Imperial College London

Building a Molecular Toolkit for Gene Editing and Transgenesis in the Farmed Insect Black Soldier Fly

James Daniel Fawcett

A thesis submitted for the degree of

Doctor of Philosophy

Department of Life Sciences, Imperial College London

Primary supervisor:

Dr. Nikolai Windbichler

Co-supervisors:

Dr. Angela Meccariello

Dr. Thomas Farrugia, Beta Bugs Ltd.

November 2024

Statement of Originality

I hereby declare that the work described in this thesis is original and was performed and reported solely by myself, unless otherwise stated. Any external contributions have been cited and acknowledged.

James Daniel Fawcett

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International Licence (CC BY-NC-ND).

Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Current agricultural methods are expected to fall short in addressing food security threats, with a pressing need for higher yields and greater resilience to climate variability. A promising solution for sustainable food security is the revalorization of food waste. Insect-based bioremediation is an innovative approach to tackling food waste, with several insect species demonstrating exceptional efficiency in converting low-value waste into valuable resources. The black soldier fly (BSF) stands out for its ability to effectively process waste into energy-dense larvae suitable for animal feed. Beyond this, BSF holds potential for producing fertilizers, biofuels, and many other valuable products, attracting interest from a burgeoning insect farming industry.

While black soldier fly is already an efficient farmed insect, a need exists for enabling technologies that allow for genetic optimization towards desirable traits, such as larger weight and shorter development times. However, foundational research is necessary to develop more refined tools for genetic manipulation in BSF. The endogenous application of the CRISPR gene editing system could be particularly beneficial, streamlining the validation of advantageous mutant genotypes and accelerating the improvement of desirable traits.

In this study, a molecular toolkit was developed for BSF to enable CRISPR production from an endogenous transgenic source, aimed at enhancing gene editing efficiency and accessibility for both the research and commercial sectors. This toolkit included the development and optimization of an efficient genetic transduction method, the embryo microinjection technique, which has yet to be thoroughly documented in the literature. Overall, this project aimed to build on the molecular toolkit currently available to BSF researchers, enabling more complex genetic investigation and manipulations to drive the development of novel, optimised BSF breeding lines. Enhancing the genetic potential of this insect could provide significant benefits as an innovative solution for managing food waste and strengthening sustainable food security.

Acknowledgements

Throughout this PhD project, I've been fortunate enough to meet and work with many wonderful people, without whom the whole experience would have been a little less memorable and felt far less rewarding. Firstly, I need to thank Nikolai Windbichler for setting this project in motion, for hosting me in his lab at Imperial College London, and for giving me ample space to grow as an independent researcher and fully shape this project into my own. I owe a huge debt of gratitude to Angela Meccariello for the comprehensive introduction to the weird world of insect biotechnology, providing crucial support and words of wisdom as I built my own insect colony, and for embodying the drive, energy and focus it takes to really thrive in academic research.

To all the members of the Windbichler and Meccariello labs, particularly Julia, Pepe, Paolo, Maria-Grazia and Pei-Shi, thank you all so much for making me feel welcome and supporting me in and outside of the lab; from this whole experience, I think I'll miss our visits to the Queens Arms and Eastside Bar the most. Sima and Jonathan, I can't thank you enough for keeping everything up and running while I was away, for the countless useful discussions and suggestions, and for always being there to talk with when I needed the help. A special thanks again to Moeez Khan for his kind offer to help with immunoblotting experiments, I really appreciate it. Thanks to all of my fellow trainees, and especially Kim and Daniel, on the BioDesign Engineering CDT program for the brilliant exchanges of ideas and perspectives, as well as for all the laughs. My gratitude to Geoff Baldwin and David Andrews for building a program with such great talent and promise, and for securing my place on it. Lastly, thanks to Tony Southall and Simon Lillico for providing an excellent viva experience, and for their thorough feedback and meaningful advice.

This study could not have gone forward without the generous support and sponsorship from Beta bugs Ltd, who facilitated the project with direction, materials and an abundance of knowhow. In particular, I'd like to thank Sam Grainger for offering his time and expertise during my visits to Roslin and for his help with establishing the Imperial BSF colony. I'd also like to thank both Izzi Swinscoe and Maria Martinez Castillero for the frequent opportunities to catch up, ask questions, and run through new ideas. I need to impart my sincerest gratitude to Thomas Farrugia for his immense support throughout the project as well as after it; all of your support, guidance and mentorship at this early stage of my career has meant a great deal to me, and I'm very much looking forward to engaging in more projects together in the future.

This project aside, I've been incredibly lucky to have had several great experiences in scientific research across a number of different labs and fields, and feel that I owe a lot of my own successes to great mentorship. Firstly, I want to thank Magnus Johansson for hosting me at the Örebro School of Medical Sciences, which is where I first developed my love for

the lab. A huge thanks to both Wessam Melik and Sezin Günaltay-Schenk for their kind friendship and mentorship, for building my confidence in the lab and for helping me to understand that I was good enough to stay there. Thanks to David Nagel from Aston University for his sage knowledge and peerless humour relating to all things molecular biology. Thank you to Celine Alkemade, Gijsje Koenderink and the entire student team at AMOLF research institute for the fun times and camaraderie. Finally, I owe thanks to Enrique Asin Garcia and Vitor Martins dos Santos at Wageningen University for the excellent induction to the world of CRISPR and synthetic biotechnology, and for the helpful feedback and assistance leading towards in my first two publications.

Outside of the lab, I wouldn't be able to immerse myself in the work that I love without the unwavering support of my friends and family. I'm very aware that working with edible insects and genetically-modified bugs is fairly eccentric, and occasionally disgusting, so I truly appreciate that I am free to explore these and other curios knowing that I have a mountain of support behind me. To all the friends who I've bounced ideas off and ranted towards, know that I'll always appreciate it. I owe a huge thanks to my mum and dad for their steadfast belief in me over the years, enabling me to follow my heart and explore a number of options in research, and for never suggesting that I fall into a less-than fulfilling career. Finally, this whole thesis would have been impossible without the help and backing of my wonderful wife Arsheen. Thanks for being my loudest supporter, for sticking by me through rain or shine, and for all the kindness that you show to me and everyone around you.

Table of Contents

Abstract iv

List of Figures x

List of Tables xii

Abbreviations xiii

Chapter 1: Introduction 1

1.1 – Challenges to Food Security and the Role of Insects in Sustainable Agriculture 1

1.1.1 – An Outlook on Global Food Security 1

1.1.2 – Challenges for Future Food Sustainability 3

1.1.3 – Approaches for Sustainable Food Systems 5

1.1.4 – Reducing and Revalorizing Food Waste 7

1.1.5 – Insects: Nature's Essential Waste Recyclers 10

1.1.6 – The Role of Insects in Sustainable Food Systems 12

1.1.7 – Innovations in Insect Farming 14

1.2 – The Black Soldier Fly as a Model Farmed Insect 16

1.2.1 – Black Soldier Fly: A Life History 16

1.2.2 – The Emerging BSF Market 19

1.2.3 – BSFL as Efficient Waste Processors 20

1.2.4 – BSFL Animal Feeds for Nutrition and Welfare 23

1.2.5 – BSFL Frass as Organic Fertilizer 26

1.2.6 – High-Value Secondary BSFL Byproducts 27

1.2.7 – BSFL for Bioremediation 30

1.2.8 – BSFL Rearing as a Sustainable Process 31

1.2.9 – BSFL Mass Rearing and Potential for Automation 33

1.2.10 – Regulatory Landscape, Market Challenges, and Future Opportunities for the BSF Industry 35

1.3 – Genetic Strategies for Improving Black Soldier Fly 39

1.3.1 – Genetic Innovations in Enhancing Agriculture 39

1.3.2 – BSF Selection for Improved Traits 41

1.3.3 – Upgrading Agriculture with Gene Editing Technologies 44

1.3.4 – CRISPR Technologies for Precision Breeding 47

1.3.5 – Emerging Tools for Multiplexed Gene Editing 50

1.3.6 – Strategies for Insect Transgenesis and Transformation 52

1.3.7 – Advances in Genetic Engineering in BSF 56

1.3.8 – Future Directions for Engineering BSF 59

1.3.9 – Research Gaps and CRISPR-Aided Breeding Strategies 62

1.4 – Aims and Objectives 65

1.4.1 – Project Aims and Summary 65

1.4.2 – Research Questions and Objectives 66

Chapter 2: Materials and Methods 67
2.1 – Animals and Husbandry 67
2.1.1 – Animal Strains 67
2.1.2 – Black Soldier Fly Diets 67
2.1.3 – Black Soldier Fly Breeding and Maintenance 67
2.1.4 – Animal Microscopy and Imaging 69
2.2 – Black Soldier Fly Microinjection and Transgenesis 70
2.2.1 – Microinjection Slide Preparation 70
2.2.2 - Microinjection Mix Preparation 70
2.2.3 - Microinjection Needle Preparation 71
2.2.4 – Embryo Collection for Microinjection 72
2.2.5 – Embryo Alignment for Microinjection 74
2.2.6 – Embryonic Microinjection Procedure 74
2.2.7 – Post-Injection Rearing of G0 Larvae 76
2.2.8 – G0 Cross for Transgenic Line Establishment 76
2.3 – Molecular Biology 77
2.3.1 - Bacterial Transformation and Culture 77
2.3.2 – Genomic DNA Extraction 78
2.3.3 – PCR 79
2.3.4 – Agarose Gel Electrophoresis 80
2.3.5 – Gel Extraction and PCR Cleanup 81
2.3.6 – Nucleic Acid Quantification and Sequencing 81
2.3.7 – Plasmid DNA Assembly 82
2.3.8 – Construct Cloning Strategy 83
2.3.9 – Guide RNA Synthesis 85
2.3.10 – Inverse and Splinkerette PCR 85
2.3.11 – Tissue RNA Extraction 86
2.3.12 – RNA Sequencing 87
2.3.13 – cDNA Synthesis and RT-PCR 88
2.3.14 – Immunoblotting 89
2.4 – Insect Cell Culture 90
2.4.1 – S2 Cell Maintenance 90
2.4.2 – S2 Cell Transfection 91
2.4.3 – Fluorescence Activated Cell Sorting 92
2.5 – Software and Bioinformatics 93
2.6 – Statistics 94
Chapter 3: Developing an Endogenous CRISPR Toolkit for Gene Editing and Transgenesis in Black Soldier Fly 95
3.1 – Introduction 95
3.2 – Results and Discussion 99
3.2.1 – Selecting Pigmentation Target Genes for Mutagenesis 99
3.2.2 – Identification of Larval Growth Regulatory Genes for Multiplexed Editing 105
3.2.3 – Selection of Pol III RNA Promoters for Guide RNA Expression 109

3.2.4 – Guide RNA Expression Construct Assembly 113
3.2.5 – Identification of Germline-Specific Regulatory Elements for CRISPR Nuclease Expression 120
3.2.6 – CRISPR Nuclease Construct Assembly 124
3.2.7 – Preliminary Assessment of CRISPR Toolkit Activity in Insect Cell Culture 126
3.3 – Conclusion 133
Chapter 4: Establishing an Efficient Platform for Genetic Transduction and Transgenesis in Black Soldier Fly 134
4.1 – Introduction 134
4.2 – Results and Discussion 137
4.2.1 – Establishing a Method for BSF Maintenance 137
4.2.2 – Optimising Breeding Cage Lighting for Viable Egg Production 142
4.2.3 – Expanding the BSF Insectary Rearing Space 145
4.2.4 – Developing a Preliminary Protocol for BSF Embryo Microinjection 148
4.2.5 – Establishing an Embryo Collection Window for Microinjection 150
4.2.6 – Early Embryo Morphology and Developmental Staging for Microinjection 151
4.2.7 – Optimising Embryo Alignment and Microinjection Site 153
4.2.8 – Optimising Post-Injection Neonate Survival and Rearing 156
4.2.9 – Refining Microinjection Needle Parameters 156
4.2.10 – An Improved Embryonic Microinjection and Transgenesis Protocol in Black Soldier Fly 160
4.2.11 – Characterization of Promoters for Constitutive Transgenic Marker Expression 166
4.2.12 – Production of a BSF *White* Knock-Out Line by Cas9-RNP Microinjection 172
4.3 – Conclusion 175
Chapter 5: Evaluating a Transgenic Cross Strategy for Endogenous Gene Editing in Black Soldier Fly 176
5.1 – Introduction 176
5.2 – Results and Discussion 179
5.2.1 – G1 to G3: Establishing Stable Transgenic Lines 179
5.2.2 – Analysis of Transgene Inheritence Bias 183
5.2.3 – *piggyBac* Integration Site Mapping and Assessing the Impact of Transgene Position on Fitness 186
5.2.4 – Cas9 x gRNA-*white* CRISPR Cross Resulted in No Detectable Mutagenesis 193
5.2.5 – MAD7 Multiplexing: CRISPR Mutagenic Crosses and Amplicon Sequence Analysis 196
5.2.6 – Confirming Nuclease Transcription and Translation in BSF Tissues 199
5.2.7 - Enabling Endogenous CRISPR Activity in Cas9-Expressing Embryos via gRNA Plasmid Microinjection 202
5.3 – Conclusion 210
Chapter 6: Discussion and Future Directions 211
Conclusion 218
Bibliography 220
Appendix 245

List of Figures

Figure 1: The Black Soldier Fly Life Cycle. **17**

Figure 2: The CRISPR-Cas9 mechanism for indel formation via NHEJ-based DNA repair. **49**

Figure 3: Cas9 and MAD7 nuclease profiles. **51**

Figure 4: The embryonic microinjection technique for insect genetic transduction. **54**

Figure 5: A transgenic breeding approach for BSF to enable efficient gene editing activity in the transheterozygote offspring of CRISPR-encoding parental lines. **64**

Figure 6: Multiple protein alignment for the identification of the H. illucens *white* gene. **99**

Figure 7: Gene structure analysis and target selection in H. illucens *white*. **100**

Figure 8: Confirmation of Cas9 gRNA-w activity via *in vitro* Cas9 cleavage assay. **102**

Figure 9: Alternative pigmentation gene targets for mutagenesis in BSF. **103**

Figure 10: Candidate growth regulatory gene (GRG) targets for increased larval mass or reduced larval development time. **106**

Figure 11: Multiple nucleotide alignment of Dm_U6:3, the two BSF-endogenous U6 promoters identified in this study, and the three BSF U6 promoters identified in Kou *et al.* (2022). **112**

Figure 12: Entry gRNA expression constructs and type IIS cloning. **115**

Figure 13: Entry crRNA expression constructs and type IIS cloning. **116**

Figure 14: The *D. melanogaster w* gene structure, Cas9 and MAD7 target sites and CRISPR RNA expression constructs produced for S2 cell transfection. **118**

Figure 15: Construct schematics for the CRISPR RNA expression plasmids used in S2 cell culture and BSF transgenesis experiments. **119**

Figure 16: Germline tissue enrichment and differential gene expression of endogenous germline regulatory candidate genes in BSF. **121**

Figure 17: The protein, mRNA and gene structure of the *Nanos* and *Vasa* genes and regulatory regions identified for the cloning of the germline restricted nuclease expression constructs. **123**

Figure 18: Construct schematic for the CRISPR nuclease constructs used in S2 cell culture and BSF transgenesis experiments. **125**

Figure 19: CRISPR toolkit S2 transfection and analysis workflow. **127**

Figure 20: Cas9-transfected S2 amplicon analysis via T7-endonuclease I assay and trace deconvolution. **128**

Figure 21: Amplicon sequencing analysis of *Act5C*-Cas9 and U6:3 gRNA-*w* transfected S2 cells. **130**

Figure 22: MAD7-transfected S2 amplicon analysis via T7-endonuclease I assay, trace deconvolution and blunt-clone colony PCR. **132**

Figure 23: Black soldier fly mating, egg laying, egg collection and neonate rearing. **139**

Figure 24: Larval, pupal and adult stages in BSF. **141**

Figure 25: Morphological differences between fertilised and non-fertilised BSF egg clutches. **143**

Figure 26: Egg clutch production and viability was greatly enhanced when employing UVA-component LED lighting as opposed to using visible-spectrum LEDs alone. **145**

Figure 27: A self-contained BSF rearing space with environmental feedback and controls. **147**

Figure 28: Early morphological signatures of embryo development in BSF. **152**

Figure 29: Embryo alignment methods used for microinjection. **155**

Figure 30: Agar-based feed vials allowed for the efficient self-transfer of hatching neonates and safer rearing of larvae at low densities. **157**

Figure 31: Needle parameters for embryo microinjection. **159**

Figure 32: Protocol optimisations and embryo features identified in BSF microinjection setup. **162**

Figure 33: Transient and stable expression of fluorescent marker transgene at the embryo and larval stages. **163**

Figure 34: The *Hr5IE1*-GFP promoter-marker combination permits clear expression of transgenic marker in multiple tissues throughout all BSF development stages. **168**

Figure 35: The *D. melanogaster PUb* promoter was effective for flourescent marker expression at both larval and adult stages. **169**

Figure 36: The synthetic *Opie2* promoter was effective for marker production at the larval, pupal and adult stages. **171**

Figure 37: BSF *w* mutants have a clear eye pigmentation phenotype at the larval and adult developmental stages. **173**

Figure 38: Transgene inheritance bias is only apparent in the *Nanos*-Cas9 transgenic line. **185**

Figure 39: The sites of genomic integration of four transgenic *piggyBac* constructs determined by inverse PCR. **189**

Figure 40: No *white* target site mutation was detected in the F1 transheterozygote progeny of the *Nanos*-Cas9 x U6:3 gRNA *w* cross. **195**

Figure 41: No *white* target site mutation was detected in the F1 transheterozygote progeny of the *Vasa*-MAD7 x U6:3 crRNA *w* x2 cross. **197**

Figure 42: No *Maf1* target site mutations were detected in the F1 transheterozygote progeny of the *Vasa*-MAD7 x U6:3 crRNA GRG x6 cross. **198**

Figure 43: RT PCR and Western blot of WT and Cas9-transgenic line ovarian tissues, carcass and embryos. **200**

Figure 44: Injection data and amplicon sequencing analysis of *Vasa*-Cas9 embryo microinjection with gRNA-*w* plasmid. **205**

Figure 45: Allele frequency and alignment following amplicon sequencing of non-injected and U6:3 gRNA-*w* injected *Vasa*-Cas9 tissues. **207**

Figure 46: Allele frequency and alignment following amplicon sequencing of U6:1 gRNA-*w* and U6:2 gRNA-*w* injected *Vasa*-Cas9 tissues. **209**

List of Tables

Table 1: Pigmentation gene candidates identified for mutagenesis during this study. **104**

Table 2: Larval growth regulatory gene (GRG) candidates identified for mutagenesis during this study. **110**

Table 3: Chronological embryo microinjection experimental survival data before process optimisation. **164**

Table 4: Chronological embryo microinjection survival and transgenesis data for established transgenic and gene-edited lines. **165**

Abbreviations

AEL	-	After egg laying (*time elapsed*)
AMP	-	Antimicrobial peptide
BSF	-	Black soldier fly, *Hermetia illucens*
BSFL	-	Black soldier fly larvae
CDS	-	Coding sequence
CRISPR	-	Clustered regularly interspaced short palindromic repeats
crRNA	-	CRISPR RNA (unprocessed transcript, gRNA precursor)
DSB	-	Double-strand break
FAO	-	Food and Agriculture Organisation
FCR	-	Feed conversion ratio
FL	-	Food loss
FW	-	Food waste
GE	-	Gene editing
GHG	-	Greenhouse gas
GMO	-	Genetically modified organism
GOI	-	Gene of interest
GRG	-	Growth regulatory genes
gRNA	-	guide RNA
GWAS	-	Genome-wide association studies
GWP	-	Global warming potential
HDR	-	Homology-directed repair
Indel	-	Insertion-deletion
IVT	-	*in-vitro* transcribed
ITR	-	Inverted terminal repeat
KO	-	Knock-out
NGS	-	Next generation sequencing
NHEJ	-	Non-homologous end joining
PAM	-	Protospacer adjacent motif
PAH	-	Polycyclic aromatic hydrocarbons
pBac	-	*piggyBac* transposable element
QTL	-	Quantitative trait loci
ROS	-	Reactive oxygen species
RNP	-	Ribonucleoprotein
SDG	-	UN sustainable development goal
shRNA	-	Short-hairpin RNA
SNP	-	Single nucleotide polymorphism
UTR	-	Untranslated region

Chapter 1: **Introduction**

1.1 – Challenges to Food Security and the Role of Insects in Sustainable Agriculture

1.1.1 – An Outlook on Global Food Security

One of the most pressing challenges faced by society in the following century will involve building and maintaining global food security. While hunger refers to an acute and immediate lack of food, food insecurity involves the broader issue of a regular lack of access to safe and sufficient nutrition necessary for healthy development[1]. Food insecurity is pervasive and chronic in many areas, and can perpetuate states of long-term poverty, conflict, inequality and susceptibility to disease[2]. Building robust and sustainable food security in vulnerable populations will require international cooperation to tackle the economic, political, logistical and humanitarian challenges involved with bolstering global food production and distribution systems[3].

The second of the 17 sustainable development goals (SDG-2) set by out the UN in 2015 is to “End hunger, achieve food security and improved nutrition and promote sustainable agriculture” by 2030[4]. Despite a steady decrease in undernourishment between 2000 and 2014, recent reports by the UN Food and Agriculture Organisation (FAO) indicate that progress towards food security has stagnated, with food insecurity seeing an increase in many parts of the world between 2019 and 2022; factors such as increased conflict, the COVID pandemic, the locust crisis, and economic and climate shocks have all contributed to recent food scarcity and high prices[1].

