its receptor play essential roles in the development and reproduction of the brown citrus aphid, *Aphis (Toxoptera) citricidus*. Insect Molecular Biology. 2018;27(2):221-33.
106. Li A, Sadasivam M, Ding JL. Receptor-Ligand Interaction between Vitellogenin Receptor (VtgR) and Vitellogenin (Vtg), Implications on Low Density Lipoprotein Receptor and Apolipoprotein B/E: The first three ligand-binding repeats of VtgR interact with the amino-terminal region of Vtg. Journal of Biological Chemistry. 2003;278(5):2799-806.
107. Wu ZX, Yang LB, He QJ, Zhou ST. Regulatory Mechanisms of Vitellogenesis in Insects. Frontiers in Cell and Developmental Biology. 2021;8.
108. Roy SG, Raikhel AS. The small GTPase Rheb is a key component linking amino acid signaling and TOR in the nutritional pathway that controls mosquito egg development. Insect Biochem Mol Biol. 2011;41(1):62-9.
109. Roy S, Saha TT, Zou Z, Raikhel AS. Regulatory Pathways Controlling Female Insect Reproduction. Annual Review of Entomology, Vol 63. 2018;63:489-511.
110. Dhara A, Eum J-H, Robertson A, Gulia-Nuss M, Vogel KJ, Clark KD, et al. Ovary ecdysteroidogenic hormone functions independently of the insulin receptor in the yellow fever mosquito, *Aedes aegypti*. Insect Biochemistry and Molecular Biology. 2013;43(12):1100-8.
111. Kokoza VA, Martin D, Mienaltowski MJ, Ahmed A, Morton CM, Raikhel AS. Transcriptional regulation of the mosquito vitellogenin gene via a blood meal-triggered cascade. Gene. 2001;274(1):47-65.
112. Ling L, Raikhel AS. Amino acid-dependent regulation of insulin-like peptide signaling is mediated by TOR and GATA factors in the disease vector mosquito *Aedes aegypti*. Proceedings of the National Academy of Sciences. 2023;120(34):e2303234120.
113. Charles J-P, Iwema T, Epa VC, Takaki K, Rynes J, Jindra M. Ligand-binding properties of a juvenile hormone receptor, Methoprene-tolerant. Proceedings of the National Academy of Sciences. 2011;108:21128-33.

114. Li M, Mead EA, Zhu J. Heterodimer of two bHLH-PAS proteins mediates juvenile hormone-induced gene expression. *Proceedings of the National Academy of Sciences*. 2011;108:638-43.
115. Jindra M, Palli SR, Riddiford LM. The juvenile hormone signaling pathway in insect development. *Annu Rev Entomol*. 2013;58:181-204.
116. Ojani R, Fu X, Ahmed T, Liu P, Zhu J. Krüppel homologue 1 acts as a repressor and an activator in the transcriptional response to juvenile hormone in adult mosquitoes. *Insect Molecular Biology*. 2018;27(2):268-78.
117. Gujar H, Palli SR. Juvenile hormone regulation of female reproduction in the common bed bug, *Cimex lectularius*. *Scientific Reports*. 2016;6.
118. Guo W, Wu Z, Song J, Jiang F, Wang Z, Deng S, et al. Juvenile Hormone-Receptor Complex Acts on Mcm4 and Mcm7 to Promote Polyploidy and Vitellogenesis in the Migratory Locust. *PLOS Genetics*. 2014;10(10):e1004702.
119. Wu Z, Guo W, Xie Y, Zhou S. Juvenile Hormone Activates the Transcription of Cell-division-cycle 6 (Cdc6) for Polyploidy-dependent Insect Vitellogenesis and Oogenesis*. *Journal of Biological Chemistry*. 2016;291(10):5418-27.
120. Wu Z, Guo W, Yang L, He Q, Zhou S. Juvenile hormone promotes locust fat body cell polyploidization and vitellogenesis by activating the transcription of Cdk6 and E2f1. *Insect Biochemistry and Molecular Biology*. 2018;102:1-10.
121. Wu Z, He Q, Zeng B, Zhou H, Zhou S. Juvenile hormone acts through FoxO to promote Cdc2 and Orc5 transcription for polyploidy-dependent vitellogenesis. *Development*. 2020;147(18).
122. Wang J-L, Saha TT, Zhang Y, Zhang C, Raikhel AS. Juvenile hormone and its receptor methoprene-tolerant promote ribosomal biogenesis and vitellogenesis in the *Aedes aegypti* mosquito. *Journal of Biological Chemistry*. 2017;292(24):10306-15.
123. Saha TT, Roy S, Pei G, Dou W, Zou Z, Raikhel AS. Synergistic action of the transcription factors Krüppel homolog 1 and Hairy in juvenile hormone/Methoprene-tolerant-mediated gene-repression in the mosquito *Aedes aegypti*. *PLOS Genetics*. 2019;15(10):e1008443.
124. Swevers L. An update on ecdysone signaling during insect oogenesis. *Current Opinion in Insect Science*. 2019;31:8-13.
125. Schwedes C, Tulsiani S, Carney GE. Ecdysone receptor expression and activity in adult *Drosophila melanogaster*. *Journal of Insect Physiology*. 2011;57(7):899-907.
126. Yao T-P, Segraves WA, Oro AE, McKeown M, Evans RM. *Drosophila* ultraspiracle modulates ecdysone receptor function via heterodimer formation. *Cell*. 1992;71(1):63-72.
127. Upadhyay SK, Singh H, Dixit S, Mendu V, Verma PC. Molecular Characterization of Vitellogenin and Vitellogenin Receptor of *Bemisia tabaci*. *PLoS One*. 2016;11(5):e0155306.
128. Shen Y, Chen YZ, Lou YH, Zhang CX. Vitellogenin and Vitellogenin-Like Genes in the Brown Planthopper. *Front Physiol*. 2019;10:1181.
129. Li Z, Zhang S, Liu Q. Vitellogenin Functions as a Multivalent Pattern Recognition Receptor with an Opsonic Activity. *PLOS ONE*. 2008;3(4):e1940.
130. Sun W, Li H, Zhao Y, Bai L, Qin Y, Wang Q, et al. Distinct vitellogenin domains differentially regulate immunological outcomes in invertebrates. *Journal of Biological Chemistry*. 2021;296:100060.

131. Wu B, Liu Z, Zhou L, Ji G, Yang A. Molecular cloning, expression, purification and characterization of vitellogenin in scallop *Patinopecten yessoensis* with special emphasis on its antibacterial activity. *Developmental & Comparative Immunology*. 2015;49(2):249-58.
132. Singh NK, Pakkianathan BC, Kumar M, Prasad T, Kannan M, König S, et al. Vitellogenin from the Silkworm, *Bombyx mori*: An Effective Anti-Bacterial Agent. *PLOS ONE*. 2013;8(9):e73005.
133. Salmela H, Amdam GV, Freitak D. Transfer of Immunity from Mother to Offspring Is Mediated via Egg-Yolk Protein Vitellogenin. *PLOS Pathogens*. 2015;11(7):e1005015.
134. Huang H-J, Lu J-B, Li Q, Bao Y-Y, Zhang C-X. Combined transcriptomic/proteomic analysis of salivary gland and secreted saliva in three planthopper species. *Journal of Proteomics*. 2018;172:25-35.
135. Wang Y, Lu C, Guo S, Guo Y, Wei T, Chen Q. Leafhopper salivary vitellogenin mediates virus transmission to plant phloem. *Nature Communications*. 2024;15(1):3.
136. Leipart V, Montserrat-Canals M, Cunha ES, Luecke H, Herrero-Galán E, Halskau Ø, et al. Structure prediction of honey bee vitellogenin: a multi-domain protein important for insect immunity. *FEBS Open Bio*. 2021;n/a(n/a).
137. Guo J-Y, Wan F-H, Ye G-Y. Oogenesis in the *Bemisia tabaci* MEAM1 species complex. *Micron*. 2016;83:1-10.
138. Anderson TA, Levitt DG, Banaszak LJ. The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein. *Structure*. 1998;6(7):895-909.