Moderate food insecurity refers to irregular or unpredictable access to insufficient or low-nutrition food, while severe insecurity involves no access to food and going a day or longer without eating, posing serious risks to welfare[3]. By UN estimates, 2.33 billion people, around 28.9% of the global population, faced moderate to severe food insecurity in 2023, an increase from 25.9% in 2019 and 22.4% in 2014[1,5]. Geographically, populations in Africa faced a disproportionately high food burden, with 58.0% of people experiencing moderate to severe food insecurity, twice the global average[6]. Around one quarter of the populations

of Asia (24.8%), Latin America and the Caribbean (28.2%) and Oceania (26.8%) were similarly food insecure in 2023[1]. In general, rural communities were more likely to be affected than those in urban areas, relating to lower incomes, limited availability of key resources, and poor access to infrastructure and government support programmes[7]. Around 864 million people, 10.7% of the global population, experienced severe food insecurity in 2023[1].

Historically, a focus on reducing calorie deficit has not placed sufficient emphasis on the importance of dietary protein, fat, fibre and micronutrients for balanced nutrition, which is difficult to achieve for those with limited food choices[8,9]. This is especially salient for rural communities of smallholders and subsistence farmers, who represent a large proportion of the chronically food insecure population throughout much of Asia and sub-Saharan Africa[10]. In addition to energy from calories, having access to sufficient dietary protein, vitamins and minerals is crucial for early development and immune function[2].

Diverse nutrition in childhood is essential, with malnourishment during pregnancy and in the first 5 years of life consistently leading to low birth weights, child wasting, stunted growth, impaired cognitive development and a compromised immune system in children, all of which increase susceptibility to diseases like anaemia, rickets and malaria infection[3]. Although around half of all child deaths in 2023 were associated with malnutrition, less than 2% of deaths were attributed directly to starvation, with the vast majority suffering from diseases stemming from poor nutrition[11]. Despite subsistence farmers often having access to an energy surplus through a limited choice of cereal crops, the so-called “hidden hunger” linked with protein, fat and micronutrient deficiencies is implicated in a significant amount of child mortality throughout food insecure populations[12].

In 2024, the FAO recognised that the world is not on track to achieve any of the seven global nutrition targets associated with SDG-2 by 2030[13]. At least 600 million people are projected to be chronically undernourished in 2030, with food production systems at a severe risk of further destabilisation from conflict, climate shocks and increasing cost-of-living. Global trend assessments show that the prevalence of global food insecurity and undernourishment has recently deteriorated further away from the target goal, while the prevalence of stunting and wasting in children under 5 is decreasing, but at a slow rate[14]. In 2022, a joint statement made by the FAO, IMF, World bank and World Trade Organisation acknowledged a state of world food crisis, which as of 2024 is still ongoing[15,16].

Despite these current challenges, the risk to future food security is positioned to become much greater in the years to come. The global population is forecasted to increase to 8.5 billion by 2030, and 9.7 billion by 2050, necessitating an increase in food production capabilities to support this[17]. Although estimates vary, a conservative projection suggests that food output will require an expected increase of 56-62% by 2050 when compared to 2010 levels of production[18,19]. Despite substantial increases in crop and livestock yields in the previous century, resulting from improved irrigation, development of effective crop fertilisers and intensive farming practices, the rising gains in crop performance are now beginning to plateau; with the current productivity trend, we will be unable to keep pace with the required agricultural output to sustain the population level expected by 2050[20].

1.1.2 – Challenges for Future Food Sustainability

Despite the clear need for significant expansion to our food production capabilities, our agricultural systems as they currently exist will likely not be sufficient to close the food gap. Between now and 2050, the ability to produce food is likely to be weakened by the dual threats of climate variability and the depletion of resources such as freshwater and arable land, both of which are being exacerbated by the high greenhouse gas (GHG) emissions associated with intensive farming[3,21,22].

One challenge to meeting food production needs is the arable land gap. A combination of climate change and human activity means much of the land made available for agriculture is now unsuitable for cultivation, through drought and desertification, degraded and depleted soils, and high susceptibility to agricultural pests[23]. Because of unsustainable farming practices, such as crop monoculture and overdependence on synthetic fertilisers, food production is estimated to be responsible for 32% of soil acidification on land and 78% of eutrophication around the world, each having a significant impact on local ecosystems and land productivity[24]. A 2022 UN land report identified that up to 40% of all land suitable for agriculture could now be classified as degraded, chiefly due to soil erosion linked with intensive agricultural practices and deforestation[25].

Agriculture alone is responsible for 80% of deforestation globally, with cattle, palm oil and soybean cultivation being the main drivers for further land clearance[26]. Despite increasing food productivity in the immediate term, unsustainable practices such as this ultimately undermine food security, both directly by disrupting livelihoods and food production chains

in local communities, and indirectly by reducing regional and overall ecosystem biodiversity[27,28]. Agriculture-associated land clearance and pollution have been linked to a 70% projected loss in terrestrial biodiversity, and deforestation negates the climate mitigation potential of forest lands that would otherwise act as a net carbon sink[29]. Under the current trajectory, soil degradation is predicted to affect 90% of land globally by 2050[24,30]. Productive land, therefore, represents a critical resource to be protected, managed and replenished if soil-based food production is to continue at pace.

Another challenge to future food security is the substantial water footprint associated with current production methods. Agriculture accounts for around 70% of the world's freshwater use, primarily used for irrigation of dry soils, and the sector relies heavily on unsustainable levels of groundwater extraction[31]. In high demand, limited water supplies are vital to all processes in the agricultural, industrial and domestic sectors, with efficient use and allocation of water being crucial for avoiding drought and crop failures in food insecure areas.

In addition to excessive land and water use, food production is also estimated to be responsible for around 29% of total GHG emissions[32]. While this includes emissions from agricultural machinery and fertiliser production, and takes the effects of deforestation on reduced carbon capture into account, the majority of these GHG emissions result directly from food cultivation itself. The major agricultural GHGs that contribute to climate change are carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O)[33].

Although only 10% of anthropogenic CO_2 emissions are expected to originate from food production, the sector is collectively responsible for around 68% of CH_4 and 60% of N_2O annual emissions[34–36]. CH_4, mainly produced through livestock rearing and rice cultivation, and N_2O, a byproduct from fertilised crop soils, have around 21 and 310 times the global warming potential (GWP) of CO_2 respectively, expressed as CO_2 equivalent (CO_2-eq), and as such have a much higher climate impact[37].

Looking ahead, changing patterns in food consumption across the world is also cause for concern. Despite accounting for only 18% of the global calorie supply, and 25% of the protein supply, livestock cultivation occupies around 77% of all agricultural land[38]. The majority of farmed animals are reared on pastureland, grazing in permanent grasslands and marginal areas with poor soil and unfavourable climatic conditions, thus making productive use of land that would otherwise be unsuitable for cultivation[39]. However, up to 40% of arable land

is currently dedicated to growing crops, including cereal grains and oilseed cake, exclusively for use in animal feeds, ultimately resulting in a significant loss of agricultural resources compared to using the same land to produce crops for direct human consumption[40]. Compared to cereal grains, beef is associated with a 100 times higher land footprint, 20 times higher water usage and 20 times more GHG emissions per kilogram of food produced[41–43].

Nevertheless, meat still represents the most accessible source of protein, minerals and micronutrients in many low and middle-income countries (LMICs), with these otherwise being difficult to obtain from grown or foraged sources by those with a more limited food choice[44]. The consumption of meat has been heavily implicated in our evolutionary success as a species, and was found to be closely associated with improved perinatal health, cognitive development and immune response in children from LMICs[45].

Globally, rising incomes and increasing urbanisation means that meat consumption is expected to increase by up to 75% by 2050, and as such can be expected to have a major impact on land and water use and GHG emmissions[30]. While collective and individual efforts need to be made in moving towards meat and livestock-free dietary nutrition wherever possible, we have to be realistic about how this can be achieved worldwide in a limited timescale, and should therefore explore alternative strategies for sustainable production and climate mitigation to offset this expected increase in meat consumption[46].

1.1.3 – Approaches for Sustainable Food Systems

Enhancing global food security by increasing agricultural output, while simultaneously preserving environmental resources and taking more active measures to mitigate climate change seem to be fundamentally at odds with one another. A long-standing focus in the research community on increasing crop and livestock yields to plug the calorie and protein gaps unfortunately ignores several other crucial factors in ensuring food security, such as the importance of balanced nutrition beyond subsistence calories[10]. We should also recognise that large parts of our agricultural production system will need to be significantly adapted to meet the coming challenges, and redesigned with long-term sustainability in mind[2].

The effects of climate change are not now distant theory, but a lived reality of many in poverty stricken and in food insecure areas, with the potential for climate events to cause cascading

effects on production and supply chains threatening food security in developed and developing countries alike. At the same time, we should not be oblivious to the negative contributions of human behaviour in the ongoing state of food insecurity, with conflict, poor governance and class, gender and social inequalities all being key drivers in malnourishment and entrenched poverty endemic in many disadvantaged populations[16,47].

For many stakeholders, protecting sustainable food security at a local level may even require consideration of a conscious decoupling from many aspects of the international system of food production and trade, and a focus on building self-sufficiency and resilience to destabilising world events like regional conflicts, climate shocks and pandemics that can stymie global trade[48]. Far from being an isolationist position, technology and policy could perhaps be best utilised in many food-insecure communities by developing more decentralised methods of food production, taking advantage of natively-produced resources and a close network of local supply chains. This model could effectively reduce dependence on uncertain or risky sources of supply from international partners, bolster and diversify local economies and grant disadvantaged populations higher agency and control over their own food production and security[16,49].

Moving forward, solutions will require the development and adoption of sustainable food production methods, with governments and the research community both needing to prioritise quick and effective implementations to begin to address this challenge[50]. Inevitably, this will include the development of technologies and management techniques to increase the efficiency of current agricultural production platforms, without a significant escalation in resource usage or expansion into new land[51]. A vital area of development will revolve around enhanced breeding strategies for crops and livestock, both for higher produce yields, and for increased resilience to both disease and abiotic stresses brought about by climate change. Any productivity gains, however, will need to be coupled with improving sustainability in agricultural practices, and ecosystem renewal projects to translate the benefits into lasting food security[52].

Many of the potential solutions to building sustainability into our agricultural system rely on extensive scientific, economic and political investment to realise over the long-term. However, there are several steps that can be taken immediately to help reduce the risks to food security going forward, including education strategies and a change in habits regarding how we consume and interact with our food systems[53]. The first step to this is to identify and

eliminate inefficiencies within our current methods of food production, with the UN proposing that the most impactful opportunity for this being the reduction of food waste.

1.1.4 – Reducing and Revalorizing Food Waste

The circular economy is a model intended to address sustainable methods of resource production, such as food, in a way that minimises the generation of waste and makes the most of the resources used in the production process[54]. This usually involves the revalorisation of less useful by-products from other processes into new uses, ideally making the most of the feedstocks and energy used to create it to minimising loss as unused waste. Considering this, one way to reduce the current burden on food production is to reduce the agricultural losses from uneaten food wastes (FW), which makes up a staggering loss of efficiency from all agricultural processes.

Approximately one third of produced food by weight is lost as FW, with 1.3 billion metric tons (Mt) of uneaten, but usable biomass being thrown away annually across the globe[55]. As well as squandering the limited water, land and feed resources used to generate it, the loss of food as FW represents a global economic loss of $1 trillion USD per annum, with the production, transport and later processing of FWs being responsible for 8-10% of global GHG emmisions[56–58]. Considering the processes used from food production through to supply chain, consumer delivery and waste management, FW generation and elimination is the cause of an incredible 50% of GHG emissions resulting from the whole food production system[59]. A clear target for reducing loss of resources, UN SDG 12.3 calls for halving consumer, retail and supply chain-driven FW worldwide by 2030, with a modest reduction of 25% by 2050 predicted to reduce the food, land and GHG gaps in food production by 12%, 27% and 15% respectively[19,58].

With many strategies for enhancing food security inaccessible to the average person, reducing resource loss from FW is a clear step for individuals to take for making our agriculture much more sustainable. One review estimates that the population of developed countries, around 1.4 billion people, discard 670 Mt of food each year, while developing countries, a population of 6.2 billion, discard 630 Mt[60]. In the UK, food loss is around 50% at the point of consumption, equal to the combined losses at the food production, processing and distribution stages, equating to 13.1 Mt of food and 27.0 Mt of CO_2 going to waste in 2021[50,61]. With 50% representing the average for FW lost at the point of consumption in all

developed economies, any future improvements to agricultural yields and food retainment through more efficient processing and transportation methods would be effectively negated if our current attitudes towards waste generation remain[60].

These numbers highlight that if those in developed economies, which generally enjoy higher rates of food security, could reduce the amount of food lost to FW, it would take considerable pressure off the global food production system and allow for redirection of agricultural resources into food production in places where they are better utilised. On the production side, achieving this could involve improved processing and packaging techniques to maximise the stability and lifespan of perishable foods, as well as retail and consumer practices such as more considered purchases to avoid food excess and spoilage, donation of uneaten food and using FW as animal feed.

While FW usually refers to unconsumed, but otherwise edible and nutritionally available portions of food, a significant fraction of agricultural waste comes from non-edible biomass and process biproducts, such fibre and lignin-rich plant matter that is nutritionally poor or metabolically inaccessible[62–66]. Termed food loss (FL), this encompasses all non-edible biomass lost during food production, processing and in the supply chain, although the definition can be sufficiently stretched to include all biomass generated during production of food that ultimately doesn't contribute to the final product[67].

With this, FL can include non-harvested biomass such as crop roots or branches, or inedible parts of livestock removed from carcass during meat preparation. Palm oil, for example, represents only around 10% of the oil palm tree, with the remaining 90% of the plant being treated as waste, around 295 Mt annually[68]. While this fraction of total food waste is inevitable and cannot be addressed through consumption behaviour, they nevertheless represent a net loss of energy and resources used to generate food, and so represent an inefficiency in the agricultural process.

Although FW and FL should be reduced as much as possible, a certain level of resource loss to waste will be unavoidable. Under the ideals of circular economy, the nutrients and calories in the non-edible fractions of FW and FL should be processed into a more bioavailable substrate and reintroduced as a feedstock into other production cycles, maximising the yield of usable end products from the initial resource and energy investment[69]. Developing varied industrial processes to efficiently re-valorise organic wastes

with minimal input of additional resources could supplement the growing bioeconomy and represent a sustainable solution to long-term food security[70,71].

Certain organic wastes are suitable feedstocks for production of fertiliser or biofuels, whilst less accessible wastes may be used in domestic compositing or anaerobic digestion[72]. Household FW revalorisation processes have also been developed for the production of bioethanol, hydrogen or biogas, hydrochar and even biobased plastics such as PLA and PVA[72]. Some FWs may even be suitable for use in livestock feed, following considerable FW processing and safety assessment[73]. However, the inconsistency in FW composition and accessible nutrient content means that the vast majority of wastes are unsuitable for revalorisation, and end up in landfill or incineration plants. Although a significant amount of energy and nutrients are stored in waste biomass, the separation and processing required to liberate these resources is not usually cost-efficient, if even possible.

To be a sustainable and circular process, waste revalorisation needs to serve two purposes: to reduce bulky, low-value wastes into more manageable quantities, and to concentrate nutrients from the waste into an accessible, high-value resource that can be easily utilised in other production streams[74]. Having essentially negative value, FW is a plentiful feedstock found wherever humans inhabit, with FW disposal posing a significant economic, logistical and environmental challenge in rural areas and urban municipalities alike[75]. As FW represents a large reservoir of inaccessible calories, techniques that allow for the recovery of sequestered useful resources that require minimal additional inputs are highly sought after. The revalorisation of wastes with low nutritional value needs to be equally suitable for a number of diverse, unprocessed and unsorted waste streams to make the small return on energy investment worthwhile as a process. This necessitates a remediation method that is rapid, low-cost and requires minimal technology and energy input to achieve at high scales, before FW can be efficiently exploited as a resource and the energy within it can be sustainably recovered.

1.1.5 – Insects: Nature's Essential Waste Recyclers

Insects, the largest class of animals from the phylum Arthropoda, are one of the most diverse groups of species on the planet. So far, around 1 million distinct species of insect have been described, making up more than half of all identified eukaryotic species[76]. Still, only a small fraction of all insect species on earth are thought to have been classified, with estimates suggesting that there may be up to 5.5 million distinct species globally[77]. Terrestrial arthropods have a collective biomass of around 1 billion metric tons, around as much as all humans (400 Mt) and reared livestock (600 Mt) combined[78].

Insects are ectothermic invertebrates with a chitin-based exoskeleton, with the anatomy of the mature insect stage being divided into three morphological regions: the head, with a pair of compound eyes and antenna, the thorax, having three pairs of legs and two pairs of wings, and a segmented abdomen containing the digestive and reproductive systems[79]. During development, holometabolous insects undergo a complete metamorphosis in 4 distinct life cycle stages[80]. Embryos are in the form of laid eggs, which is proceeded by the mobile larval stage after hatching.

As invertebrates, feeding larva pass through several exoskeletal moults, or instars, during their growth and development, acquiring enough nutrients to support further metamorphism[81]. Insects then enter the non-feeding, quiescent pupal stage, in which the external and internal structures of the organism undergo drastic remodelling. Pupa often develop a hard shell or casing to protect and encapsulate this metamorphic process during this dormant period.

Finally, the imago, or adult stage, of the insect emerges from the pupal casing, characterised by a reproductive activity and high mobility, such as flight ability[79]. Depending on the species, the size of the adult body does not change, with most insects feeding in only small amounts or not at all at. The adult stage is therefore focused entirely on migratory and reproductive functions, in which they first search for a mate, then a suitable location near a source of larva-accessible food to deposit large numbers of fertilised eggs.

Although human attitudes towards the presence of insects often involves fear, aversion or disgust, the vast majority of species have no negative effect on human activities. Less than 1% of insects can be considered detrimental to human health or productivity, including all biting, stinging and disease vector species, as well as all destructive agricultural, structural

and resource pests[76]. Conversely, the activity of most species has a substantially beneficial effect on the functioning of our ecosystem, with insect feeding cycles playing a crucial role in nutrient recycling[82].

More than 90% of green plant biomass is unconsumed while alive in terrestrial ecosystems, and so is passed on to decomposers[83]. Dead, decaying or inactive biomass is responsible for sequestering large amounts of energy and nutrients available throughout the biosphere, with the health of the global ecosystem relying on the rapid turnover of these resources to support new growth. By feeding on dead plants and organisms, insects represent a vital link for the return of carbon, nitrogen and stored energy to the soil, allowing these nutrients to be assimilated by other species and re-enter the food chain.

Many polyphagous insect species are opportunistic scavengers, feeding on dead animal matter, or detritivores, consuming decomposing animal and plant detritus[84]. While true decomposition is mainly performed by saprotrophic bacteria, fungi and protists, most organic waste matter enters the remineralisation cycle as highly structured biomass, both physically and metabolically inaccessible to these microbes. In the first stage of remineralisation, insects greatly accelerate the process of organic decomposition by physically fragmenting waste biomass into smaller accessible materials, increasing the surface area available for absorption by plant and microscopic decomposers.

The substantial diversity in insect species has allowed for a high degree of specialisation regarding feedstock niche, meaning insects are collectively able to feed on almost any type of organic waste imaginable[76]. Evolutionarily, insects have become remarkably efficient at liberating the useable energy and nutrients from low-value organic matter not otherwise suitable for animal consumption. Termites, as one example, play a significant role in the breakdown of tough lignocellulosic biomatter in dead wood, while numerous other insect species are able to colonise and thrive in animal manures and decaying carrion, despite the high cooccurrence of pathogenic microbes and volatile byproducts from bacterial fermentation and chemical decomposition[84,85].

A high burrowing activity in feeding insect larva helps pulverise and aerate waste biomass into a porous, nutrient-rich humus. When mixed into soils, this greatly improves soil fertility and structure, regulating water retention while allowing drainage, promoting root growth and supporting the activity of soil decomposers microfauna[84]. Detritivorous insects and microbial decomposers have developed a synergistic relationship in supporting each other’s growth,

and both in turn are thought to be crucial for the rapid turnover and redistribution of sequestered energy and nutrients in plant, fungal and animal ecosystems.

Recent investigations into ecological decline has highlighted a worrying decrease in global insect populations, resulting from increased land clearance, pesticide use and pollution[86]. As well as making up the majority of pollinating species and being important to many animal diets, the central role of insects in organic decomposition and nutrient recycling should not be ignored, with ecosystem turnover and sustainability being severely at risk should insect numbers continue to dwindle[87]. At the same time, the incredible organic processivity of insect larva is being considered for applications in the revalorisation of human wastes.

1.1.6 – The Role of Insects in Sustainable Food Systems

The use of insects in agriculture and industry is not a novel phenomenon. Apiculture, the farming of honey from bees, and sericulture, the production of silk fibres from silkworms, have both been practiced since before recorded history[88]. Both species are considered domesticated, showing clear evolutionary signs of selective breeding for higher production yields and more efficient rearing practices[89,90]. In addition, carmine red pigments and shellac resins have been produced from cochineal and lac insects for several centuries, with both seeing periods of high international trade and demand[91,92]. More recently, research communities have begun exploring insect-derived peptides with unique functional properties for applications as engineered biomaterials. Insect cuticle-derived resilin is an elastomeric protein of interest in flexible tissue engineering, and fibrous spider silks are being developed for various uses in the biomedical, textile and construction industries[93,94].

Originally considered a novelty, the apparent benefits and utility of insect-based bioconversion for low cost, efficient waste revalorisation are now becoming more clear[95]. As with ecosystem wastes, the juvenile larval stage of many insects are able to efficiently break down and assimilate tough and low-nutrient FWs from various municipal and agricultural sources[96]. The larvae of many species are hyperphagic and have a close to exponential growth rate in optimal feed conditions, the purpose of this stage being to quickly accumulate sufficient size and energy stores to fuel metamorphosis into adult flies[97,98].

As an efficient bioconversion system, insects can rapidly convert a diverse range of organic substrates into larval biomass with a high protein, lipid and mineral content, with the unincorporated processed substrate left behind having a potential application as organic

plant fertiliser[99,100]. The most valuable product from insect-based bioconversion is the energy-dense larva, with their composition making them highly suitable as feedstock for numerous products including fuels, materials, animal feeds and even human foods[101].

The challenge in effective FW revalorisation is not just in waste reduction, but efficient bioconversion and concentration of stored energy into a high-value resource that is suitable for other uses, with nutritional status being a priority for improving our agricultural productivity[75]. For a reliable and scalable process, the incorporation of insects as a waste revalorisation strategy would benefit from the selection of a limited number of species more optimal for cultivation. The ideal farmed insect, or “mini-livestock”, should therefore meet three criteria; feed efficiently on a number of common FWs, yield a valuable or nutritious larval product, and be amenable to cheap mass-rearing strategies[102].

Several species that may be suitable for mass rearing have been identified by the nascent insect farming and food sector. Entomophagy, or the use of insects as food, is deeply routed in human evolutionary and cultural history[103]. As many agricultural societies began to move away from foraged sources of food, relying instead on more dependable production of crops and livestock, attitudes towards the use of insects in food have fallen out of fashion in many parts of the developed world[104]. However, entomophagy still plays a substantial role in the diets of several communities and cultural groups in LMICs, with edible insects considered a valuable source of nutrition and often even a delicacy[105].

Around 2000 species of edible insects have been recorded in modern use, with species of beetle, caterpillar, ant and cricket making up the majority of products in a protein market that serves 2 billion people throughout South America, Africa, Asia and Oceania[103]. By weight, many of these have a higher protein and mineral content than soybean, maize, beef, chicken or fish, and are considered safer that livestock products regarding zoonotic disease transmission[106]. Insects are more commonly eaten in the Tropics, with their large size, dense groupings and year-round presence making them an easy food to forage. Strategies for semi-cultivation in natural environments are commonly employed to increase insect collections, with the surplus being sold in local markets as a reasonable source of income[107]. In some communities, entomophagy can provide up to 10% of protein, fat and minerals in the diet, contributing directly towards better food security in these areas[104].

1.1.7 – Innovations in Insect Farming

However, the global market for foraged insect protein is inherently limited to certain seasons and geographies[108]. Additionally, an inability to effectively reconstitute a natural feeding or breeding environment makes the rearing of many species to high density in captivity difficult. For more reliable and productive insect farming, techniques for the mass-rearing of amenable insects are necessary[95,103]. Modern waste-fed insect farming facilities employ a limited number of species, with the most common including the black soldier fly, house cricket, palm weevil, waxworm, and the lesser mealworm[82]. The original considerations for the use of these species included factors such as food conversion rate, biomass accumulation over time, susceptibility to disease, genetic traceability, robustness on different waste feedstocks, cost of rearing and suitability towards farming automation[101].

With short lifecycles, insect larva reared to high densities using locally-derived FW address the dual problems of waste accumulation and protein shortage in food insecure communities. In its simplest form, a semi-cultivation process can be achieved using small bins of domestic and farm wastes by smallholders, revalorising spent crop waste and producing a steady source of energy dense-feed for livestock like swine, poultry and fish[109].

The barrier to scale-up in insect farming is low, with minimal feedstock or land investment needed to increase protein production. With insects reared in open containers of FW, a small dedicated insect farming facility only requires basic equipment like containers, plastic sheeting and nets, with electricity for lighting and climate control being mostly optional[103]. Agnostic to location, insects are well-suited to indoor or urban farming, allowing facilities to be localised close to areas with high FW generation to minimise transport costs. Facility space can be maximised through stacking trays of feeding larva, analogous to a vertical farming setup.

These factors all lower the cost-of-entry for enterprising insect farmers, with process scale up only requiring incremental increases in containers and space, and not depending on significant capital investment in heavy equipment. This dramatically reduce the risks associated with the venture, with an insect rearing setup being realistically achievable even with limited resources. If desired, domestic-scale farms can be easily translated into small local enterprises, empowering smallholders and subsistence farmers supporting local economies[110]. The produced larva can be sold at local markets for a quick and consistent

source of cash flow, external to the lengthy seasonal cycles associated with crop and livestock production[103,111].

With the practice providing clear socioeconomic benefits, insect farming is now seeing employment in several food-insecure LMICs by rural farm communities, as well as in urban areas prone to a high burden of FW[109]. Insect rearing requires little prior knowledge or training, and adoption is being encouraged through education drives and the dissemination of rearing manuals and equipment starter kits[112,113]. Being non-labour intensive as a practice, insect farming has a high social accessibility, and is understood to help reduce inequality by proving a path to livelihood for women, the elderly and people with disabilities[103,111]. On many fronts, FW valorisation through insect farming is a highly sustainable and circular process, and addresses many of the challenges faced by the global south regarding local food production and security[74,103,114].

Interest in insect farming is also growing in developed countries. The sector offers significant promise for generation of cheap and nutritious livestock feed, with suitability towards process scale-up being the major advantage from a cost-efficiency standpoint. As the major contributor towards FW, developed countries have a high availability of waste feedstock, with governments searching for more sustainable methods of waste management[115,116]. Following the entry of several pioneering companies, the nascent insect farming industry is rapidly becoming well established for serving the agricultural and pet feed markets[117].

In EU member states, the demand for insect protein is expected reach 1 Mt by 2030, with the global market value for insect protein predicted to increase from $271 million USD in 2020 to $1.5 billion USD by 2028[118]. At present, both research and commercial stakeholders are investigating methods for increasing farmed insect process efficiencies and larval yields, and have begun exploring the potential for emerging secondary markets, with valuable insect-derived materials possibly having applications across multiple sectors and industries[119]. While early insect farming communities initially worked with a range of different reared species, the vast majority of enterprises across all production scales have now moved towards adopting the black soldier fly.

1.2 – The Black Soldier Fly as a Model Farmed Insect

1.2.1 – Black Soldier Fly: A Life History

In the past 20 years, *Hermetia illucens* (*Linnaeus 1758*), or the black soldier fly (BSF), has attracted considerable academic and industrial interest as the ideal farmed insect species[120,119,121]. A large and distinctive common fly (Diptera) of the Stratiomiydae family, BSF is synanthropic, or a non-domesticated species that nevertheless historically associated with humans to benefit from their activities and environment[122]. Although the origin is disputed, BSF is considered likely indigenous to the tropical, subtropical and temperate zones of the Neotropics, a biogeographical region spanning from modern-day Argentina to the southern United States. Now a cosmopolitan species, BSF likely first spread following contact with the new world and increasing worldwide trade, spreading passively by ship[123]. A flurry of reported sightings throughout the early 20th century suggest a rapid colonisation of BSF to permissive climates across Europe, North America, Africa, and Asia, with a global distribution of genetically-diverse captive and wild populations seen today[122].

During this phase of expansion, BSF was considered a pest species, being erroneously associated with the disease myiasis, a parasitic maggot infestation in humans[124]. Population suppression and control strategies were developed to reduce BSF colonisation of domestic and agricultural wastes. The first positive indication for the utility of the insect came from the finding that the presence of black soldier fly larva (BSFL) in organic waste reduced the prevalence of other pest species, particularly the house fly *Musca domestica*[125]. BSFL were later recognised and routinely employed in the field of forensic entomology, with their predictable colonisation and growth rates being used to determine postmortem interval, or time of death, of human remains[124,126].

Although several studies showed dried BSFL as a suitable feed for chicken and swine, the need to isolate the species from the wild restricted research to a small group of suitably positioned entomologists. The first documentation of a BSF rearing method in 2002, later amended to allow for continuous rearing indoors independent of season or climate, was the paradigm shift allowing for BSF rearing to be achieved on an industrial scale. Within only the last 10 years, BSF mass rearing has become a thriving area for both academic research and development and innovative commercial ventures, with a community of BSF research

becoming established and considering the future directions of BSF technologies. However, the area is still young, and much basic developmental biology in BSF has yet to be studied.

Adult BSF are large flies between 15 – 20 mm in length, with their heavily tanned thoracic and abdominal cuticle giving them their eponymous colour, with occasional white colouration throughout the leg sections (**Figure 1A**)[127]. The flies have large compound eyes and antenna at the head, as well as a characteristic translucent dorsal and ventral cuticle at the first abdominal segment, allowing light to pass through the fly completely.

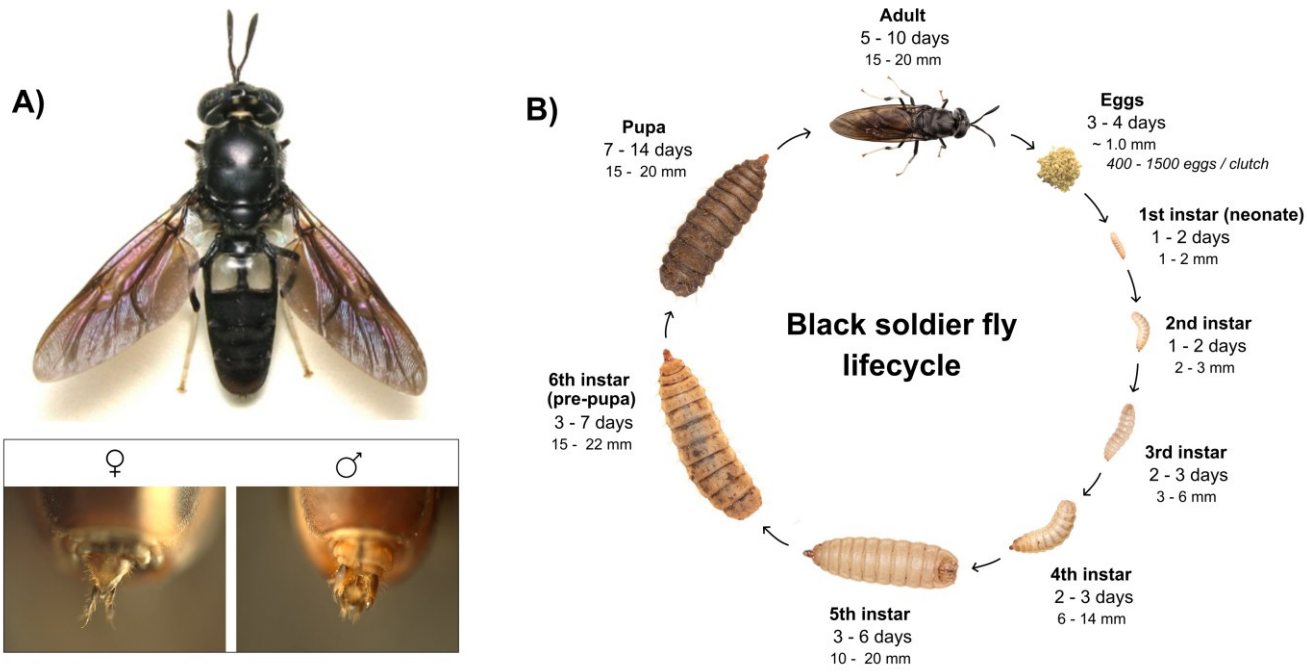


Figure 1: The Black Soldier Fly Life Cycle. A) A full size dorsal view of an adult BSF female fly (image adapted from De Smet et al., 2018[128]), and the morphology of the female and male external genitalia as the clearest indication of sexual dimorphism (imaged during the current study). The female retractile tubular "fork" oviduct extends during egg laying, and connects with the male "claw"-like aedeagus structure during mating[129]. **B)** The life cycle stages of BSF generally last between 5 – 6 weeks. The timeline and sizes provided are a rough estimate for BSF development in optimal feeding and environmental conditions, with both size and duration of each stage being highly plastic under suboptimal conditions. Size estimates and instar durations were averaged from a number of studies, with the relative sizes depicted here not being to scale and intended for illustration purposes only[130–132].

While the functional purpose of this “window” tissue is poorly understood, it is suggested that it plays a role in Batesian mimicry; the fly closely resembles the organ pipe mud dauber, an aggressive predatory wasp, in size and appearance, with the thin strip of black cuticle separating the two dorsal window sections giving the fly the appearance of narrow, wasp-like waist[133]. Alternative theories for the window tissue suggest that it permits the passage of sunlight for thermoregulation, is involved in fly communication or even a remnant of a since-extinct bioluminescence behaviour[127].

Despite appearing and sounding like wasps in flight, adult BSF are docile, poor fliers, unable to sting or bite, and generally human avoidant[134]. BSF are a non-pest, non-vector species, with the adult fly chiefly serving a reproductive purpose. Although flies can drink through their mouth parts, they tend not to feed, having assimilated enough nutrients during their long larval stage to sustain the rest of the lifecycle through considerable fat body energy stores[135,136]. BSF only exhibit sexual dimorphism at the adult stage, with male and female flies being distinguished by their genital structures (**Fig. 1A**).

Breeding in the species resembles a complex courtship, with males demonstrating highly active flight and crowding activity to attract females to a mating site, followed by chasing, in-air interception and capturing before copulation attempt is made on a flat surface[137]. Flies show a characteristic reverse coupling when mating successfully, with the flies facing opposite directions while their genitals are connected for a ~30-minute period. Mating activity is observed around 2 days post fly emergence from the pupal casing in captivity, with gravid females laying their eggs a further 2 days post mating (**Fig. 1B**).

BSF tend to lay their entire compliment of eggs in a single session as a large, compact clutch, containing anywhere from 400-1500 fertilised eggs, with the female dying soon after[138]. Females use their ovipositor to sequentially deposit the eggs in dark, dry cracks or crevices near a source of decaying organic matter, for protection against excess moisture or predation. Eggs are around 1 mm in length, and hatch around 3.5 - 7 days after being laid, depending on the ambient temperature[137].

After hatching, neonate BSF, or the first larval instar, crawl towards the decaying matter and begins feeding. The larvae, or BSFL, feed insatiably on a wide range of substrates, passing through 6 larval instars and exoskeletal moults and growing from 1 mm up to 27 mm long and 6 mm wide[120,99]. Larval growth typically lasts 16-36 days, including a 7-day wandering

prepupal stage in which feeding ceases and larva crawl away from the food source to find a dark, dry plate to pupate[120].

The prepupal to pupal transition is marked by a reduction of movement and a gradual hardening and tanning of the cuticle, progressing from soft and beige in colour to a hard, dark brown pupal casing. This is followed by an immobile pupal stage wherein the BSFL metamorphosise into adult flies, lasting from around 10 days up to months in cold conditions[136,138]. The life cycle lasts around 40 days, although the total development time is highly plastic and influenced by larval diet and environmental conditions[120].

1.2.2 – The Emerging BSF Market

BSF, and specifically BSFL, have been found to be especially suitable as a farmed insect for FW remediation and the production animal feed, organic fertiliser and valuable commodities[139]. This species is incredibly efficient at bioconversion, revalorising bulk organic matter into energy-dense and nutritionally-enriched larva ideal for livestock, fish and pet feed[140].

Research interest in this insect has thrived in the previous decade with the advent and adoption of reliable mass-rearing protocols, indicated by the proliferation of primary literature and reviews[119]. In the PubMed and Scopus databases, over 95% of articles mentioning “black soldier fly” or “*Hermetia illucens*” in the title or abstract were published after 2017, with 54% being published between 2022 and 2024[141,142]. A similar trend can be seen with BSF-related patent applications throughout the same timeframe[143].

Using a basic system of classification, the topics in the primary literature can be roughly divided into three groups of equal proportion: the suitability and yield of BSFL as meals or fertilizers for common crops and livestock; the performance and nutritional values of BSFL reared on various feedstocks; and a third, mixed group that includes studies on BSF basic and developmental biology, optimization of BSFL rearing or processing, applications for secondary larval byproducts, and various assessments of policy, regulation, safety, and welfare[141,142].

Commercial interest is growing too, with the number of niche startup enterprises and larger production-scale facilities increasing to serve the animal feed sector. According to market research, BSF products had a global market share of around $300 million USD in 2023, and

this is expected to increase to $4 billion USD by 2033, with a high compound annual growth rate of ~31%[144]. In terms of volume of BSFL protein, the market is expected to reach 8.23 Mt by 2033, with the majority of producers focusing on the aquafeed, animal feed, fertiliser and pet food markets. The main drivers for the predicted growth and success in this sector are expectations of an increased demand of meat and fish, rising soy and fish meal prices, and the adoption of government incentives enabling research innovation, supporting BSFL enterprises and promoting consumer acceptance.

In 2024, around 55% of the global market share for BSFL products resides within the European market, with EU countries hosting the highest number of large production plants[144]. However, it is harder to establish the contribution of small, non-industrialised BSFL operations currently being carried out across food insecure LMICs, or estimate the trajectory and adoption of BSFL farming in these areas going forward. While the bulk of insect farming currently caters to the animal feed sector, the emerging potential for high market diversification should not be ignored, with small enterprises looking to employ BSFL as a platform for fuel, pharmaceutical and cosmetic production.

The growing popularity of BSFL farming for sustainable agriculture and industry can be explained by several characteristics that make them highly desirable for the above-mentioned uses. This includes their remarkable feed efficiency, their suitability as food and feed, their utility for non-feed products and bioremediation, and their sustainability attributes when compared with other common protein sources. Looking at each of these in detail is useful for understanding the potential of the new BSFL industry going forward, and to highlight some of the current challenges faced by the sector.

1.2.3 – BSFL as Efficient Waste Processors

Firstly, BSFL have a remarkable waste processivity. At peak feeding activity, larva can consume up to 500 mg of waste per day[145]. When using stacked feeding trays, BSFL are able to process 31 kg of municipal food waste per m^2 per day, with a total waste reduction of 50-80% by weight depending on it's composition[146–148]. This high waste throughput can be maintained as a continuous process by adding fresh waste, water and young larva as the quality of the processed feedstock decreases and satiated prepupa migrate out of the system[149].

Being short lived and cold blooded, insects reared in a temperate environment have a lower metabolic burden than larger, warm-blooded livestock species that have to expend energy to maintain a constant body temperature[108]. BSF are therefore able to convert and assimilate significant amounts of the nutrient and energy content of feedstocks into larval biomass. The efficiency of this bioconversion process can be measured as the feed conversion ratio, or FCR, for different species and feedstocks[150]. For BSFL, FCR can be calculated as the weight of the feed consumed divided by the weight of larval biomass produced, with a low FCR indicating a high bioconversion efficiency. In practice, low FCRs translate to a higher yield of biomass for a given amount of feedstock, and brings corresponding economic and environmental benefits for more efficient bioconversion processes.

BSF has an incredible tolerance for several low-nutrition waste feedstocks, ranging from domestic food and agricultural wastes to industrial organic runoff, abattoir and fishery offal, animal manure and even sewage sludge and wastewater[151–155]. Across multiple feeding studies, the larva have been shown to efficiently reduce the weight and nutrient content of various waste formulations, with limited impacts to their own growth and survival[156].

A combination of BSF-endogenous and exogenous factors allow larva to thrive in these low-nutrient environments. The production and activity of digestive amylase, protease and lipase enzymes in BSFL guts is around twice as high as found in the similarly polyphagous larva of *M. domestica*[157]*.* The expression levels of these enzymes show a remarkable plasticity to suit the relative concentrations of their target substrates, allowing an adaptable digestion efficiency for each macronutrient depending on its abundance in the current feedstock[131].

Levels of binding and transport proteins for sugars, amino acid, lipid and micronutrients are similarly variable, facilitating high absorption of digestate into the gut for a balanced nutrition[131]. In larva fed with nutrient-poor diets, the microvilli of midgut columnar cells were found to be much longer than in those fed with high-quality feed, suggesting further adaptability in cell morphology to increase the gut surface area available for absorption. Additionally, BSFL maintain a high content of glycogen and saturated fats relative to other insects, meaning that energy can be efficiently sequestered and mobilised as needed in variable nutrition conditions[158].

The intestinal microbiome of BSFL also plays a significant role in their ability to utilize low-value feedstocks, as evidenced by their poor growth and survival on autoclaved feeds and

in sterilise conditions[159]. Although a highly active area of BSF research, there is limited consensus on whether the high digestibility of different feeds by BSF stems from the core microbiome species found consistently across feed conditions, or if this is more dependant on the variable fraction of intestinal microfauna that exists transiently and associates with different types of feed[160,161].

Regardless, several frequently identified bacteria genera are suspected to play a role in aiding BSFL metabolism in the gut. A core microbiome of *Actinomyces*, *Enterococcus, Dysgonomonas* and *Providencia* species are all likely to help degrade complex carbohydrates and plant fibres, including lignocellulose, xylene and hemicellulose, into more readily digestible substrates[162,163]. Additionally, *Lactobacillus* species present in the gut are expected to aid in detoxifying ingested xenobiotic compounds and stimulate the production of antimicrobial peptides, protecting BSF from pathogens[164]. Several probiotic strains, including *Bacillus subtilis*, *B. velezensis* and *Rhodococcus rhodochrous* have also been identified, prompting higher feed conversion rates and shorter larval development times when supplemented into BSF feedstocks[162].

In BSF, the achievable feed conversion ratio or FCR for different feedstocks can vary significantly, depending on resource factors like feed composition, quality and methods of preparation, and process factors such as rearing practices and environmental determinants such as temperature, humidity and feed aeration[135]. According to one study, calculated FCRs for BSFL can be as low as 1.25 – 1.40 on a dry weight basis with high protein, high fat diets[165,150]. In real terms, this means that 1 kg of live BSFL can be generated from as little as 1.25 kg of feed by wet weight, representing an incredible conversion of feed to biomass. For comparison, energy-dense concentrate feeds have been expected to lead to relative FCRs of 1.4 in salmon, 2.3 in chicken, 4.0 in pork and 8.8 for cereal beef[166].