139. Sharrock WJ, Rosenwasser TA, Gould J, Knott J, Hussey D, Gordon JL, et al. Sequence of lamprey vitellogenin: Implications for the lipovitellin crystal structure. *Journal of Molecular Biology*. 1992;226(3):903-7.
140. Mathers TC, Wouters RHM, Mugford ST, Swarbreck D, van Oosterhout C, Hogenhout SA. Chromosome-Scale Genome Assemblies of Aphids Reveal Extensively Rearranged Autosomes and Long-Term Conservation of the X Chromosome. *Molecular Biology and Evolution*. 2021;38(3):856-75.
141. Mathers TC, Mugford ST, Percival-Alwyn L, Chen Y, Kaithakottil G, Swarbreck D, et al. Sex-specific changes in the aphid DNA methylation landscape. *Molecular Ecology*. 2019;28(18):4228-41.
142. Li Z, Zhang S, Zhang J, Liu M, Liu Z. Vitellogenin is a cidal factor capable of killing bacteria via interaction with lipopolysaccharide and lipoteichoic acid. *Molecular Immunology*. 2009;46(16):3232-9.
143. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar Gustavo A, et al. InterPro in 2022. *Nucleic Acids Res*. 2023;51(D1):D418-D27.
144. Dalla Benetta E, Chaverra-Rodriguez D, Rasgon JL, Akbari OS. Pupal and Adult Injections for RNAi and CRISPR Gene Editing in *Nasonia vitripennis*. *J Vis Exp*. 2020(166).
145. Rispe C, Legeai F, Nabity PD, Fernández R, Arora AK, Baa-Puyoulet P, et al. The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC Biology*. 2020;18(1):90.
146. Johnson KP, Dietrich CH, Friedrich F, Beutel RG, Wipfler B, Peters RS, et al. Phylogenomics and the evolution of hemipteroid insects. *Proceedings of the National Academy of Sciences*. 2018;115(50):12775-80.
147. Camacho ER, Chong J-H. General Biology and Current Management Approaches of Soft Scale Pests (Hemiptera: Coccidae). *J Integr Pest Manag*. 2015;6(1):17.

148. Mongue AJ, Michaelides S, Coombe O, Tena A, Kim D-S, Normark BB, et al. Sex, males, and hermaphrodites in the scale insect *Icerya purchasi**. *Evolution*. 2021;75(11):2972-83.
149. Gavrilov-Zimin IA. Development of Theoretical Views on Viviparity. *Biology Bulletin Reviews*. 2022;12(6):570-95.
150. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021.
151. Mathers TC, Mugford ST, Wouters RHM, Heavens D, Botha A-M, Swarbreck D, et al. Aphidinae comparative genomics resource (Version v1) [Data set]. Zenodo. 2022.
152. Thorpe P, Escudero-Martinez CM, Cock PJA, Eves-van den Akker S, Bos JIB. Shared Transcriptional Control and Disparate Gain and Loss of Aphid Parasitism Genes. *Genome Biology and Evolution*. 2018;10(10):2716-33.
153. Mathers TC, Wouters RHM, Mugford ST, Biello R, van Oosterhout C, Hogenhout SA. Hybridisation has shaped a recent radiation of grass-feeding aphids. *BMC Biology*. 2023;21(1):157.
154. Liu Q, Goldberg JK, Mugford ST, Saalback G, Martins C, Singh A, et al. The salivary proteome of the green peach aphid/peach-potato aphid (*Myzus persicae*) (Sulzer, 1776) (Hemiptera, Aphididae). Zenodo. 2024.
155. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792-7.
156. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006;22(21):2688-90.
157. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. 2024;52(W1):W78-W82.
158. Wickham H AM, Bryan J, Chang W, McGowan LD, François R, Grolemond G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. Welcome to the Tidyverse. *Journal of Open Source Software*. 2019;4(43):1686.
159. Grosjean P. SciViews-R. UMONS, MONS, Belgium.2022.
160. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. *Nature Biotechnology*. 2024;42(2):243-6.
161. R K. pheatmap: Pretty Heatmaps. R package version 1.0.13 ed2025.
162. Lambert TJ. FPbase: a community-editable fluorescent protein database. *Nature Methods*. 2019;16(4):277-8.
163. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nature Methods*. 2012;9(7):676-82.