With much of FCR being influenced by bioavailability of the feedstock, the advantages of BSFL in efficient waste reduction becomes apparent when considering the use of low-quality diets. In the above study, a low FCR of 1.25 was achieved using a 1:1 mixture of poultry litter and slaughter waste, a considerably low-value waste feedstock[165]. Larval performance in this feed was high, seeing high larval weight, short development time and a 100% survival rate post feeding. This corresponded to a bioconversion ratio of 31.84%, representing the fraction of feedstock assimilated in larval biomass on a dry weight basis, and a ~40% reduction in waste biomass.

In other studies, BSFL were able to assimilate 59% of the nitrogen, or protein, content of domestic FW, and 80% when fed on a basic diet of chicken feed[167]. In terms of energy metabolism, BSFL were found to retain 53 – 63 % of carbon assimilated through feeding as larval mass, with the rest lost to respiration[168,169]. This is more favourable than livestock like chicken, in which only 35 – 40% of assimilated energy is directed towards growth, the rest being lost through metabolism and maintenance[170].

Additionally, a larger proportion of BSFL biomass can be digested when compared to other livestock and animal feeds. When eaten as whole or dried larva, up to 80% of BSFL represents digestible food and feed, with chitin fibre being the majority of the inedible fraction, while only 55% of chicken and swine biomass, and 40% of cattle, is suitable for direct consumption[171–173]. Around 85% of soybean and 90% of oil palm plant biomass do not ultimately contribute to animal feed, and eventually contribute to crop waste[68,174]. In production terms, BSFL can be ready for harvest from as low as 20 or so days post feeding initiation, while wheat and cattle is around 240 days, and soy and chicken is around 120 days[173,174].

1.2.4 – BSFL Animal Feeds for Nutrition and Welfare

For waste revalorisation purposes, a high bioconversion efficiency is only beneficial if the recovered energy is directed towards a more useful and valuable product. Importantly, the waste-fed BSF larval biomass itself represents a nutritious source of nutrients that is well-suited for use in different animal feeds[139]. Typically, FW-raised BSFL have a relative dry weight content of 38-62% crude protein with high digestibility, and a consistent amino acid composition enriched in essential lysine and methionine when compared with cereal, soy and fishmeal-based animal feeds[136,175–178].

Although variable based on diet, larva have a generally high lipid content of 14-49%, and are rich in saturated fatty acids such palmitic, lauric and oleic acid[179]. BSF oils have been found to be highly suitable for animal nutrition, and have been recommended as a sustainable dietary alternative to fats extracted from palm-oil processing[180]. In production populations, larval diet can be adjusted to increase the content of the relative protein and lipid fractions. For example, feeds that include fish offal and seaweed allow BSFL to accumulate omega 3 and 6 fatty acids, despite not being able to produce these naturally, increasing their value as a source of essential dietary fat[181–183]. A high content of indigestible

chitin polysaccharide serves as a source of fibre that is well-tolerated in livestock[184]. BSFL also have a rich profile of essential micronutrients for balanced nutrition, with prepupa being enriched in calcium, phosphorous, iron, essential trace minerals and vitamins such as B2, B7 and B12[185].

In multiple studies, inclusion of BSFL products in livestock and fish feeds has been linked with general improvements in final body weight, meat quality, feed conversion ratio, and immune response and welfare[186,187]. Larva and their processed products have been assessed for their potential to replace anywhere between 10% and 100% of conventional soybean or fishmeal protein supplement in many commonly farmed species[145]. Generally, inclusion of between 25-50% of insect protein alongside conventional protein sources has been best associated with increased meat yield and quality in livestock, as well as enhancing production of secondary products such as eggs from chickens and milk from and dairy cows[188,189]. Alternatively, complete replacement of soy and fishmeal with BSFL meal has also been successfully reported, with no adverse effects found regarding yield, quality and welfare characteristics in broilers, pigs and Nile tilapia[188,190–192].

Post-feed BSFL processing steps such as defatting, drying and milling allows for further enrichment of the specific macronutrient fractions, with commercial insect meals comprising a dried, protein-rich powder with long shelf life and suitability for mixing into traditional animal fodders[193]. Alternatively, dehydrated, freeze-dried and even live BSFL can be fed to animals directly; although less economical than processed insect meal in terms of space and storage, using dried and live larva as feed may be more desirable when considering the welfare and psychological well-being of the animal[194]. Poultry show a preference and high palatability towards live BSFL, with insect foraging more closely approximating their natural diet and feeding behaviours[195]. Live insect feed has correspondingly been associated with improved activity, feed intake and stress response in farmed animals, with a clear reduction in frustrated, fearful and aggressive behaviours observed with BSFL-fed chicken, duck, swine and fish[195–198].

BSFL feeds are also associated with imparting immunological benefits when included in animal diets. Lacking an adaptive immune function, insects rely on an innate immune system largely comprised of antimicrobial peptides (AMPs) for protection against pathogens[199]. Short cationic peptides, AMPs inhibit bacterial, fungal and viral colonisation by disrupting

cell membrane integrity, with their lipid-targeting mechanism of action though to be an effective barrier to the evolution of AMP resistance mutations[200].

As a species, BSFL thrive in unhygienic areas and substrates with high microbial and pathogenic loads, which is thought to explain the high activity and diversity of their AMP gene complement[201]. Particularly enriched in cecropins, defensis and peptidoglycan recognition proteins, more than 50 AMP genes have been recorded in BSF, one of the highest throughout all insect species[199]. With promising antimicrobial activity, several BSF AMP genes have already been characterised for their inhibitory activity against common pathogenic bacteria in *in vitro* panels and when expressed endogenously in transgenic silkworm[202,203].

In addition to AMPs, BSFL meals have a high chitin and lauric acid content, with both compounds being linked to beneficial immune and anti-inflammatory responses when included in animal feeds[146,204]. Larva-fed chicken showed higher lymphocyte production and increased survivability against *salmonella* infection, and seabass had increased anti-oxidant activity and an enhanced resistance to pathogenic *Vibrio alginolyticus*[205,206]. As well as serving an antibiotic function, BSFL feeds can have a probiotic effect on livestock microbiome homeostasis and support gastrointestinal health. In finishing pigs, larval diets were linked to a balancing and enrichment of beneficial gut microbes, while simultaneously decreasing the abundance of opportunistic *Streptococcus* pathogens[207].

With their strong immune characteristics, waste-fed BSFL are considered a hygienic species when processed correctly, and are considered low-risk for the transmission of zoonotic pathogens when used as livestock feed[208]. BSFL were able to inhibit and reduce pathogenic *Staphylococcus aureus*, *Salmonella enterica* and *Escherichia coli* in pig manure and domestic FW[209,210]. Despite not being a reservoir for food-associated pathogens, FW-fed BSFL may still allow the passive or mechanical transmission of certain parasites present on larval surfaces or in the intestine, with post-processing washing and heating steps recommended to reduce disease risk[211].

1.2.5 – BSFL Frass as Organic Fertilizer

The major byproduct by weight, frass is a granular mixture of undigested feed, larval faeces and shed exoskeletons left behind following a BSFL feeding process. Despite this, dried frass has no foul odours, is hygienic and generally well-suited for use as an organic plant fertiliser and soil additive[212]. Depending on the feedstock, BSF frass has a relative nitrogen, phosphorous and potassium (NPK) content ideal for plant growth and comparable with commercial fertilisers[213]. Additionally, it contains high amounts of secondary macronutrients (calcium, sodium and magnesium), as well as the trace nutrients iron, copper, manganese and zinc for balanced soil nutrition[212]. Easy to store and transport, frass decomposes into soil more quickly than inorganic fertilisers and is readily incorporated in plant roots.

As well as nutrition, BSF frass contains several bioactive stimulatory compounds that increase soil quality and support plant growth, especially in low-nutrient soils found in drought conditions. Frass is enriched in humic and fulvic acid, which has been found to improve soil structure and water retention, precipitate micronutrients for easier root absorption, and reduce soil salinity and oxidative stress[212]. These humic compounds have a buffering activity, with frass being neutral to slightly alkaline in pH[214]. Taken together, frass supplementation can help balance soil pH and regulate soil acidity.

Due to the larval exuviae content, frass contains a high concentration of chitin, known to be a potent stimulator of plant immunity through activating defence-related signalling cascades, prompting the secretion of antimicrobials and reinforcement of cell wall structures[215,216]. BSF frass was found to reverse the effect of wilt diseases in tomato, lettuce and common bean plants by reducing levels of the pathogenic soil fungi *Fusarium oxysporum* and *Rhizoctonia solani*[217,218].

Chitin is also effective in the control insect pest species, with BSF frass fertiliser significantly impacting both egg and larval survival in wireworm and root-knot nematodes[219,220]. Despite this, chitin-rich frass promotes overall growth of beneficial soil microbes, such as nitrogen-fixing bacteria and rhizobacteria, with increased microbial biomass and activity found in frass-fertilised soils[221]. Frass itself can have a beneficial microbial composition, being highly enriched in *Bacillus*, *Lactobacillus*, *Actinobacteria* and *Pseudomonas* species associated with improved plant growth and soil conditioning[214].

BSF-produced frass allows for a quicker reduction of organic waste when compared to conventional composting methods, and releases less than a third of the GHG emissions associated with composting[222]. The growth and health of several frass-fertilised crop species was comparable to the gains seen with commercial fertiliser use, and was significantly higher than non-fertilised conditions, although yields can be variable depending on plant type and amount of frass used[218,223]. Cheap to produce in bulk, BSF frass could be a suitable replacement for inorganic fertiliser used with several crop species, ameliorating the high sustainability costs associated with fertiliser production and soil eutrophication[224]. Able to supplement and replace the composition of peat-rich soils by up to 20%, the suitability of frass in greenhouse growth mediums and soilless agriculture may also be useful to explore going forward[225].

1.2.6 – High-Value Secondary BSFL Byproducts

In addition to their suitability in animal feeds, BSFL are rich in secondary products and metabolites that have been indicated for a diverse range of applications throughout many areas of food, health, and industry[111]. The protein, lipid and chitin fractions of BSFL can be extracted post-feeding for a cheap source of enriched bioactive materials. Chitin, and it's derivative chitosan, are biodegradable, non-toxic fibres with a high biological compatibility[226]. While currently produced from the inedible fraction of crustaceans, such as crabs and shrimp, an increasing demand for these fibres as a future biomaterial could benefit from a sustainable, ethical source[227]. BSF, particularly at the adult stage, are rich in these materials, and several processing methods have been described for the extraction of chitin and chitosan at high yields for use in other industries[228].

With strong antimicrobial activity, BSF chitin could be used as a food preservative and packaging, with the ability to make an edible protective coating for fruits and vegetables[227]. When mixed into foods, it can act as an emulsifying, thickening and flavouring agent that inhibits the growth of bacteria and fungi[229]. At higher concentrations, chitin functions as a biodegradable plastic with good mechanical properties, allowing for moulding into drinking straws and cups with high heat resistance and transparancy[228,230]. BSF-derived chitin has already been indicated for use in the textile industry, where it could be used to functionalise cotton, silk, wool and polyester fibres for to enhance shrink resistance, moisture control and dye stability[231].

The research community is also assessing purified chitin for medical uses, with high potential for dietary supplementation in cholesterol reduction, antimicrobial wound dressings and for drug coating and encapsulation matrices[232]. The material is also being considered for use in biodegradable scaffolds and tissue engineering, with high suitability being demonstrated for biocompatible implants and skin, cartilage, bone and blood vessel reconstruction[233]. Finally, chitin and chitosan extracts have also proved useful as flocculants for the coagulation of materials, with potential uses ranging from beer clarification to waste water treatment[234]. Chitin extracts have already been successfully employed to remove oils, soluble organic pollutants and heavy metal ions in waste water, and improve the quality and safety of drinking water[235,236].

In terms of antimicrobial activity, BSF AMP extracts have proved to be an effective antimicrobial agent against biofilm formation and persistent infections that respond poorly to antibiotic treatments[164,202]. They have a broad targeting spectrum, and their *in vitro* effects have already been demonstrated with a number of common human pathogens including *Klebsiella pneumoniae* and multi-drug resistant *P. aeruginosa*[202]*.* The lipid fraction of BSFL also contains species with antimicrobial and anti-inflammatory activity, such as lauric acid and isoprenoids that have been shown to reduce inflammatory bowel disease when fed to mice[237]. AMPs and medium-chain fatty acids like lauric acid produced from BSF have been emphasized as promising alternative compounds to antibiotics for avoid pathogenic resistance in humans and livestock[238].

Additionally, BSF extracts contain bioactive compounds that could be explored for applications in pharmaceuticals and cosmetics. As a medicinal component, isolated BSF peptides have potential for use as an ACE inhibitor in reducing hypertension, as well as in regulating feeding behaviour as a therapy for type 2 diabetes[239,240]. The lauric acid content of BSF is compatible to those in coconut and palm oil extracts, with BSF preparations having been used as a ROS scavenger to limit oxidative stress in human tissues[241]. Larva also contain high concentrations of oleic, linoleic and palmitic acids, with their inclusion suggested for moisturizing, skin barrier and UVB protection functions in BSF-derived skin products[242,243]. These fatty acids, and other BSF-rich elements such as flavonoids, are already commonly employed in cosmetics as emollients, stabilisers and emulsifiers, and are understood to be beneficial for cracked skin repair and ailments such as atopic dermatitis, psoriasis and eczema[244].

BSF oil extracts, either complete or fractionated, have also been shown to be effective lubricants and surfactants for use in industrial processes, and are more sustainable than alternative petrochemical and vegetable oils[245,246]. Finally, BSF oils are of significant interest for the development of biofuels. While biofuels can be criticised as unsustainable, they are generally more efficient, safer and emit lower amounts of GHG than fossil fuels, and so represent an acceptable medium-term solution for reducing transport emissions where EV infrastructure is poorly supported[247]. Although more sustainable than petrochemical fuels, biofuel from plant sources including soya, palm and microalgae take up a significant amount of agricultural resources to produce, with production of the feedstock used in fuel extraction production responsible for a significant amount of emissions, land and energy use[248].

The lipid content of BSFL make them highly suitable for the extraction of biodiesel, with simple processing steps making the production process significantly more sustainable in terms of water and chemical usage compared to crop feedstocks[249]. One tested method for biofuel extraction in BSFL was found to require 46% less energy input when compared to leading green extraction methods used in microalgae, with most energy usage being associated in dewatering the larva post-feeding[250,251]. When produced from organic waste, as opposed to from crops grown specifically for biofuel production, BSFL-based fuel production is significantly closer to a closed loop sustainability process. With BSFL representing a high oil content and production requiring low amounts of space, it has been estimated that significantly more fuel (145,800 kg of biodiesel per hectare per year) could be produced in the same space and time frame as when using soy (562 kg), palm (4,747 kg) and microalgae (86,515 kg) feedstocks[248].

Similarly, biofuels production with BSFL could be decentralised away from geographically-limited areas of crop production, providing a significant benefit in reducing fuel transport costs and in increasing local supply[251]. A case study of local BSF biofuel production from FW in Mexico modelled that mass adoption of the technology could reduce landfill-related emissions by 86%, with BSFL-derived biofuel production generating 41% lower GHG emissions than when made from waste cooking oils, and 78% lower GHG than using fuels generated from soy and palm[252]. The use of BSFL to produce valuable commodities holds considerable promise, with higher adoption of BSFL in animal feed production and waste management potentially creating research and commercial space for exploring these markets.

1.2.7 – BSFL for Bioremediation

In addition to efficiently recycling beneficial nutrients from wastes, BSFL have a significant bioremediatory potential for many toxic compounds and pollutants. When compared with other Diptera, BSF is significantly enriched in genes from the cytochrome P450, Glutathione S-transferase and ABC transporter families, which are strongly associated with the detoxification of xenobiotic compounds[253]. In BSFL feeding studies, cytochrome P450 was implicated in the detoxification of food-borne mycotoxins produced from *Aspergillus* and *Fusarium* fungi, a common source of contamination that causes crop spoilage and significant disease in humans and livestock[254].

Feed dosing studies have demonstrated that BSFL are capable of breaking down several common pesticides and insecticides often retained in crop waste, with minimal bioaccumulation or impediment to survival[255]. BSF can similarly remediate several classes of antibiotics used in livestock feed and enriched in manures, such as tetracycline and ciprofloxacin, as well as a number of persistent pharmaceutical compounds commonly found in domestic wastes[256–259]. The species is even able to remediate petroleum-based plastic contaminants in feedstocks, with undetermined elements of the microbiota allowing for the partial degradation of polyethylene (PE) and polystyrene (PS) fibres[260]. While research on plastic metabolism and assimilation in BSF remain limited, the microbiomes of several insect species are increasingly being explored for their potential applicability in microplastic waste remediation. The yellow mealworm *Tenebrio molitor,* for example, has been shown to effectively degrade low, linear low and high density PE microplastics, with the larvae even being capable of breaking down and assimilating solid PS and polyurethane wastes as their sole feed source[261–263]. Whether the microbial consortia responsible for plastic degradation in other insect species can be effectively reconstituted in the more resilient, processive and substrate-tolerant BSF remains a promising avenue for future research[264].

Remarkably, BSFL are also able to tolerate feeding on several persistent organic pollutant (POP) species, adverse chemicals or carcinogens often present in industrial wastes. Larva were able to reduce the amount of 4 different polycyclic aromatic hydrocarbon (PAH) compounds in high-dose feeds by up to 80%, with PAH remediation by BSFL being significantly more rapid and efficient when compared with other common treatment methods[265]. Although unable to effectively degrade these compounds, dioxins,

polychlorinated biphenyls and mineral oil hydrocarbons were found to not bioaccumulate in BSFL when included in feeds, instead remaining in the processed frass[266,267].

Finally, the species can be used to bioremediate wastes containing heavy metals such as arsenic, mercury and zinc with low retention in larval biomass[268]. Cadmium, and lead and chromium to a lesser extent, shows significant bioaccumulation in BSFL fed under high doses, representing a point of concern regarding the safety of larval bioremediation with certain high-risk wastes[269]. However, high bioaccumulation of certain pollutants could still be exploited for effective waste treatment, still allowing for significant reduction and conditioning of mixed organic wastes followed by the removal and disposal of the pollutant-enriched larva, leaving behind process frass significantly depleted of these metals.

1.2.8 – BSFL Rearing as a Sustainable Process

When considering the use of BSFL in animal feeds, the main market competition comes from the dominant cheap feeds based on soy or fish meal. Around 70% of protein meal used in all agricultural feeds is soy meal, which is produced in bulk for a low price at the cost of causing unsustainable levels of land clearance and high shipping-related GHG emissions[270,271]. Although produced in far lower quantities than soy, up to 18% of all fishing landings are thought to go towards the production of fish meal for animal feeds, with 90% of that fraction being deemed otherwise safe for direct human consumption[272]. Although accounting for only 4% of the animal feed consumed today, aquaculture is by far the fastest growing market within the agricultural sector, with a 4.5 fold increase in fish farmed by weight seen since 2000[273]. Around 75% of all fish meal production now contributes directly to aquacultural feed, with the use of fish meal expected to increase in line with continued growth of global aquaculture.

Although not currently competitive with soy and fish meal on price and production level, BSFL protein meals are associated with significantly increased sustainability metrics when compared to these and other common protein sources. Regarding water use, the production of 1 kg of BSFL has a calculated footprint of ~1.4 m^3, or 1400 L of water, compared to 2.1 m^3 required to grow 1 kg of soy[42,274]. While soy can be mostly provisioned by rainfall, BSFL may not need to be supplied at all, with most FW diets containing water in sufficient amounts. Comparatively, chicken, swine and cattle require 4.3 m^3, 5.9 m^3 and 15.4 m^3 of water respectively[42].

Life cycle assessments have identified BSFL waste processing and protein production process as high sustainability in terms of GHG emissions and overall GWP. CH_4 production from anaerobic digestion in waste landfill accounts for 18% of all anthropogenic emissions, with N_2O emissions in wastes accounting for around 3%[275]. As a waste-treatment process, composting with BSFL was associated in one study with a 47 times lower GWP than if the same waste were conventionally composted in landfill, mainly through significant reduction in production of CH_4 (0.4 g per ton) and N_2O (8.6 g per ton)[222].

Still, 68% of the GWP identified from BSFL-treated waste was linked to direct emissions from uneaten residues that continued to compost following BSFL removal, with post treatment of spent feedstocks likely to mitigate this[222]. BSFL are thought to produce little to no CH_4 from intestinal bacteria, with aeration of the feedstock mediated by BSFL movement helping to reducing the amount of CH_4 produced through anaerobic fermentation. The recorded N_2O emissions of the BSFL composting process, as well as 34% of recorded CO_2 emissions, were primarily associated with digestion of insect matter by microorganisms, again not being associated directly with the larval feeding process[276]. With CH_4 and N_2O being the main GHGs contributing to GWP in agricultural systems, BSFL waste treatment may prove very effective for substantially reducing landfill-associated emissions[222].

In term of protein production, BSFL rearing has also been linked to a low GWP of 0.91 kg CO_2-eq per kg of protein, lower than soy (1.7 kg CO_2-eq), and significantly lower than chicken (19-37 CO_2-eq), pork (21 – 54 CO_2-eq) and beef (77-175 CO_2-eq)[42,277]. Soybean cultivation is heavily reliant on chemical fertilisers, the production of which contributes around 14-17% of total GHG emissions per year[278]. When considering negative GWP contributions, namely by accounting for the reduction in overall emissions made by the omission of fertiliser use when producing BSFL protein compared to soy, the GWP of BSFL protein production was significantly net negative (-432 kg CO_2-eq / kg of protein).

In another life cycle assessment, an estimated input of 15.1 MJ was required per kg of BSFL produced, with FW transport to the processing site and larval drying procedures being the main energy contributors in protein production[279]. BSFL therefore required much more than energy to produce than an equal weight of soya (4.1 MJ), but far less than chicken (80-152 MJ), pork (95-237 MJ) and beef (177-273 MJ)[42,279]. However, this assessment assumed that soya and BSFL were produced and consumed over the same area, which affects the interpretation of energy requirements; while BSFL can functionally be produced anywhere,

soya is usually farmed in a limited geographic area, and requires an energy-intensive supply chain for transport to the site of consumption. With this in mind, the energy requirements of soy can vary drastically depending on their transport chain.

Finally, BSFL rearing is associated with a low land use. 1 kg of protein could be effectively produced in as little as 0.05 m^2, with soy requiring 8.65 m^2, chicken 41-51 m^2, pork 46-63 m^2 and beef 142-254 m^2[279,42]. Putting this into context, soyabean cultivation is estimated to take up 133 million hectares to produce 348.9 Mt of soy annually, around 8.4% of the worlds cropland area and roughly the size of Mongolia[280,281]. The same amount of BSFL protein could be produced in 0.77 million hectares, an area smaller than Puerto Rico, with the shorter larval cultivation time allowing for even higher production over a prolonged period. All together, BSFL protein production and FW treatment shows remarkable promise as a more sustainable and circular practice compared with our current agricultural and waste treatment processes.

1.2.9 – BSFL Mass Rearing and Potential for Automation

Almost as important as their feed efficiency and nutritional value, BSF are a species are highly amenable to mass-rearing and automation. Insects have short life cycles and high fecundity, allowing for rapid scaling to meet variable substrate demand with little loss. BSF in particular have a high productivity metric, allowing for reliable production of over 1000 offspring from a single breeding pair in artificial rearing setups[137].

In practical terms, BSF are hygienic and easy to manage, with adults being poor fliers and the larva preferring to remaining in food sources. Non-feeding prepupa demonstrate a self harvesting ability, which can be easily exploited to remove the majority of biomass from spent feeds[282]. BSFL are capable of sanitising their substrate during feeding, inactivating pathogenic microbes and aerating feeds to reduce moisture and offensive odours[103,283–285].

Showing tolerance to a wide range of temperature and humidity conditions, BSFL are highly adaptable to different rearing setup designs of varying levels of sophistication. In addition to seasonal rearing outdoors, BSF breeding can be accomplished year-round through the use of basic climate controls and artificial lighting[134]. With minimal operational input and only general purpose equipment requirements, BSFL rearing has a low barrier for entry and is suitable for a ranging scale of enterprises in LMICs[286].

In addition to this, BSFL are also compatible for rearing at scale in high-throughput waste processing and protein production. With high fecundity and a brief life cycle generation time of around six weeks, BSFL biomass density can be rapidly scaled up and down over short timescales to suit the needs of the rearing facility and the amount of feedstock available[287]. This endows a remarkable process flexibility when compared similar mass-production setups, with the loss of larva not translating to a significant loss in resource investment if feedstock supply fluctuates.

With a low land footprint, insect farms are location agnostic and are suited for indoor, urban and vertical farming[288]. As such, facilities can be integrated close to existing supplies of water, energy and waste feedstocks, dealing with FW at source and reducing the burden on the waste management and transport infrastructures[289]. With a high adaptability regarding process scaling, staffing and operational costs can similarly be scaled to better match the level of FW available. However, low staff and resource requirements mean that BSFL production plants can be cost-effective throughout standard operation, performed by a small workforce with low training requirements[290].

In production facilities, several cost-effective process optimizations have been developed and employed to increase waste throughput and automate production. This includes mechanised components, such as roller conveyors and robotic systems for sorting larva and feedstocks, rotary drum reactors for large batch feeding processes, and automated feedback systems for environmental monitoring[291,292]. Various systems and workstations can be arranged as a sequential production line, starting with the pre-conditioning of the input feedstock and flowing through to larval separation, drying and milling operations at the end of the process[293].

With the BSF larval developmental stages being identifiable by age, size and colour, machine learning models have also been developed for the identification and tracking of BSF with cameras to better determine health, feeding and harvesting status[294]. Along with large capital equipment, other automation technologies are being developed to suit highly mobile or decentralised rearing facilites[295]. To remove the need for transporting low-value, high-density FWs to a central plant for processing, automated and modular rearing setups have been compartmentalised into small shipping containers or cabinet “smart bins”, suitable for temporary transport to sites of FW such as farms or breweries[293,296].

1.2.10 – Regulatory Landscape, Market Challenges, and Future Opportunities for the BSF Industry

Despite studies showing high digestibility of BSFL meal for a plethora of examined species, it is only regulated as livestock feed in a limited number of species in the EU and the UK[297,298]. In these regions, both live and processed BSF insect products are permitted for swine and poultry feed, encompassing pigs, broiler and laying chickens, turkey and quail. Processed insect products, such as hydrolysed BSF protein and fat extracts, are also permitted for ruminant livestock feeds, including sheep, goat and cattle. The types of waste feedstock used to rear BSF for feed purposes are also controlled, with vegetal and dairy wastes being generally acceptable and meat or manure-containing wastes not approved[299].

Subtle differences exist between UK and EU regulations regarding what types of processed BSF products can be fed to different species of livestock, with inclusion of live larva in feeds seeing the highest restriction. This ultimately stems from concerns from regulators regarding the utilisation of recycled animal byproducts in the feeds of certain farm animals, which has previously been associated with the spread of epidemics such as foot-and-mouth disease and bovine-spongiform encephalopathy[300].

Live and processed BSFL products are also regulated for use in aquaculture, which may grow into the dominant market for insect meal[297]. These products have been found favourable for nutrition in many commonly farmed fish, including trout, salmon, carp, seabream, seabass and shrimp[184,193,206,301–303]. While livestock and aquacultural feeds prioritise minimal, low-cost feeds for a viable business model, a promising secondary market for BSFL products is in organic, hypoallergenic pet feed[111,119]. At present, live and processed larva are sold to cater to dogs, cats, rabbits, rodents, wild birds, reptiles and amphibians, emphasising sustainability, nutrition and animal preference in regard to insect meals[304–306,111].

Looking forward, there is even a potential market for the use of BSFL products in human foods in the future. The nutritional and sustainability features of BSFL meal makes it equally appropriate for human diets as with agricultural feeds, with entomophagy being practiced across a variety of cultures[307]. EU regulation permits four farmed insects, including cricket and mealworm but not BSF, as a novel food for human consumption, with BSFL having been submitted for further evaluation[308]. However, BSFL are currently under limited licence as

“novel food” in the UK, allowing for the regulated sale of these products for human consumption[298].

While western attitudes are found to be generally averse to eating insects, surveys have indicated that acceptance of entomophagy increases when the inclusion of insect products in food can be made less obvious[307]. Researchers have already begun exploring the techno-functional properties of BSFL for suitability in human cuisine, with prototype burgers and sausages being constructed from larval biomass as an example for meat analogue products[309,310].

For better palatability, BSFL can also be refined into insect flours and dietary oil extracts. Bread, butter and margarine have been produced from ground BSF protein and lipid fractions with favourable taste and nutrition profiles, with these compounds perhaps being suitable for other grain-based foods like pastas, biscuits, soups and pastries[311,312]. However, the potential allergenicity of BSFL meals may prove a point of concern for regulators, with primary sensitisation to chitin, tropomyosin and arginine kinase by skin contact and inhalation being documented with BSFL use and within entomophagous populations[313,314].

As the market for BSFL as feed matures, the main test for success will involve its ability to displace the dominance of soy and fish meal use in cheap agricultural feeds[315]. Soy is by far the more dominant of the two protein sources, with production increasing exponentially over the last 30 years and being crucial part of the economy of several LMICs[271,316]. However, this has come at an enormous environmental and social cost, with exploitation, inequality and conflict rife among communities affected or displaced by the trade, and the impact of soy-associated deforestation and pollution being felt across the globe.

Despite having a long record of affordability, the unsustainability of soy production and transport has begun to be reflected in recent price fluctuations, with the destabilising effects of the COVID-19 pandemic only highlighting potential weak spots throughout international supply chains[317,318]. Protracted global conflicts have increased inflationary pressure concerning several key agricultural products, with this “agflation” making access to feeds and fertilisers much more volatile[319,320]. Going forward, the changing climate is suspected to greatly affect the reliability of soy production for all stakeholders concerned[321].

With UK food security and agricultural sectors potentially at risk, concerted efforts should be made to decrease dependence on distant, unsustainable sources of feed production and look towards producing more sustainable feeds closer to home, and more in-line with

net-zero ambitions[139]. Additionally, the benefits of BSFL as a feed replaced are not limited to farming at industrial-scales. In LMICs, feed represents around 70% of the total production costs for livestock farmers, with the increasing price and decreasing availability of soy and fish meal products limiting the livelihoods of many smallholders[322]. Independent assessments have each suggested that even a partial replacement of soy and fish meal use with BSFL protein in Kenya, Malawi and Uganda would provide huge economic and agricultural benefits in each country, and be sufficient in each case to lift millions of people above the poverty line[290,323,324].

A recent market analysis indicated that the UK imported 1.3 Mt of soy in 2022 for use in agricultural feed, while public pressure to reduce soy import is increasing due to its associated environmental and social impacts[298]. The authors suggested that diverting only 25% of either retail, farm or manufacturing wastes towards BSFL production could replace 28% of UK soy imports destined for livestock feed, with BSFL-fed meat having only a moderate price premium of 5% in poultry and 9% in beef products.

As another option, pivoting towards replacing fish meal in the aquacultural feed market rather than focusing on the soy market may be more realistically achievable, at least in the short term, due to the comparatively high cost and low utilisation of fish meal[325,326]. In addition, a greater diversity of farmed fish species are permitted for rearing with BSFL feeds than terrestrial animals under current UK legislation, allowing BSF producers to specialise in catering to these niches[297].

Despite low rearing costs, and even lower feedstock costs when using FWs, BSFL protein production is not yet able to compete with the current price of soy, which enjoys an entrenched global supply chain and dominates the feed market[327]. In the same market analysis, it was estimated that a 70% decrease in the current cost of BSFL production in the UK would be required to outcompete soy on a protein weight basis[298]. Globally, BSFL farmers are now focusing on methods to increase larval growth efficiencies and reducing overall production costs, which is crucial to the survival of the insect farming sector in the long term[328].

Although ultimately vulnerable to market trends, adapting regulations could help the nascent UK and EU BSFL sectors become more competitive with the conventional agricultural feeds. For example, a significant limiting factor regarding production cost and scalability is the restrictions around the type of waste feedstocks that can be used for BSFL rearing.

Currently, around 4 Mt of FW is estimated to be available for UK-based BSFL processing per year, comprising mainly of “cleaner” waste types such as brewer’s grains and dairy, vegetable or bakery by-products[298]. Conversely, around 95 Mt of organic waste, including animal manures, wastewater sludge and anaerobic digestate, are not permitted for use in BSFL production due to their possible animal by-product or pollutant content.

While the safety of feeds ultimately used to produce our own food is understandably paramount, policies for effective waste separation, reclassification and pre-processing could free up a significant amount of biomass from lower-risk waste types for BSFL and frass production. For example, if able to access either household, hospitality or retail food waste alone, which jointly contribute 5 Mt of waste in the UK per year, the price of BSF meal is estimated to drop below £400 per ton, approaching the approximate price of soy at £300 per ton[298]. Although assessing and ensuring safe access to these feedstocks poses significant research and regulatory challenges, steps such as facilitating the easier access and transport of cleaner sources of FW to insect farms can be taken to enable growth in the sector and begin to integrate BSFL treatment methods in our waste management infrastructure.

1.3 – Genetic Strategies *for* Improv*ing* Black Soldier Fly

1.3.1 – Genetic Innovations in Enhancing Agriculture

With the all the beneficial characteristics described, the rearing of BSF as mini-livestock demonstrates clear potential as a sustainable technology with numerous practical, economic, environmental and welfare-related advantages over conventional agricultural feeding practices[329]. As a non-domesticated species, BSF has already proved remarkably versatile and highly efficient for a number of different uses. However, as with all crop and livestock species employed for agriculture, desire will grow for increasing yields, simpler rearing strategies and new areas for value creation as BSF farming sees higher adoption. Considerable steps have already been made in optimising the BSFL rearing process for automated and cost-effective rearing at scale[135,330]. However, gains through optimised husbandry are likely to soon reach a ceiling, with commercial BSF colonies already starting to shifting interest towards augmenting the BSF genetic background for further yield improvements[331].

Whether intentional or not, all domesticated and agricultural species have been changed throughout history to better suit human needs, with breeding strategies employed to select and enrich for desirable traits[332]. More specifically, selection in breeding involves the mating of members of a species that have beneficial genotypes, identifiable through links to an observable phenotype, with the aim of increasing the strength or prevalence of that phenotype over time[333].

For edible species, this generally involves increasing useful yield through increasing size or nutritional value, or decreasing the time taken to harvest. Termed yield or economic traits, some desirable characteristics that related to BSF could include increased larval body weight, reduced larval development time, lower FCRs and increased bioconversion rates, and higher protein or lipid compositions[333]. Beneficial yield traits could also apply with non-larval traits, with examples being increased egg production or fertilisation rate.

In addition to yield, humans have selected for domestication or fitness traits to help facilitate or enhance the rearing process[332]. For example, crops and livestock have been bred over

time to perform optimally in a number of different climates, increasing their robustness to environmental factors like temperature, or increasing their resilience to diseases[334]. Additionally, beneficial behaviours and temperaments more conducive to rearing in captivity have also been enriched through breeding. These fitness traits would all apply equally well to BSF, with additional desirables including resistance to desiccation, low-nutrient waste tolerance and enhanced abilities to digest different feedstocks.

Artificial selection, or selective breeding, has been practiced in some form with most crop and animal species since prehistory; genetic evidence indicates signs of intentional selection applied to chickens from as far back as 4,500 yeas ago, with pigs and cattle some ~10,000 years ago, and in dogs from at least 14,000 years ago[335,336]. However, it took until the practice of farm enclosure in the 18th century for selective breeding to become tracible and feasible at scale, with the intentional crossing of different livestock breeds with desirable characteristics helping to propel the agricultural revolution[337]. The birth of genetics as a discipline in the 20th century then gave way to quantitative genetic theories as a way of understanding the principles behind selection, followed more recently by molecular genetics[338].

Quantitative genetics allows for directional enrichment of continuous traits or phenotypes such as height or body mass. Quantitative, or complex, traits are polygenic in nature, with multiple genes each likely to cause a small effect and so difficult to isolate[339]. These genes are often epistatic, interacting with each other and environmental factors to influence the observed phenotype, and pleotropic by contributing to numerous other traits. The relative contributions of complex genetic and environmental factors leads to both genetic and phenotypic variation for different traits throughout wider populations, with quantitative methods employed to model their distribution and inheritance to enrich for desirable characteristics.

Quantitative trait loci (QTLs) are regions of the genome that influence variation of a complex trait, with multiple QTLs often indicated as contributing towards specific phenotype[340]. Molecular markers within QTLs, such as microsatellites or single nucleotide polymorphisms (SNPs), can be mapped to help identify other QTLs and their relative contributions to influencing the complex trait in question. With low cost next generation sequencing (NGS) technologies, it is now more feasible to gather whole genome sequencing data from a range of individuals with varying phenotypes and compare the frequencies of SNPs across a large

group[341]. These genome wide association studies (GWAS) are powerful methods for revealing the genetic loci that contribute towards trait variation, within which the potentially causative genes and regulatory sequences can be interrogated for enrichment of traits in selective breeding programs.

With epistasis, epigenetic and pleotropic factors all contributing to the eventual phenotype, the main limitation for disentangling quantitative genetic data is not necessarily the identification of QTLs, but in identifying the causal DNA sequences themselves; sequences that contribute towards variation are often found in protein non-coding regions, altering the context and relative level of gene expression but the function of the gene itself[342].

Despite this, quantitative genetic approaches have seen high success and employment in enriching desirable phenotypes for a number of different crop and livestock species. Selective breeding programs have allowed for a 43% reduction in FCR and a 243% increase in body weight in broilers over a 45-year period[343]. In a similar timespan, pigs have been associated with a 27% decrease in FCR and 37% increase in lean meat by weight. Breeding approaches in dairy cows has seen a 67% in milk production throughout this period, with hens now laying 30% more eggs per year and 80% more per kg of feed since the 1960s. Not all genetic gain went towards increasing product yields; the GWP of chicken, pigs and dairy cows, through a reduction in GHG emissions, also declined by 24%, 15% and 16% respectively in this period.

1.3.2 – BSF Selection for Improved Traits

With little evidence of intentional BSF breeding and domestication before establishing how to rear them in captivity 20 years ago, the species is likely fertile ground for the intentional enrichment of traits through artificial selection[134,328]. Cheap to rear, with generation times of around 6 weeks and a high fecundity of 1000 offspring or more, these insects are an ideal candidate for realising performance gains through quantitative genetics in a short space of time.

Interestingly, quantitative genetic approaches have also been applied with several commonly reared insect species. In general, body mass has been found to be a medium to highly heritable trait in insects, with development time being moderately hertiable[333,344]. A 10-year breeding program in European honey bees (*Apis mellifera*) led to the genetic gains of 17% increased honey yield, 15% increased resistance to *Varroa* mites, and a 15%

reduction in defensive behaviour against humans[345]. In silkworm (*Bombyx mori*), cocoon weight was increased by 30% and cocoon shell weight by 46% over a 2 year breeding period[346]. Over an 8 year period, the pupal body weight of the farmed insect yellow mealworm (*Tenebrio molitor*) was increased by 66%[347].

Most insects have evolved following the r-selected life history strategy, characterised by high fecundity and short life spans to optimize their reproductive output and evolutionary success[348,349]. While this may be beneficial for scaling and speeding up the process of artificial selection, it can also lead to faster accumulation of negative traits without the regular introduction of genetic variation to the gene pool. This is particularly pertinent for farmed insects, in which many colonies are established from a limited population of wild-caught founders that quickly experience genetic bottleneck and drift effects in a closed breeding cycle[349,350]. A study on allozyme heterozygosity in a captive population of *Drosophila melanogaster* found that, despite maintaining a large breeding pool of over 3,500 flies per cage, up to 62% of genetic diversity was lost over 26 generations, equivalent to 1.6 years given a 23-day generation interval[331,351]. Artificial selection for a desired trait can be complicated by the unintentional enrichment of negative traits, masked in genetically diverse populations, leading to inbreeding depression and low fitness[331].

There is already some promising data regarding the improvement of BSF genetics with artificial breeding programs. In a study employing automated rearing procedures over 10 generations, the BSFL line under selection saw a 39% increase in larval body mass, a 32% increase in protein content, and a 21% increase in fat content[333]. Due to the rapid generation time of around 6 weeks BSF, this corresponded to an average annual improvement in larval body mass of 17.5 - 21.5% under production settings, significantly higher than encountered with chicken (5%, generation time of 20-24 weeks) and yellow mealworm (8%, generation time of 18-24 weeks).

Interestingly, there was no significant decrease in larval FCR, usually noted accompanying increased body weights with other livestock species, suggesting that these larvae may have simply eaten more feedstock than the non-selected population. With the majority of GHG emissions coming from uneaten feedstock in BSFL processing, reducing FCR is highly desirable for reducing GWP potential[352,353]. Several other BSF breeding programs have been described with similar outcomes, with enriched traits in question being increased larval mass, pupal size and adaptation to a specific feedstock[354]. A strategy of mass selection is

commonly employed, allowing for mixed groups of larva to be reared and mated as a pool, with random sampling and selection of a limited number of BSF that are most enriched for the desired trait to contribute to the next generation[355].

However, quantitative genetic breeding strategies in BSF are associated with some significant challenges. For one, insects are difficult if not impossible to reliably track and assess over their lifecycle, with identifying markings hard to implement due to frequent exoskeletal moulting[356]. In turn, this makes it difficult to establish genetic parameters, such as larval weight and length, across multiple sampling points. It also makes it impossible to account for the pedigree and family association for each insect selected to seed the next generation, hindering the enrichment of multiple desirable traits and potentially amplifying non-desirable ones[354].

Larva can be reared separately for more reliable tracking, although this drastically reduces the practical sample size and hinders interpretation[344,356]. The high number and short generation time of BSFL works as a disadvantage when it comes to phenotyping, leaving a very limited time to perform an assessment for a large number of larvae simultaneously before their life cycles become desynchronised and the window for breeding closes. While feasible at smaller scales, tracking the growth metrics of individual larva is significantly more time and labour-intensive than when rearing larva as a group, and generates data that are challenging to generalise to mixed population rearing setups that are much more commercially representative. Feeding larva in isolation does not account for the effects of crowding behaviour on growth performance, with larval density being known to strongly influence their body temperature in addition to the consistency and moisture content of the feeding substrate[356].

Another issue is that feeding and phenotyping experiments remain highly unstandardised across different labs and production facilities, with variations in feed batch size, larval density, and feed composition and availability[328]. Disentangling genetic from environmental influences in the phenotype is especially difficult in insects of small size and short generation times, with small variations in environment like temperature and feed aeration potentially playing an oversized role in insect performance. Choosing an appropriate experimental scale is also important, with some trials utilising small population sizes for establishing genetic parameters failing due to the lower temperature, gas accumulation and dried out feed associated with rearing few BSF in smaller containers[356]. Finally, genome sequencing

data regarding BSF and other insect species is limited at present, making identification of QTLs and comparison of variability between different genetic backgrounds difficult [333].