164. Bentham AR, Youles M, Mendel MN, Varden FA, Concepcion JCDI, Banfield MJ. pOPIN-GG: A resource for modular assembly in protein expression vectors. *bioRxiv*. 2021:2021.08.10.455798.
165. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, editor. *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press; 2005. p. 571-607.
166. Terradas G, Macias VM, Peterson H, McKeand S, Krawczyk G, Rasgon JL. The Development and Expansion of in vivo Germline Editing Technologies in Arthropods:

- Receptor-Mediated Ovary Transduction of Cargo (ReMOT Control) and Beyond. *Integrative and Comparative Biology*. 2023;63(6):1550-63.
167. Shirai Y, Piulachs M-D, Belles X, Daimon T. DIPA-CRISPR is a simple and accessible method for insect gene editing. *Cell Reports Methods*. 2022;2(5):100215.
168. Takenaka A, Konno H, Kikuta S. In vivo direct cell-penetrating peptide mediated protein transduction system in *Acyrtosiphon pisum*. *BMC Research Notes*. 2023;16(1):231.
169. Derossi D, Joliot AH, Chassaing G, Prochiantz A. The third helix of the Antennapedia homeodomain translocates through biological membranes. *Journal of Biological Chemistry*. 1994;269(14):10444-50.
170. Bulcha JT, Wang Y, Ma H, Tai PWL, Gao G. Viral vector platforms within the gene therapy landscape. *Signal Transduction and Targeted Therapy*. 2021;6(1):53.
171. Huang Q, Chen AT, Chan KY, Sorensen H, Barry AJ, Azari B, et al. Targeting AAV vectors to the central nervous system by engineering capsid–receptor interactions that enable crossing of the blood–brain barrier. *PLOS Biology*. 2023;21(7):e3002112.
172. Peyret H, Lomonossoff GP. Specific Packaging of Custom RNA Molecules into Cowpea Mosaic Virus-like Particles (VLPs). In: Schillberg S, Spiegel H, editors. *Recombinant Proteins in Plants: Methods and Protocols*. New York, NY: Springer US; 2022. p. 103-11.
173. Porta C, Spall VE, Loveland J, Johnson JE, Barker PJ, Lomonossoff GP. Development of Cowpea Mosaic Virus as a High-Yielding System for the Presentation of Foreign Peptides. *Virology*. 1994;202(2):949-55.
174. Reinbold C, Gildow FE, Herrbach E, Ziegler-Graff V, Gonçalves MC, van den Heuvel JFJM, et al. Studies on the role of the minor capsid protein in transport of Beet western yellows virus through *Myzus persicae*. *Journal of General Virology*. 2001;82(8):1995-2007.
175. Engler C, Kandzia R, Marillonnet S. A one pot, one step, precision cloning method with high throughput capability. *PLoS One*. 2008;3(11):e3647.
176. A K. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0. 2023.
177. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015;67(1):1 - 48.
178. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. 2009;6(5):343-5.
179. Villaverde A, Carrió MM. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol Lett*. 2003;25(17):1385-95.
180. Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM. Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophys J*. 2012;102(8):1907-15.
181. Gronenborn AM, Filpula DR, Essig NZ, Achari A, Whitlow M, Wingfield PT, et al. A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G. *Science*. 1991;253(5020):657-61.
182. Bao W-J, Gao Y-G, Chang Y-G, Zhang T-Y, Lin X-J, Yan X-Z, et al. Highly efficient expression and purification system of small-size protein domains in *Escherichia coli* for biochemical characterization. *Protein Expression and Purification*. 2006;47(2):599-606.

183. Song SJ, Diao HP, Moon B, Yun A, Hwang I. The B1 Domain of Streptococcal Protein G Serves as a Multi-Functional Tag for Recombinant Protein Production in Plants. *Front Plant Sci.* 2022;13:878677.
184. Mandrioli M, Rivi V, Nardelli A, Manicardi GC. Genomic and Cytogenetic Localization of the Carotenoid Genes in the Aphid Genome. *Cytogenetic and Genome Research.* 2016;149(3):207-17.