Classical selective breeding is inherently an incremental process, with a large amount of phenotyping being required over multiple generations to see economically-relevant gains in a short timeframe[357]. Additionally, one of the biggest hinderances in any breeding program is a lack of genetic variation extant within the sample population. In essence, the achievable diversity in a selectively bred strain is effectively limited to the combination of variance found within the founding population, something that is poorly documented in BSF[358].

When looking to find and enrich breeding properties, wild populations are often a valuable source of phenotypic diversity that can be introduced into more intensive breeding programs to offset a low founding gene pool. Recently, 150 populations of reared and wild-living BSF were characterised on a global scale, revealing distinct genetic clusters, areas of reduced diversity and potential early signs of domestication[358]. While this data is invaluable for sourcing diversity for a breeding setup, often the cost of population-wide genetic analysis for enrichment of traits is prohibitive to smaller enterprises[349].

In addition, many potentially beneficial genotypes can be inferred from studies in other species, with the bottleneck to genetic improvement being the ability to maintain a colony large enough to avoid inbreeding depression and make significant progress when selecting for multiple traits. One solution to verifying, and even generating potentially beneficial genotypes, could be to induce diversity in a small breeding population through intentional genetic changes through the use of genome editing technologies.

1.3.3 – Upgrading Agriculture with Gene Editing Technologies

Recently, the UK has joined several other countries in relaxing regulations on the use of gene editing (GE) technologies, such as CRISPR, for agricultural applications in plant crops[359,360]. Crucially, while mutagenesis of endogenous material (gene editing, GE) is now expected to be permissible for certain uses, insertion of foreign genes into products (transgenesis) will likely remain under different, more stringent regulations. For translational agricultural research, an emphasis is placed on the improvement of farmed organisms using GE technologies rather than transgenic insertions, with genomes produced through GE means having a higher potential to have arisen “naturally” through the selection of rare beneficial mutations in breeding programs[361].

The 2023 Precision Breeding Act aims to distinguish farmed species subjected to GE from genetically-modified organisms (GMOs) in terms of regulatory acceptance in the UK and facilitate their adoption into the food supply chain[359,362]. Despite being originally written to apply to both plants and vertebrate farm animals, the Act is currently being implemented in a phased approach, limiting GE use to plant crops while animals are still subject to previous legislation[363]. At present, it remains unclear if or when the Act will be expanded to implement precision breeding in vertebrate animals such as livestock and fish, and if invertebrate species will be covered under this regulation at all.

Already in somewhat of a regulatory quagmire even without the implementation of GE or transgenic technologies, the future legislation of precision breeding in novel feed insects like BSF remains uncertain[360]. However, considering the common use of transgenic products as livestock feed in many jurisdictions, both public and governmental acceptance of genetically enhanced BSF may be possible in the near future, with a lower legislative barrier to overcome when use is restricted to animal feeds and not for direct human consumption[359].

To increase the speed of genetic improvement towards increasing yields in BSF, the application of both GE and transgenesis technologies should be considered. GE has been successfully employed for phenotypic improvement in crops, most commonly for removing deleterious alleles[364]. Several commercial plant strains have been created using CRISPR, with traits such as increased thermal and desiccation tolerance, enhanced disease resistance, improved nutrient composition and higher overall yields[365].

CRISPR can be directed to alter the level and context of gene expression by disrupting the gene regulatory sequences in non-protein coding regions, one example being the ablation of transcription factor binding motifs[366]. Sequences that mediate alternative splicing mechanisms can also be targeted to ensure preferential inclusion or skipping of alternate exons[367]. The technology can also be used to fix deleterious polymorphisms or mutations in the protein coding sequence (CDS) of a gene, which has been employed for imparting disease resistance in different crop and animal species[368].

However, tuning gene expression and introducing beneficial mutations using GE techniques requires a deep understanding of the function and gene structure of the target gene of interest (GOI) in question. Far easier to achieve, the more common strategy employed in GE is to induce null or “knock out” (KO) mutations by disrupting the reading frame in the CDS of the target gene[369]. To increase the possibility of GE leading to a loss of gene function,

regions in the CDS that are highly conserved in orthologues genes across several species are preferentially targeted, with sequence disruption around critical amino acid codons highly likely to disrupt the activity of the protein when expressed[370].

CRISPR-Cas9 induced KO mutations have been applied in livestock breeding, although methods of producing stable hereditary mutations in animals are more complex than those available for crop species[371]. Successful examples of this include the KO of the myostatin (MSTN) gene, a negative regulator of muscle growth that leads to a large muscular phenotype when mutated in cattle, sheep, pigs and goats[372–374]. Targeting growth factors such as IGF1 and IGF2 in the insulin signalling pathway has similarly lead to increased growth and meat yield in livestock[375,376]. The majority of GE so far applied with livestock has focused on similar production traits, such as muscle growth, or on manipulating target GOI to make animals more resilient to disease[377].

Large gene sequences can be even fully deleted from the genome through the simultaneous targeting of sites flanking the gene. CRISPR has been used to biomimic natural alleles from other crop strains or species, with the repair outcomes at a target locus altering the native gene and recreating the function of the desirable allele[378]. This method has been employed for the *de novo* domestication of wild crops with high abiotic tolerance, but low production traits, to increase product yields[379].

In addition to creating ideal genotypes informed by *a priori* studies of gene function, GE studies can aid agricultural research by identifying and validating promising candidate GOI identified through breeding programs that may or may not produce a desirable phenotype when inactivated. Marrying the selective and GE-assisted breeding technologies, candidate GOI can by elucidated via GWAS, QTL and SNP analyses, although these require considerable amounts of sequence profiling from individuals within different populations[380]. These methods only capture genetic information relating to the range of trait variation extant in nature, while more desirable phenotypes may be underrepresented or not present in the wild.

However, rapidly inducing diversity for a select few alleles may be possible through GE. Using CRISPR tools, multiple GOI could be targeted specifically for inactivation, with the edited animal capable of passing the edits onwards to progeny. If mating several gene edited individuals, these hereditable edits can stack to give unique genotype mixes that would be difficult to accomplish with conventional breeding. As most valuable yield-specific traits are

likely impacted by the interaction of several genes, this method is a rational attempt to interrogate a limited number of trait-contributing genes in one population or individual[381].

The minimal time requirement and targeting specificity afforded by GE methods could avoid the issues of linkage drag, inbreeding depression and the need for large breeding pools over multiple generations associated with contemporary selective breeding programs[382,383]. In addition, multiplexed GE could allow for simultaneous targeting of multiple genes or pathways that may otherwise have epistatic or synergistic effects, giving control over an exact defined genotype, otherwise isogenic to related non-edited genetic backgrounds[384].

1.3.4 – CRISPR Technologies for Precision Breeding

The use of GE for inducing genetic diversity and creating beneficial genotypes holds considerable promise for increasing the value of BSF products and processes[264]. As a species of only recent interest, however, few GE and transgenesis tools have been applied in BSF, with more fundamental research required. With simple life cycles, short generation times and amenability to transgenesis, CRISPR technologies have been extensively tested in several different insect systems, including the model fruit fly *D. melanogaster,* the commercial silkworm *B. mori,* and the malarial mosquito vector *Anopheles gambiae*[385–388].

To enable GE techniques for BSF breeding for the wider community, the basic CRISPR components need to be fully characterised in this system, both when supplied exogenously and or expressed from an endogenous source. The molecular toolkit available in the species can also be expanded to include methods of stable transgene introduction into the BSF genome. Finally, the most suitable methods for inducing both GE and transgenesis in BSF need to be explored, with hereditable changes to the genome needed to retain these modifications over generations.

The classic CRISPR-Cas9, originally identified in *Streptococcus pyogenes* as part of the bacterial immune system, can be used for GE purposes by leveraging the host's native DNA repair mechanisms to introduce stable genetic mutations at a specific target site[389]. In CRISPR-directed GE, the CRISPR components first create a double stranded break (DSB) in the genomic DNA. This starts a cycle of error-prone DNA repair *via* the non-homologous end joining (NHEJ) repair mechanism and re-cleavage of the repaired DNA by the CRISPR components[390]. This DNA cleavage and repair cycle makes likely the eventual accumulation of frameshifting insertion-deletion (indel) mutations at the target site[391].

Classical NHEJ is the major DNA DSB repair pathway in eukaryotes[392]. NHEJ can occur at any point in the cell cycle, including G1 where only single copies of the chromosomes are present. Importantly, NHEJ repair can occur without a repair template, and as such is more likely to result in error-prone repair outcomes. The major alternative DNA template-dependant repair mechanism, homology direct repair (HDR), is usually restricted to the G2 phase, where sister chromatids are in close proximity[393]. In this scenario, HDR can be employed to accurately repair the broken chromatid through a templated mechanism, using the intact gene copy as a guide sequence.

In CRISPR-Cas9 cleavage-induced NHEJ, a ribonucleoprotein (RNP) complex of a Cas9 nuclease protein and a target-specific guide RNA (gRNA) is introduced to the intended mutant target site (**Figure 2**)[394]. The gRNA sequence is comprised of a “core” or “scaffold” sequence, used to bind with the Cas9 nuclease, and a 20 base-pair “variable” or “spacer” sequence that directs the RNP complex to an intended DNA target sequence. The variable sequence corresponds to a 20 nucleotide stretch on the genome that is immediately upstream of a NGG (5’-3’) DNA motif, known as the protospacer adjacent motif (PAM)[395].

Once recruited to the sense strand on the double stranded DNA (dsDNA) target site through base pairing *via* the gRNA spacer sequence, the RuvC and HNH nuclease domains in the Cas9 protein create a DSB in the genome, approximately three bases upstream of the PAM site. Immediately following DSB formation, degradation may occur at the exposed DNA ends, called resection[390]. In the first phase of NHEJ, a Ku70-Ku80 heterodimer is recruited to each of the exposed DNA ends, protecting them from further damage and providing a docking site for the DNA-dependant protein kinase catalytic subunit (DNA-PK_{cs})[396]. DNA end processing can occur if they are found incompatible for immediate ligation. For this, the endonuclease *Artemis* is recruited by DNA-PK_{cs}, where it removes single stranded extensions from the DNA end, which are then filled in by DNA polymerase λ or μ[397]. The ligase IV/XRCC4 and XLF complex then ligates the blunted ends together to fix the DSB, liberating the NHEJ repair complex.

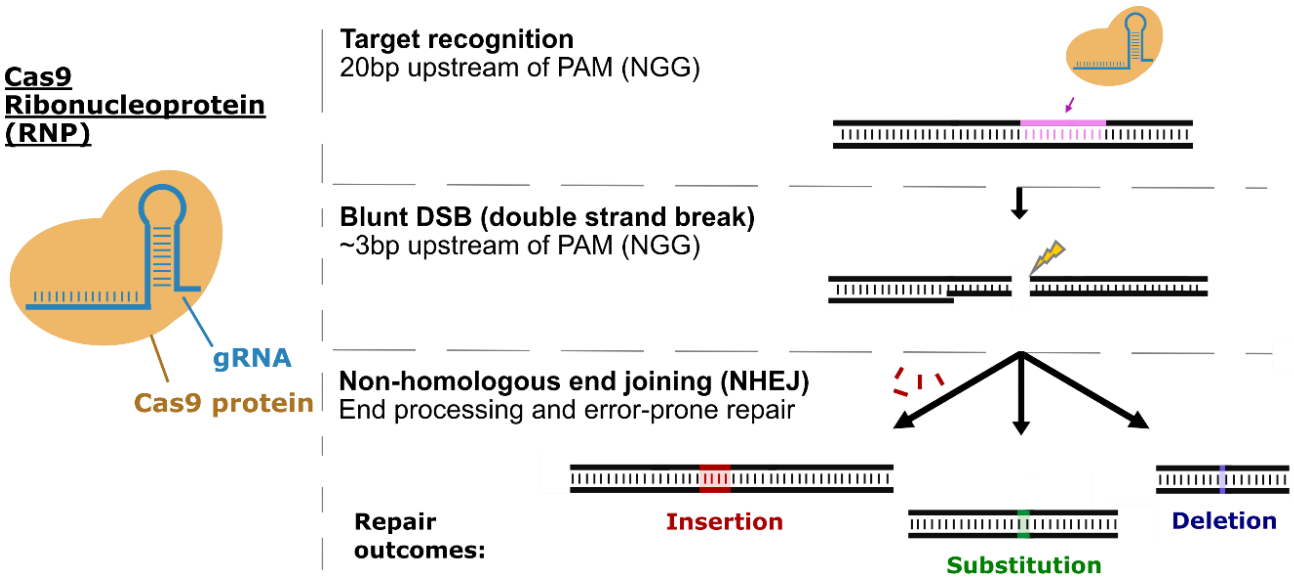


Figure 2: The CRISPR-Cas9 mechanism for indel formation via NHEJ-based DNA repair. The Cas9 ribonucleoprotein (RNP) complex is formed from the CRISPR Cas9 nuclease protein and a short, single-stranded guide RNA (gRNA) that encodes homology to a DNA target sequence[389]. Once a cognate DNA sequence, or one that corresponds to the 20 nt variable sequence in the gRNA, in is encountered in the genome, the Cas9 protein mediates a blunted double-stranded break (DSB) three bp upstream of the PAM sequence. For Cas9, the PAM recognition sequence required in the genome for DNA cleavage follows the motif NGG, where N can equal any DNA base. The DSB is generally fixed via templated methods, such as homology-directed repair (HDR), or non-templated methods such as non-homologous end joining (NHEJ). In NHEJ, the exposed DNA ends are remodelled by a series of host factors and eventually re-ligated, reconstituting the original target sequence and allowing repeated rounds of cleavage by Cas9. Without a repair template, NHEJ can often lead to inaccurate repair outcomes, with the removal or addition of nucleotides leading to base substitutions or insertion and deletion mutations (indels). Accumulating indels at the target site eventually result in the inability of Cas9 to recognise and re-cleave the target, ending the GE cycle and leaving a genomic sequence with remodelling around the target PAM site region[390].

Endonuclease, polymerase and ligase components can be recruited to the DNA-PK$_{cs}$ in any order, giving NHEJ repair flexibility for different initial conditions at each exposed DNA end until a sufficient scenario for ligation is encountered[398]. Processing is likely to be iterative and require multiple modifications by the NHEJ repair factors, leading to potentially different repair outcomes depending on the order and frequency of recruitment before both strands are ligated. When cleaved with Cas9, the nuclease is likely to generate a clean, blunt cleavage that can be repaired without error in most cases[390].

However, small 5' or 3' overhangs may also be created at the cut site, which could face resection or be filled in by *Artemis* or the NHEJ polymerases respectively, leading to the generation of indel mutations[399]. In this instance, nucleotides are incorporated or removed

at the exposed DNA ends, which is converted by the NHEJ mechanism to double-stranded base pair in the case of insertions, or omission of the original sequence in deletions. Indel formation is close to a stochastic process, with the mutation size needing to be a multiple of three nucleotides to preserve the original reading frame in protein coding regions[400].

Accurately repaired target sites will become susceptible to repeated cleavage, with the cleave and repair cycle continuing until the target site is repaired in a way that the Cas9 complex cannot recognise it for further cleavage. This tendency pushes the system towards fixation of non-perfect repair outcomes, at which point the cleavage cycle can be broken, leaving an edited sequence at the CRISPR target site[401].

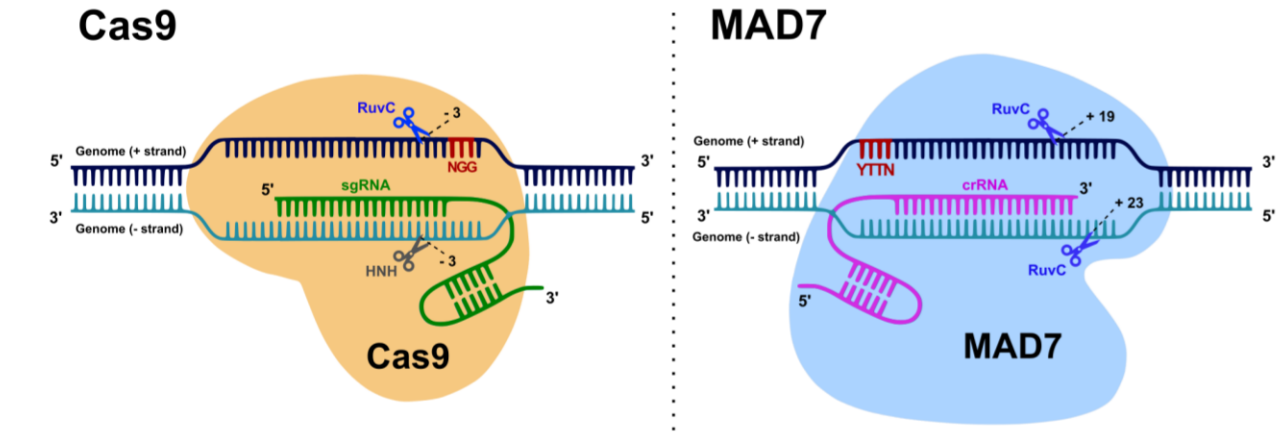
Creating stably inherited mutations in animals is not a trivial process, requiring laborious microinjection procedures and several generations to validate any genetic modification. The value of GE interventions, however, can be greatly increased through multiplexed targeting of several GOI simultaneously. As described, this can be used to induce genetic diversity at a number of genetic loci, leaving the potential to identify desirable phenotypes from a mated population of organisms encoding different combinations of editing outcomes. This may be especially useful for rapidly assessing multiple candidate GOI in a determined QTL region, with the effect of indel mutations in each gene possibly being revealed in the resulting phenotype.

1.3.5 – Emerging Tools for Multiplexed Gene Editing

Despite the use of complex multiple gRNA expression vectors, Cas9 has an overall limited ability to create edits at multiple targets simultaneously[402]. However, several other CRISPR-associated nucleases, such as Cas12a, have been shown to demonstrate an improved ability to mediate multiplexed editing[403]. With no evidence for use of non-Cas9 nucleases in BSF engineering to date, the validation of alternative CRISPR components for their suitability in the BSF genetic toolkit could prove valuable for generating diverse editing outcomes.

The engineered Cas12a-like nuclease MAD7, or *ErCas12a*, is capable of efficiently processing tandem arrays of gRNAs and mediating targeted cleavage at multiple sites (**Figure 3**)[404]. Multiple targets can be designated for MAD7 activity from a gene expressing a single CRISPR-RNA (crRNA) array. In this context, the crRNA array is a polycistronic precursor RNA containing a series of alternating direct repeat (DR) sequences and variable

spacer sequences[405]. The 35 nt DR sequence is analogous to the scaffold sequence in Cas9-associated gRNAs, and is identical throughout the crRNA. These are interspaced with 21 nt spacer sequences, each encoding a region of homology that directs MAD7 to the different genomic targets. MAD7 utilises the PAM sequence YTTN (5'-3'), with Y encoding either a C or T, immediately upstream of the spacer homologous region on the genome[406].



Type:	**Class 2, type II**	**Class 2, type V**
PAM:	**NGG** (GC-rich)	**YTTN** (AT-rich)
Cleavage profile:	**Blunt** (~3 bp upstream of the PAM)	**Staggered** (~19, 23 bp upstream of the PAM)
Spacer (nt):	**20**	**21**
Size (kDa):	**~160**	**~148**
RNAse activity:	**Not present**	**Present**
Multiplexing:	**Limited**	**Flexible**
Licensing:	**Restricted**	**Open access**

Figure 3: Cas9 and MAD7 nuclease profiles. MAD7 is an engineered variant of the Cas12a nuclease originally isolated from the bacterium *Eubacterium rectale*, and so is occasionally referred to as ErCas12a[407]. MAD7 and the *S. pyogenes*-derived Cas9 nuclease have fundamental differences in their protein domain structures, interacting with different gRNA sequence motifs, utilising different PAM sequences and mediating different DNA cleavage mechanisms. While the RuvC and HNH domains of Cas9 mediate a blunt DSB 3 bp upstream of the NGG PAM site, a single RuvC-like domain in MAD7 mediates a staggered DSB at 19 and 23 bp downstream of the YTTN PAM site, on the PAM containing and antisense DNA strands respectively. Unlike Cas9, MAD7 possesses both RNA and DNA processing activities, and so can be employed in processing its own gRNAs from a long polycistronic precursor CRISPR RNA (crRNA). This makes MAD7 more suitable for multiplexed editing.

MAD7 has both RNAse and DNAse domains, allowing for the nuclease to process the crRNA transcript into individual gRNAs, each one comprising of one DR and variable sequence. Unlike Cas9, MAD7 has a staggered DNA cleavage profile, leaving 5' overhangs of 4nt by cutting 19 and 23 bp upstream of the PAM on the sense and antisense strands respectively. Cas12a-like nucleases such as MAD7 typically create larger indels than with Cas9 repair outcomes, with the staggered cleavage product being more susceptible to resection[408]. Additionally, the large distance between the PAM and cleavage site means that smaller indels are likely tolerated for target recognition by MAD7, prompting further cycles of cleavage and NHEJ until the target is sufficiently altered enough to be unrecognisable.

With an indel rate comparable to the commonly used Cas9, MAD7 is advantageous for use in the research and commercial communities. Unlike Cas9, MAD7 is provided under a royalty-free licence for academic and R&D purposes, with products developed from the technology similarly not requiring a licence if the nuclease is not present in the final product[409]. In context, MAD7 could be employed to mediate efficient multiplexed GE in BSF intended for sale, with the edited strain not being subjected to additional licensing fees for using this technology.

Developing the BSF genetic toolkit with both Cas9 and MAD7 in mind, however, greatly increases the range of target sequences available for cleavage by facilitating access to both GC and TA-rich regions of the genome. Although validated for multiplexed editing in multiple expression systems, MAD7 activity has yet to be tested in insects[410–412]. Characterising this nuclease in BSF has the added benefit of increasing the democratisation and adoption of CRISPR-derived products within the agricultural industry.

1.3.6 – Strategies for Insect Transgenesis and Transformation

To better characterise the CRISPR system in BSF, a transgenic approach that involves the endogenous production of both nuclease and gRNA components as stable transgenes could be considered. In other insect species, transgenic expression constructs are often designed to incorporate genetic regulatory elements, such as promoter sequences located upstream of endogenous genes, to approximate both the level and spatial-temporal pattern of native gene expression when used to drive the production of an introduced transgene[385]. With the optimal regulatory context, BSF-mediated expression of the CRISPR toolkit could potentially achieve higher gene editing efficiencies than methods relying on the exogenous delivery of

these components, such as CRISPR-RNP embryonic microinjection. Additionally, a robust CRISPR system based on heritable transgenes could serve as a reliable platform for comparing the relative activities of various nucleases and gRNAs targeting different genomic loci, with combinations of these elements easily achieved through the crossing of separate transgenic lines.

In addition to creating indel mutations, the CRISPR system can be employed to direct site-specific integration of transgenic DNA at specific sites in the genome[413]. This commonly involves the creation of a DSB at the intended integration site and the provision of a DNA repair template, containing the transgene flanked by regions of homology to both sides of the cut site. The HDR mechanism can them mediate the repair of the DSB using the provided repair template, which effectively copies the transgenic cargo sequence into the cleavage site. Although ideal at creating site-specific integration, HDR is a much less efficient process than gene KO through the NHEJ mechanism, and has not currently been reported in BSF[414]. Although enabling HDR-directed repair mechanisms in the species will play an important role in BSF engineering going forward, a simple and efficient method of transgenesis has already been widely employed in this and numerous other insect species.

The *piggyBac* DNA transposon system allows for non-site specific integration of transgenic cargo, and has been widely employed for insect transgenesis[415]. Isolated from the cabbage looper moth, the *piggyBac* (*pBac)* transposase enzyme recognises inverted terminal repeat sequences (ITRs) flanking a DNA cargo sequence, and mediates the integration of the combined cassette into any accessible TTAA sequence motif present throughout the genome[416]. This has been exploited by researchers to mobilise transgenes into random genomic sites in different organisms, with provision of a *pBac* donor vector and a source of *pBac* transposase expression in the form of plasmid DNA. Cointroduction of both plasmids in cells results in highly efficient transgenesis of the DNA cargo element in the donor vector at one or multiple TTAA sites in the genome, with transgenic-encoding cells identifiable through expression of a transgene-embedded marker gene such as GFP[417].

Contrary to a truly random pattern of genome integration, transgenic constructs that employ the *pBac* system are typically enriched in regions of open, transcriptionally-active chromatin, and particularly within large chromatin loops associated with highly expressed genes[418]. *pBac* is also associated with a high frequency of intragenic integration, with constructs often inserting near or within actively transcribed genes. Although the practical effect of insertion

location often appears to be neutral, the specific site of *pBac* construct integration should not be overlooked as a possible contributor to any transgenic phenotype observed, especially in cases where the function of endogenous genes may be disrupted[419].

A more difficult problem involves how to effectively introduce transgenic cargo or CRISPR components into the organism. The majority of insect geneticists employ the embryonic microinjection technique, involving the physical injection of plasmids, RNA or protein directly into the yolk of developing eggs (**Figure 4**)[420]. Embryo development is understood to be highly conserved, and BSF likely follows the same processes and stages as encountered and described in the developmental model *D. melanogaster*[421].

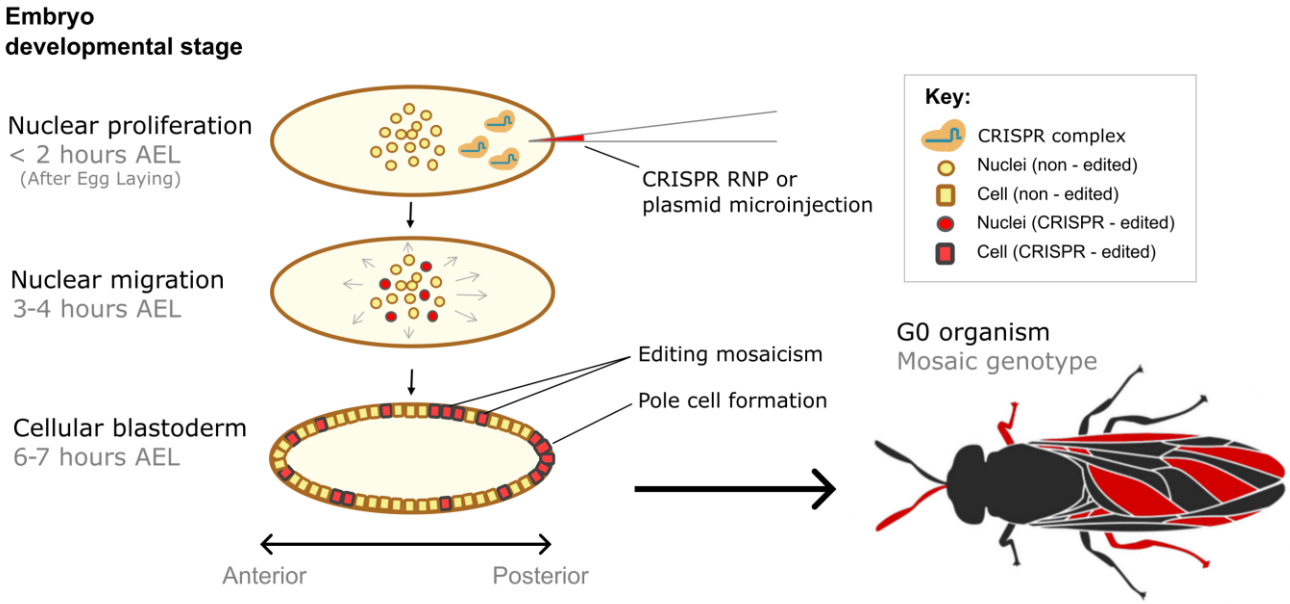


Figure 4: The embryonic microinjection technique for insect genetic transduction. The introduction of heritable gene edits or genome-inserted transgenes is mediated by direct embryo injection carried out in the initial stages of development. Eggs are injected at the posterior pole with an injection mix containing the CRISPR complex or plasmid transgenesis agent. In BSF, the first embryonic stage suitable for injection is expected to last for up to 2 hours after egg laying (AEL)[422]. In this syncytial blastoderm stage, the initial nuclei containing the insect's genome undergo a vast proliferation and expansion prior to migrating to the inner embryo walls for cellularisation. This cell layer forms the basis for the development of later tissue types, with cells located near the embryo posterior destined to become the primordial germ cells (PGCs) that give rise to the adult germline tissues and gametocytes. By injecting within the first few hours AEL, the mutagenic compounds are able to act freely on the proliferating nuclei in the syncytium, prior to their migration to the poles for cellularisation. Inevitably, the gene editing and transgenesis outcomes will be variable in the nuclear pool, leading to the development of a mosaic genotype in the organism that develops from the injected egg (generation 0, G0). However, targeting injection to the posterior pole enhances mutagenic exposure near the PGC region, increasing the likelihood that modified germ cells will be produced, which can establish a uniform mutant genotype in the offspring of the G0 generation (G1).

In the early stages of embryo development, the fertilised egg acts as a large single cell, with genomic DNA present in the yolk, or cytoplasm[423]. The egg contains cytoplasmic determinants, such as proteins and transcribed mRNAs, acquired through deposition from maternal and paternal gametes. These factors, such as the maternally deposited genes *nanos* and *vasa*, are spatially distributed across the cytoplasm in a way that coordinates early developmental patterning and establishing later cell localisation and differentiation[424,425].

Soon after being laid, the egg enters a stage of nuclear division, in which the diploid nucleus formed from the fusion of sperm and egg nuclei undergoes rapid doubling. At this stage the embryo is a syncytial blastoderm, with all nuclei contained in a common cytoplasm and not individually enclosed in cell membranes[426]. In embryonic microinjection, components such as Cas9-RNPs or *pBac* plasmid mixes are preferentially injected into the egg yolk at the syncytial blastoderm stage, where they can directly access and modify the proliferating pool of nuclei before they are sequestered into cells. With indel sequence formation and the site of *pBac* integration both being random processes, the outcomes of editing and transgenesis will likely be different across multiple edited nuclei, with many nuclei not changed at all[427].

Following proliferation, the nuclei migrate to the inner walls of the embryo, where they start to be encapsulated to form a layer of cells known as the blastoderm[423]. Depending on their location in the egg, these cells form the progenitors for different tissue types arising in the fully developed insect. In edited embryos, the disparity in genome sequence between the segregated nuclei means that the organism is now a genetic mosaic, with later tissues being endowed with non-identical genomes depending on the editing status of their progenitor cell[427]. The cells that form at the posterior pole, for example, become the germline progenitor cells that will eventually differentiate into the gametes used in adult reproductive tissues. In order to maximize the chance that the edit or transgene insertion is passed on to later generations, the posterior pole of the egg is targeted during injection, hopefully enriching the amount of edited nuclei in this region that end up establishing the germline progenitor cells[420].

Even when maximally efficient, the non-templated nature of NHEJ and the random integration behaviour of *pBac* transgenes means that the organism that develops from the injected embryo, or generation 0 (G0), will be a genetic mosaic[427]. To produce a fully transgenic organism that contains the identical mutation or transgene integration site across

all cells, the G0 organism needs to reproduce, passing on a single copy of their genome as a gamete[428]. If that gamete happens to encode a mutation or integrated transgene, the newly fertilised progeny or G1 organism should be fully heterozygous or hemizygous for the change. The edited sequence is also stably inherited in the G1 germline, meaning that it can be disseminated to later generations.

1.3.7 – Advances in Genetic Engineering in BSF

In its most simple form, the molecular toolkit for genetic engineering in BSF consists of the previously described CRISPR-Cas9, *pBac* transposon system and embryonic microinjection technologies. Using these, several efforts have already been made in the research and commercial space for editing towards valuable traits in the species. These have been enabled by considerable genetic resources in the form of multiple high-quality chromosome-scale genome assemblies. BSF has a genome of 1.01 gigabases and 16,478 genes spread out over 7 chromosomes, with the small 7th chromosome representing the sex chromosome[253,429]. Presenting as haploid in males, the sex chromosome is analogous to the X in X:Y sex determination systems, although this mechanism has been poorly characterised in BSF[430].

The first description of CRISPR-Cas9 related GE in the species was released alongside the first published genome assembly in 2020. In this, Zhan and colleagues employed embryonic microinjection of Cas9 RNPs to knock out two genes in BSF, with the intention of increasing larval yield in one case and simplifying the burden of colony maintenance in the other[253]. The *H. illucens* prothoracicotropic hormone gene (*HiPtth*), a neurohormone that regulates the timing of developmental transitions through stimulating the release of ecdysone from the prothoracic gland, was knocked out using two gRNA sequences to mediate a 889 bp deletion spanning multiple exons[431]. Surviving somatic G0 mutants showed a significant increase in prepupal weight compared to non-edited larva, extending the duration of the last feeding larval instar from 5 to over 85 days. With pupal transition effectively blocked, the considerable size of these mutants was explained due to their extended feeding period, and showed no real difference in growth rate when compared to the non-edited strain.

The researchers also produced a mutant for *vestigial*, a key regulatory gene for the development of wings in adult flies[253,432]. A large 11,460 bp deletion was made with the same techniques, with somatic mutant flies exhibiting wings that were clearly malformed or fully

absent. The given rationale was that flightless BSF would make colony maintenance a simpler process less prone to escapees, although the complex breeding behaviour in this species is widely understood to depend on the ability to fly during courtship rituals[433]. It should be noted that although the study demonstrated clear and strong somatic phenotypes for both mutants, these were each described as being the mosaic G0 generation. With the progression of the *HiPtth* mutants into the pupal and adult stages not being mentioned, it stands to reason that neither the *HiPtth* or *vestigial* mosaic phenotypes could be extended as uniform and hereditable mutant lines. Both candidate genes, however promising, are therefore unlikely to be suitable as beneficial BSF strains in the long term.

Since this first use, progress in Cas9-mediated GE in the species has been promising. In developmental research, the sex determination cascade member *intersex* was knocked out *via* Cas9-RNPs injection[430]. Regulating the female-specific developmental pathway in *Drosophila*, female mosaic mutant BSF showed clear abnormalities in genital and ovary development as adult flies, representing an early step in deciphering the sex determination pathway for the species[434].

Recently, the *H. illucens white* (*w*) and *yellow* pigmentation genes were also knocked out using similar methods, with stable inheritance of the homozygous mutant phenotypes also achieved in both cases through backcrossing[435,436]. Highly conserved, *w* encodes an ABC transporter subunit used for pigment transportation in the compound eyes of *Drosophila* and numerous other insect species[437]. Mutants for this gene have characteristically white eyes as adults compared with the typical *wild-type* red pigmentation, making *w* a common target used in genetic studies and for estimating CRISPR mutagenic efficiencies.

Interestingly, the *w* mutant phenotype showed high penetrance at the embryonic and larval stages as well as in adult flies, with the red eyespots apparent through the egg chorion during late embryo development appearing transparent in homozygous mutant BSF[435]. Another commonly pigmentation gene commonly target in CRISPR studies, *yellow* genes are hypothesised to function in dopamine synthesis and cuticle melanisation[438]. Through knocking out several genes in the *yellow* family of BSF, researchers were able to identify a *yellow-y* mutant with a light-brown colouration at the pupal and adult stage, easily distinguishable from the near-black *wild-type*[436].

In terms of yield traits, one promising mutant BSF phenotype *via* KO of *Lipid storage droplet-1* (*LSD1*) has been described[439]. A member of the perilipin family, *LSD1* regulates lipolysis and lipid accumulation, with *D. melanogaster* mutants showing an increase in lipid droplet fusion at the larval stage. In BSF, *LSD1* mutants showed a significantly increased body weight (10%) and more rapid larval development compared to the non-edited strain, as well as a lower FCR and higher bioconversion rate as detected in feeding trials[439]. This was accompanied by a general upregulation in the insulin signalling pathway, although palmitic and palmitoleic acid content was decreased. Homozygous mutants retained a high fitness after several generations, with *LSD1* and adjacent pathway members representing interesting candidate GOI for exploration in further studies.

Regarding transgenic technologies in BSF, the use of the *pBac* transposon system for random-site DNA integration is now fairly well characterised. Kou and colleagues first described the insertion of a transgenic GFP marker under the BSF-native *actin5C* promoter, with strong and global marker expression noted throughout the larval and adult life stages[422]. The authors also characterised the first instance of RNA interference (RNAi) technology in the species through the expression of short hairpin RNA (shRNA) against the GFP gene.

Three endogenous U6 promoter isolates, Pol III promoters commonly used for the expression of small RNA species, were cloned upstream of the shRNA gene in *pBac* plasmids intended for microinjection. Following their stable integration into the GFP-positive line, the relative strengths of the three promoters were compared *via* qPCR and the visible reduction in GFP fluorescence at the pupal stage. With each U6 promoter associated with varying levels of RNAi activity, this could suggest their potential for the expression of transgenically-encoded CRISPR gRNAs and crRNAs.

To explore the potential for modified BSF in human nutrition, a transgenic line was produced *via pBac* integration for the production of carotenoids[440]. Two fungal genes were integrated to facilitate the synthesis of α-carotene and β-carotene, with the goal of increasing the levels of fat-soluble vitamin A, which is often lacking in food-insecure diets. The inclusion of both genes led to a notable increase in carotenoid content, highlighting the potential for this insect as a metabolic engineering platform. The authors also characterised the synthetic constitutive promoter Hr5-IE1 for strong global transgenic expression of the DsRed fluorescent marker, while indicating a poor expression with the synthetic 3xP3 and *D. melanogaster* polyubiquitin (*PUb)* promoters. Finally, a third study characterised the pattern

of mosaic and transient marker expression seen during G0 screening and crossing, and described a protocol for identifying the site of *pBac* integration in the BSF genome[441].

Interestingly, much of the early development of more advanced mutant strains and CRISPR applications in BSF have emerged from work in the commercial sector, although hard data regarding these experiments is lacking outside of what can be gleaned from press releases and patent applications. In one example, the agritech company FreezeM claims to have engineered an enlarged BSF strain, dubbed “BSF-Titan”, with a 50% increase in larval mass, 25% decrease in FCR, and a significantly improved protein yield[442]. From patent disclosure, this appears to have been produced through Cas9-RNP directed KO of the BSF orthologue of *Semaphorin-Ia*, a gene involved in sensing fat sufficiency in *D. melanogaster,* with mutants continuing to gain weight and failing to undergo adult metamorphism[443,444]. BSFx2, a strain developed by biotech BugEra, has been described as having an oil and fatty acid twice as high as found in non-edited larva, with the company positioning this for use in the biodiesel and sustainable aviation fuel markets[445,446].

Interestingly, these companies and others have emphasised the idea of “strains as a service” in regard to edited BSF, indicating the potential for specialisation in producing improved genotypes on a bespoke basis, with larval properties being tuned to best fit one of several potential application areas[446]. Recently, several other companies have entered the scene with the expressed intention of producing transgenic strains to fit various different niche markets, with examples including the production of omega-3 enriched larva for use in aquafeeds, larva that encode appetite stimulating peptides for use in livestock feeds, and BSFL that express insulin, IGF1 and numerous other growth factors as a media supplement for use in the cultivated meat industry[447,448].

1.3.8 – Future Directions for Engineering BSF

Although the demonstrated attributes seen in the few known examples of engineered BSF strains are encouraging, these enhancements may barely scratch the surface of what could be realized with more efficient methods of GE and transgenesis in the species. Improvement in these technologies could enable the rapid assessment of multiple interesting GOI, identified from comprehensive genetic studies or literature descriptions of orthologous mutants in different species, for their potential in enhancing a number of desirable economic and fitness traits.

Reinforcing this, several commercial BSF producers have expressed an eagerness for genetic improvements leading to an increased larval biomass for maximising protein output, lower FCRs for boosting feeding efficiency, and a reduced larval development time to shorten the interval between larval harvests. In addition, the relative protein, lipid, chitin and mineral content in BSFL could be adjusted through genetic approaches to be more optimal for different use cases. This could be useful for producing animal feeds with tailored nutrient profiles for different livestock species, or for maximising yields of secondary metabolites for use in antimicrobials, cosmetic formulations and biofuels.

In particular, GE could be used for the manipulation of metabolic genes to direct energy flux towards the enhanced production of different molecules, with metabolic engineering strategies having been demonstrated to greatly increase desirable yields in microbial, plant and animal cell systems[449,450]. Furthermore, the reproductive output of BSF in terms of egg number, breeding activity and egg hatch rate represents an additional feature for genetic improvement, helping to increase economic performance in large scale rearing facilities.

In addition to yield traits, there is also an expressed interest in producing genotypes that are more tolerant to a range of rearing conditions, with the benefits of robust strains of BSF being felt in expensive rearing facilities and small-scale insect farms alike. This could include a predisposition towards better utilisation of certain low value waste types, such as manures from livestock with poor nutrition. An increased thermotolerance for either hot or cold climates could significantly reduce the economic, energy and GWP costs associated with maintaining an optimal rearing environment, and facilitate the outdoor rearing of BSF over a wider range of geographies. Indeed, selective breeding has already been employed to increase cold tolerance in BSF, with the authors reporting high larval weight and performance when reared at temperatures as low as 12°C[334].

Additionally, an increased resistance to desiccation could improve larval performance on drier wastes, and reduce the costs involved in maintaining a stable moisture content throughout the feeding process. Finally, the tuning of temperamental and behavioural traits, such as making the larval more voracious feeders or the adults display a higher breeding activity at low densities, could prove highly valuable in further boosting productivity.

However, efforts to enhance BSF are largely constrained by their genetics. Comprehensive data will be required for a deeper understanding of BSF development and metabolism, with many identified GOIs likely to produce unpredictable outcomes when mutated. The use of

CRISPR-directed indel formation, while practical for gene KO applications, is more limited for achieving subtle genome manipulations, such as adjusting or increasing the relative expression level of genes[451]. Inevitably, the physical and behavioural optimisations possible through targeted GE in BSF will be limited to fine-tuning expression of genes already present within their existing genomic architecture.

As such, several researchers and biotech companies have already started to validate BSF as a platform for transgenic expression of exogenous proteins. Looking ahead, one clear example for enhancement through transgenesis is in imparting an ability to synthesize essential fatty acids such as omega 3. With BSF unable to produce omega 3 endogenously, their use in aquaculture is significantly limited, and a strain engineered with an innate ability to produce and accumulate these oils can be expected to be highly valuable for the feed sector[452].

Without significant contribution from the microbiome, BSF is inadequate at digesting lignocellulose biomass, with this making up the majority of inaccessible plant waste[453]. Poor performance and bioconversion on several waste types has been associated with high lignocellulosic fibre content[454]. Waste pretreatment, such as grinding or boiling, can increase digestibility in these waste types, but these are high-cost interventions for a low-return revalorisation process. The value of BSF could therefore be greatly increased if able to partially break down lignocellulose natively, opening up a significant amount of plant waste for energy recovery.

Altering and enhancing the microbiome through supplementation is likely to be crucial in increasing the capacity for lignocellulose utilisation in this species, but an expanded scope for transgenic technology in BSF opens up an opportunity for encoding novel digestive and nutrient assimilation enzymes to aid with processing tough plant matter[455]. Candidates could be sourced from insects with an innate lignocellulose processing ability, such as termites, and reconstituted in BSF as a more robust and efficient platform for digesting diverse agricultural wastes[456].