185. Nováková E, Moran NA. Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Mol Biol Evol.* 2012;29(1):313-23.
186. Sharma RP, Chopra VL. Effect of the wingless (wg1) mutation on wing and haltere development in *Drosophila melanogaster*. *Developmental Biology.* 1976;48(2):461-5.
187. Swarup S, Verheyen EM. Wnt/Wingless signaling in *Drosophila*. *Cold Spring Harb Perspect Biol.* 2012;4(6).
188. Bejsovec A. Wingless Signaling: A Genetic Journey from Morphogenesis to Metastasis. *Genetics.* 2018;208:1311-36.
189. Sunkel CE, Whittle JRS. Brista: a gene involved in the specification and differentiation of distal cephalic and thoracic structures in *Drosophila melanogaster*. *Roux's archives of developmental biology.* 1987;196(2):124-32.
190. Cohen SM, Brönner G, Küttner F, Jürgens G, Jäckle H. Distal-less encodes a homoeodomain protein required for limb development in *Drosophila*. *Nature.* 1989;338(6214):432-4.
191. Judd BH. The white Locus of *Drosophila melanogaster*. In: Hennig W, editor. *Structure and Function of Eukaryotic Chromosomes.* Berlin, Heidelberg: Springer Berlin Heidelberg; 1987. p. 81-94.
192. Mackenzie SM, Brooker MR, Gill TR, Cox GB, Howells AJ, Ewart GD. Mutations in the white gene of *Drosophila melanogaster* affecting ABC transporters that determine eye colouration. *Biochim Biophys Acta-Biomembr.* 1999;1419(2):173-85.
193. Ramakrishna S, Kwaku Dad A-B, Beloor J, Gopalappa R, Lee S-K, Kim H. Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Research.* 2014;24(6):1020-7.
194. Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 2021;49(D1):D344-D54.
195. Biolabs NE. In vitro digestion of DNA with Cas9 Nuclease, *S. pyogenes* (M0386). protocolsio. 2014.
196. Komal J, Desai HR, Samal I, Mastinu A, Patel RD, Kumar PVD, et al. Unveiling the Genetic Symphony: Harnessing CRISPR-Cas Genome Editing for Effective Insect Pest Management. *Plants* [Internet]. 2023; 12(23). Available from: https://mdpi-res.com/d_attachment/plants/plants-12-03961/article_deploy/plants-12-03961-v2.pdf?version=1700989041.
197. Margaritopoulos JT, Tsitsipis JA, Goudoudaki S, Blackman RL. Life cycle variation of *Myzus persicae* (Hemiptera : Aphididae) in Greece. *Bulletin of Entomological Research.* 2002;92(4):309-19.
198. Mugford ST, Osbourn A. Saponin Synthesis and Function. *Isoprenoid Synthesis in Plants and Microorganisms.* 2012:405-24.
199. De Geyter E, Smagghe G, Rahbé Y, Geelen D. Triterpene saponins of *Quillaja saponaria* show strong aphicidal and deterrent activity against the pea aphid *Acyrtosiphon pisum*. *Pest Manag Sci.* 2012;68(2):164-9.

200. Li Y, Hao N, Ye S, Hu Z, Zhao L, Qi Y, et al. New triterpenoid saponins from *Clematis lasiandra* and their mode of action against pea aphids *Acyrtosiphon pisum*. *Industrial Crops and Products*. 2022;187:115517.
201. Goławska S, Sprawka I, Łukasik I. Effect of saponins and apigenin mixtures on feeding behavior of the pea aphid, *Acyrtosiphon pisum* Harris. *Biochemical Systematics and Ecology*. 2014;55:137-44.
202. Sylwia G, Leszczynski B, Wiesław O. Effect of low and high-saponin lines of alfalfa on pea aphid. *J Insect Physiol*. 2006;52(7):737-43.
203. Hong F, Han H-L, Pu P, Wei D, Wang J, Liu Y. Effects of Five Host Plant Species on the Life History and Population Growth Parameters of *Myzus persicae* (Hemiptera: Aphididae). *Journal of Insect Science*. 2019;19(5):15.
204. Goggin FL, Lorence A, Topp CN. Applying high-throughput phenotyping to plant–insect interactions: picturing more resistant crops. *Current Opinion in Insect Science*. 2015;9:69-76.
205. Złotkowska E, Wlazło A, Kietkiewicz M, Misztal K, Dziosa P, Soja K, et al. Automated imaging coupled with AI-powered analysis accelerates the assessment of plant resistance to *Tetranychus urticae*. *Scientific Reports*. 2024;14(1):8020.
206. Gelman DB, Hu JS. Critical feeding periods for last instar nymphal and pharate adults of the whiteflies, *Trialeurodes vaporariorum* and *Bemisia tabaci*. *Journal of Insect Science*. 2007;7(1):33.
207. Sun J, Hiraoka T, Dittmer NT, Cho K-H, Raikhel AS. Lipophorin as a yolk protein precursor in the mosquito, *Aedes aegypti*. *Insect Biochemistry and Molecular Biology*. 2000;30(12):1161-71.
208. Barberà M, Cañas-Cañas R, Martínez-Torres D. Insulin-like peptides involved in photoperiodism in the aphid *Acyrtosiphon pisum*. *Insect Biochemistry and Molecular Biology*. 2019;112:103185.
209. Cuti P, Barberà M, Veenstra JA, Martínez-Torres D. Progress in the characterization of insulin-like peptides in aphids: Immunohistochemical mapping of ILP4. *Insect Biochemistry and Molecular Biology*. 2021;136:103623.
210. Synthego Performance Analysis, ICE Analysis. v3.0 2019 [Available from: <https://ice.synthego.com/#/>].
211. Singh KS, Cordeiro EMG, Troczka BJ, Pym A, Mackisack J, Mathers TC, et al. Global patterns in genomic diversity underpinning the evolution of insecticide resistance in the aphid crop pest *Myzus persicae*. *Commun Biol*. 2021;4(1):847.
212. Sun D, Guo Z, Liu Y, Zhang Y. Progress and Prospects of CRISPR/Cas Systems in Insects and Other Arthropods. *Front Physiol*. 2017;8:608.
213. Li JJ, Shi Y, Wu JN, Li H, Smagghe G, Liu TX. CRISPR/Cas9 in lepidopteran insects: Progress, application and prospects. *J Insect Physiol*. 2021;135:104325.
214. Pausch P, Soczek KM, Herbst DA, Tsuchida CA, Al-Shayeb B, Banfield JF, et al. DNA interference states of the hypercompact CRISPR–CasΦ effector. *Nature Structural & Molecular Biology*. 2021;28(8):652-61.
215. Li Z, Zhong Z, Wu Z, Pausch P, Al-Shayeb B, Amerasekera J, et al. Genome editing in plants using the compact editor CasΦ. *Proceedings of the National Academy of Sciences*. 2023;120(4):e2216822120.
216. Richardson CD, Ray GJ, Bray NL, Corn JE. Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. *Nature Communications*. 2016;7(1):12463.

217. Blackman RL. Photoperiodic determination of the male and female sexual morphs of *Myzus persicae*. *Journal of Insect Physiology*. 1975;21(2):435-53.
218. Perreau J, Patel DJ, Anderson H, Maeda GP, Elston KM, Barrick JE, et al. Vertical Transmission at the Pathogen-Symbiont Interface: *Serratia symbiotica* and Aphids. *mBio*. 2021;12(2).
219. Hampton HG, McNeil MB, Paterson TJ, Ney B, Williamson NR, Easingwood RA, et al. CRISPR-Cas gene-editing reveals RsmA and RsmC act through FlhDC to repress the SdhE flavinylation factor and control motility and prodigiosin production in *Serratia*. *Microbiology*. 2016;162(6):1047-58.
220. Hegde S, Rauch HE, Hughes GL, Shariat N. Identification and characterization of two CRISPR/Cas systems associated with the mosquito microbiome. *Access Microbiol*. 2023;5(8).
221. Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife*. 2014;3:e04766.
222. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*. 2019;576(7785):149-57.
223. Bosch JA, Birchak G, Perrimon N. Precise genome engineering in *Drosophila* using prime editing. *Proc Natl Acad Sci U S A*. 2021;118(1).