Similarly, an increased capability for utilising plastics could be imparted to transgenic BSF[260]. Already demonstrating some innate plastic remediating potential, this could be enhanced with transgenes isolated from insects with stronger digestive abilities, such as waxworm, allowing for more complete breakdown of synthetic plastics like polyethylene[457]. By logical extension, transgenes could be imparted from a number of insect and microbial sources for

the engineering of BSF strains with an ability to better degrade different pollutants enriched in industrial wastes, such as POP and PAH species. As a proof of concept, researchers recently engineered a *D. melanogaster* strain with two *E. coli*-derived enzymes that allowed larva to degrade methylmercury, releasing elemental mercury to the atmosphere for easier collection[458]. Following up, the authors suggested reconstituting these enzymes in BSFL as a method for remediating municipal and fishery wastes with high methylmercury content[264].

1.3.9 – Research Gaps and CRISPR-Aided Breeding Strategies

As a new farmed species of high potential for sustainable waste processing, BSF stands to benefit from the development of improved genetics strategies to further optimize its desirable traits and characteristics. Several fundamental genetic techniques have been employed in the species in a short space of time, enabling the production of genetically-enhanced strains by the academic and commercial sectors. Crucially, however, the integration and analysis of the CRISPR toolkit components in transgenic BSF has not be described, with the GE of different targets currently depending on the embryonic microinjection of Cas9-RNPs.

Although this may be sufficient for a preliminary assessment of the function of candidate GOI, the mosaicism inherent with this method makes analysis difficult in the initial modified generation. Additionally, a small gain or loss in strain performance is likely to be masked in G0 insects by the impact of the microinjection technique, which can be highly disruptive and damaging to the fitness of the organism. Aside from this, microinjection is an expensive and laborious technique that lacks sufficient reproducibility in terms of volume of Cas9-RNP delivered, the specific developmental timing of injected embryos, and the overall efficiency of mutagenesis.

While some mutations can be fixed and inherited in later generations following the outcross of G0 mosaics, separating the edited from non-edited genotypes will be difficult to implement without sequencing if no evident phenotype is associated with the intended edit. Finally, the direct injection of the CRISPR components is likely to be insufficient for the reliable editing of multiple genes simultaneously, with multiplexed editing being highly valuable for examining the effect of several mutant GOI in tandem, or for inducing diversity across several loci in a *de novo* breeding pool.

Many of these issues could be addressed through the production of modular transgenic BSF strains, with the CRISPR nuclease and CRISPR RNA expression cassettes segregated

between two distinct lines. Individually, each line could be maintained without the presence of GE activity, with each component being inert in isolation. The CRISPR components could then be brought together by mating the two lines, with both transgenes then being inherited by a fraction of the progeny. With both elements present, a well-characterised and reproducible level of GE activity can occur in the transheterozygote generation (**Figure 5**).

With the high fecundity of BSFL, the number of the edited progeny can be easily scaled up by simply increasing number of flies and egg collections in the parental cross, allowing a large population for phenotyping than can be later replicated by repeating the cross. A modular source of an efficient nuclease can be retained as a transgenic line for crossing to numerous RNA-expressing lines produced to test different targets. Finally, activating GE through a CRISPR crossing strategy can help deconvolute the outcomes of multiplexed editing, with variable editing outcomes across multiple loci creating a diverse pool for the selection of improved genotypes by phenotyping.

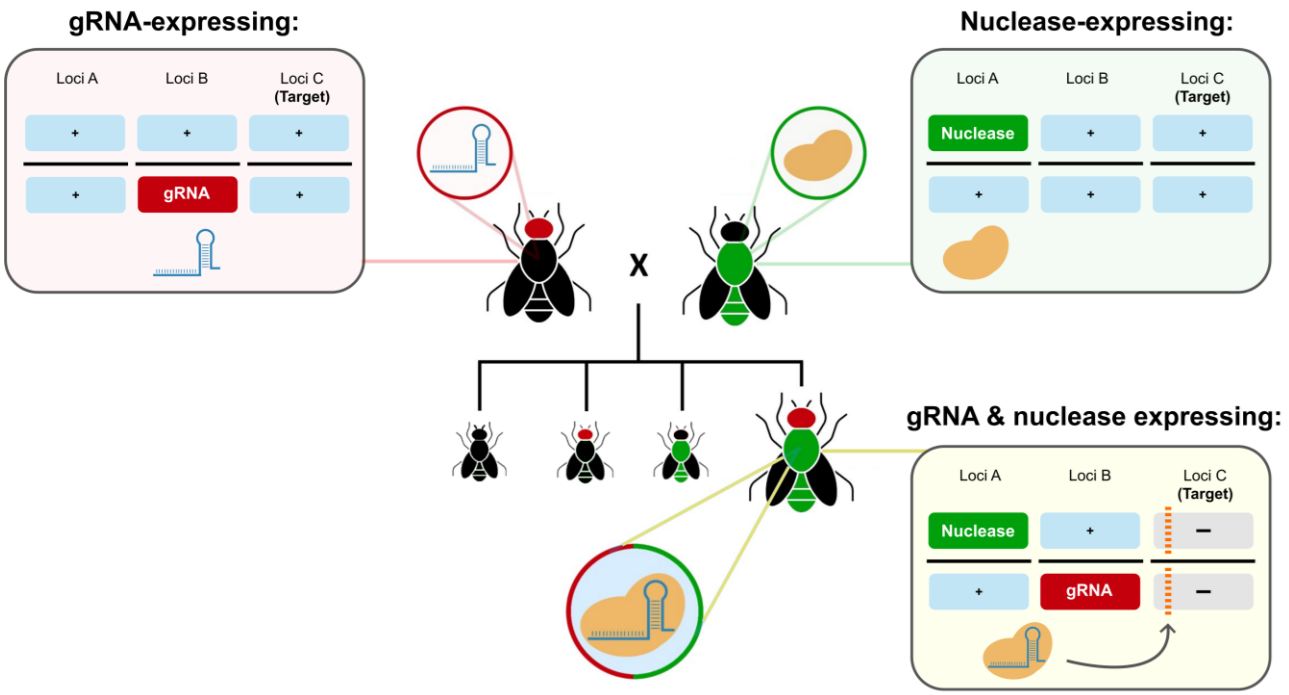


Figure 5: A transgenic breeding approach for BSF to enable efficient gene editing activity in the transheterozygote offspring of CRISPR-encoding parental lines. In this strategy, transgenic lines encoding either the CRISPR nuclease or gRNA components can be maintained in perpetuity with minimal detriment to fitness, with an absence of GE activity due to segregation of the CRISPR components. Upon genetic cross of the two lines, a proportion of the progeny will inherit both transgenes, identifiable *via* the expression of different fluorescent marker genes. Encoding both components, the transheterozygote progeny should be susceptible to GE activity at the target sites designated by the CRISPR-RNA component. Depending on the elements used for the regulatory control of the transgenes, an endogenous source of CRISPR expression could lead to a high standardised dose of the GE components in the early embryo stage, leading to a lower overall genetic mosaicism, as well as in the adult germline tissues involved in the production of gametes. By removing the dose variability and damage linked to direct microinjection-based methods of GE, the CRISPR-cross strategy should alleviate the influence of injection-related fitness costs on the observed transheterozygote phenotype. Additionally, the high repeatability and scalability associated with this method could prove to be a significant advantage when employed in detailed phenotypic analyses that require a large pool of heterogeneous mutants.

1.4 – Aims and Objectives

1.4.1 – Project Aims and Summary

The primary aim of this project was to develop and characterise an advanced molecular toolkit in the black soldier fly, showcasing the potential benefits of an endogenous source of CRISPR for enabling genetic enhancement studies in this species. The toolkit is centred around both the classical CRISPR nuclease Cas9 and the novel multiplex-capable nuclease MAD7. Both nucleases, along with their associated CRISPR RNA components, were placed under the control of a selection of candidate regulatory elements from endogenous, exogenous and synthetic sources, most of which have yet to be tested in this species.

To maximise GE activity in germline tissues while minimizing nuclease-associated toxicity and off-target mutations, the nucleases were controlled by two germline-specific promoters, isolated from germline-specific genes *nanos* and *vasa*, as well as the constitutive *Act5C* promoter[459,460]. CRISPR RNA expression was regulated by two endogenous *U6* promoters, in addition to a *D. melanogaster U6* promoter frequently employed in insect CRISPR studies[385,387]. Several candidate BSF genes were selected for mutation to allow for the straightforward tracking of editing efficiency when assessing the CRISPR toolkit components, along with multiple GOI that could serve as useful targets to explore in future genetic enhancements of BSF yield traits.

Following preliminary assessment in insect cell culture, the developed toolkit was used for *pBac*-mediated transgenesis in BSF *via* embryonic microinjection means of genetic transduction. With previous methods of embryonic microinjection in this insect being poorly described, and *pBac*-mediated transgenesis techniques having not been previously demonstrated in the species at start of this project, the development and optimisation of a protocol for the efficient establishment of transgenic BSF would also be necessary for progression in this study. Following this, the CRISPR mutagenic cross strategy outlined earlier in this project was performed and analysed for a number of combinations of transgenic lines. Finally, additional experiments were conducted to assess the potential and possible limitations of GE approaches in BSF, including the mutagenic cross strategy and other applications of the endogenous CRISPR toolkit.

1.4.2 – Research Questions and Objectives

This project was developed to address the following research questions:

1. What genetic elements are required for efficient BSF transgenesis?
2. Do BSF endogenous regulatory components support CRISPR gene editing activity?
3. Can the CRISPR cross strategy be used to achieve high levels of gene editing in transheterozygote progeny?

The experimental strategy in this study was designed to achieve the following objectives:

1. Identify suitable target genes in BSF for mutagenesis, along with endogenous regulatory elements to control CRISPR nuclease and RNA expression.
2. Establish and maintain a functional BSF insect colony.
3. Describe and optimize a reliable and efficient method for BSF transgenesis.
4. Produce and characterise transgenic BSF lines encoding the CRISPR toolkit.
5. Assess the gene editing potential of CRISPR toolkit components, both individually and in combination through CRISPR breeding crosses.

Chapter 2: **Materials and Methods**

2.1 – Animals and Husbandry

2.1.1 – Animal Strains

The Black soldier fly colony established at Imperial was founded from a population of ~3,500 pupae provided by Beta Bugs Ltd, hereby named “T4 Jan 3” or simply “T4”. The emergent T4 flies were inbred to produce a genetically homogenous stock, with the lineage treated as “Wild Type”, or “WT”, in all later experiments, as well as the parental genotype of all transgenic flies later produced.

2.1.2 – Black Soldier Fly Diets

The standard “Chick Crumb” BSF larval diet, developed by Beta Bugs Ltd, consisted of 1% sucrose (Thermo Fisher Scientific, Waltham, MA, US), 34% Mr Johnson’s Chick Crumb (Wilko, UK), and 65% tap water (w/w), which was combined and mixed until the sucrose was dissolved and the water fully incorporated. The enriched agar diet employed in this study was based off of a common diet used for rearing *D. melanogaster*, and consisted of 0.8% agar (w/v), 2% polenta (w/v), 10% brewers yeast (w/v), 8% fructose (w/v), 0.8% propanoic acid (v/v) and 0.5% methylparaben (v/v)[461]. Agar diet was prepared by combining and boiling the above mixture before pouring a 2-3 cm plug of liquid mix into 25 x 95 mm polystyrene *Drosophila* Vials (VWR, Radnor, PA, US) and allowing the agar to cool and solidify. Vials containing the solid diet were capped with cellulose acetate flugs (Scientific Laboratory Supplies, Nottingham, UK) and stored at 4°C for up to one month, and were brought up to room temperature before use.

2.1.3 – Black Soldier Fly Breeding and Maintenance

The BSF maintenance program employed in this study was adapted from an outline originally developed to support a minimal rearing setup by Beta Bugs Ltd. BSF was raised at 28.5°C +/- 1°C and 70% +/- 5% RH. BSF-breeding lights (Evoconysys, China) with a UVA

LED component were positioned 50 cm above breeding cages and the power tuned so that the light at the base of the cage read 9,000 Lux. The light operated on a 12:12 light dark cycle. BSF adults were raised in 30x30x30 cm mesh cages (Bugdorm, UK) at a density of 100-250 adult flies in a 1:1 male : female ratio. Flies were supplied with distilled water via a pot with a sponge-cloth wick, and the cage was misted with distilled water twice a day during the breeding and egg laying periods.

A pot of attractant, consisting of dried larval frass and pupal casings mixed with water, was included in the cage to stimulate egg laying near a decaying food source. A plastic egg collection device termed an “eggy” (Beta Bugs Ltd) or piled sheets of corrugated cardboard were placed on top of the attractant pot, for females to lay egg clutches within their crevices. Eggies were replaced every day and cleaned of egg clutches.

Egg clutches were gently removed from eggies with a scraping device and placed in a capped egg pot 1 inch in diameter (Amazon, UK). Eggs were incubated in the dark at 25-27°C for 3-4 days, and checked for larval emergence using a stereo microscope or 10X hand lens. Hatching clutches were immediately transferred to 20 g of standard chick crumb diet in 180 mL feeding pots. Pots were covered with topcloth and incubated in insectary conditions, with food being rehydrated and mixed each day until larvae were ~5 mm long and the substrate turned a dark brown colour, indicating spent food. Larvae were then transferred to 2 L meal pots and fed with 800 g of larval food.

Feeding larvae were sprayed with tap water and food was mixed every 2-3 days, with 800 g of larval food added when the second meal turned a dark brown. Two plastic larval escape ramps were placed in the 2 L pot, which was then placed uncovered in a large 5 L container topped with topcloth. Wandering prepupae were allowed to exit the food via the ramps and drop into the dry external container to begin pupation. The food in the 2 L pot was sprayed with water every 2-3 days to encourage prepupal movement to the dry external container to pupate. After the majority of larvae had turned to prepupae, the remaining food/frass was sieved using a bucket with 2 mm holes, and dissolved with water to liberate larvae and prepupae left in the food. All larvae and prepupae were washed in water and dried before being put in a container and left in the dark for two weeks to pupate before adult emergence.

Adults were kept in the dark to delay the onset of breeding for 2 days before being transferred to the cage. Adult flies were handled by anesthetising with CO_2. Adult breeding began 2 days after emergence, with egg laying commencing 2 days after breeding. Adults

hatched asynchronously, and were continuously added to cages to promote continued egg output. Breeding cages were generally productive for 7 days after the commencement of egg laying. After this, the eggy, attractant and water pots were removed from the cage and the cage supports gently disassembled so that the flies remained in the mesh cage, which was transported to a -20°C freezer for 24 hours to euthanise the flies. All eggs, larvae and pupae were also euthanised by freezing. All equipment was sterilised by immersion with Virkon (Thermo Fisher Scientific) and later cleaned with mild detergent.

2.1.4 – Animal Microscopy and Imaging

Transgenic BSF were isolated for crossing by screening for fluorescence marker expression at low magnification. Adult BSF were anesthetised under a CO_2 pad until fully immobilised before transferring to a petri dish by the wing with a pair of tapered fine point tweezers (Thermo Fisher Scientific). Larva were more resistant to anesthetisation by CO_2 and were instead screened by transferring them to a high-walled glass dish placed on ice until stunned. Insect microscopy was carried out using a MVX10 stereo fluorescence microscope (Evident Scientific, Waltham, MA, US) fitted with a pE-300^{lite} LED illumination system (CoolLED, Andover, UK). The fluorescent markers eGFP and DsRed was visualised, respectively, using the U-M49002XL and U-M49004XL filter cubes (Evident Scientific). Imaging was performed through a mounted DP23 digital microscope camera connected to CellSens Standard imaging software (Evident Scientific). BSF tissue dissections and embryo alignment for injection were performed using a SZX7 stereo microscope (Evident Scientific).

2.2 – Black Soldier Fly Microinjection and Transgenesis

2.2.1 – Microinjection Slide Preparation

Injection slides were prepared at least 12 hours before injection to allow the adhesive to dry. 22 x 50 mm (0.13 – 0.16 mm thick) borosilicate glass cover slips (VWR) were scored down the centre with a diamond glass cutter tool (8 mM, Draper tools, Hampshire, UK), and then snapped into two halves of ~11 x 50 mm. One half was stacked on top of the other at a slight offset and glued together with a drop of quick-drying nail polish, leaving a ~2 mm narrow overhang of the bottom cover slip exposed. Once dried, the glued coverslips were loosely adhered, overhang facing upwards, to a transparent glass microscope slide (VWR) with a drop of inert Halocarbon oil 700 (Sigma Aldrich, Dorset, UK).

2.2.2 - Microinjection Mix Preparation

Injection mixes were prepared and stored differently depending on the mix constituents. A concentrated stock of microinjection buffer was prepared in advance and added for a working mix of 0.1 mM $NaPO_4$, pH 7.2, 5 mM (DNA-only mixes) or 300 mM (Cas9-gRNA RNP mixes) KCl, and 0.05% phenol red to aid in visualisation.

For plasmid DNA injections, such as *piggyBac* construct and helper plasmids, 15 – 30 µL of total injection mix was prepared using nuclease-free water, plasmid DNA and 10X BSF microinjection buffer (1.0 mM $NaPO_4$, pH 7.2, 50 mM KCl, 0.5% phenol red solution). Total DNA concentration was set at 1.0 µg / µL, with a 600 ng : 400 ng ratio of *piggyBac* donor : *piggyBac* helper construct proving effective for producing transient marker expression following injection. Plasmid-based injection mixes were prepared at room temperature and filtered through cellulose acetate spin columns with a 0.22 µm pore size (Corning Costar Spin-X, Sigma-Aldrich) to remove large particles and precipitates that could obstruct the microinjection needle. Filtered plasmid mixes were used immediately for embryonic microinjection, or stored at -20°C for up to two weeks. Stored mixes were fully defrosted, briefly mixed by flicking and centrifuged for 5 minutes at maximum speed to sediment any precipitates before loading into injection needles.

Cas9-gRNA RNP complex injection mixes were prepared fresh on the day of injection, entirely on ice and in a RNAse-free environment. Aliquots of purified *Streptococcus pyogenes* Cas9 protein (NLS-Cas9-NLS Nuclease, GenScript, Piscataway, NJ, US) and gRNA stocks produced in-house by *in vitro* transcription were defrosted on ice and included in a 15 – 20 µL injection mix at final concentrations of 700 ng / µL and 500 ng / µL respectively. In order, nuclease-free water, 10X BSF microinjection buffer, 5X KCl supplementary buffer (1.5 M KCl, 0.22 µm-filtered), gRNA and Cas9 protein were added in an RNAse-free PCR tube on ice, with mixing by pipetting performed after each component was added. The prepared mixes were incubated at 25°C for 1 minute, then 37°C for 10 minutes to allow formation of Cas9-gRNA RNP complexes before returning to ice. Optionally, the RNP mix was filtered through a 0.22 µm spin column to reduce clogging in the injection needle. RNP injection mixes were kept on ice up until needle loading, and injections were carried out as soon as possible.

2.2.3 - Microinjection Needle Preparation

Injection needles were fabricated from quartz capillaries using a needle puller. For high survival following plasmid injection, needles required a long, gradual taper, minimal shoulder and small, fine tip. Needles were pulled to the below specification using the following equipment:

Capillaries: GQF100-70-10, World Precision Instruments (Quartz with Filament, outer diameter: 1.0 mm, inner diameter: 0.7 mm, length: 10 cm)

Needle puller: Sutter P-2000

Pulling program:

Heat **(HEAT)**	Filament **(FIL)**	Velocity **(VEL)**	Delay **(DEL)**	Pull **(PUL)**
900	5	65	150	250

Tip size was further decreased by reducing the Delay (DEL) value, or made larger by increasing DEL as needed. Needles were suitable for injection immediately through breaking the tip open by scraping against the side of a glass coverslip. Freshly-pulled needles were

stored in a dust-free, sealed container with fragile tips angled away from potential surface contact, and used for injection within one week of fabrication.

Injection needles were backloaded immediately prior to microinjection. Using a P2 pipette set to between 1.0 – 2.0 µL and a 0.5 - 20.0 µL Microloader pipette tip (Eppendorf), freshly prepared injection mix was aspirated by pressing to the second pipette stop and releasing slowly until the mix was seen to fill the thinnest part of the pipette tip. Pipetting was paused to allow the movement of liquid to stop, then the tip was removed from the injection mix tube and carefully inserted into the back of an injection needle capillary.

Holding the pipette and needle vertically, and with the pipette tip pushed to the bottom of the needle, the injection mix was slowly dispensed into the capillary, taking care to avoid formation of large air bubbles. Once filled, the capillary was firmly flicked at the bottom, avoiding the needle tip, until the small air bubble between the tip and injection mix floated to the top. The needle was then mounted to the needle holder and injection pressure source and high backpressure of 2700-3000 hPa was applied.

2.2.4 – Embryo Collection for Microinjection

At least 4 days prior to microinjection, a dedicated BSF breeding cage was initiated with a clean eggy, a freshly prepared attractant pot, and between 150 – 250 large flies aged within 2 days of pupal emergence at an equal sex ratio. The cage netting was misted at least twice daily with a spray bottle of room temperature tap water to allow the flies to drink. If set up correctly, high mating activity was evident within 2 days following cage initiation. Flies were added over several days post initiation to increase cage density.

Egg laying was observed from between 2 - 4 days following cage set up. Once the first laid clutches were detected, the eggy was collected to determine embryo fertilisation status, and replaced with a clean eggy. Cages were deemed suitable for injection as soon as laying was encountered, with sufficient egg output being reliable over the following 3 days. Large egg clutches typically indicated suitability for injection, with embryo pole retraction being apparent in young (<4 hr AEL) eggs confirming viability. Dead flies were removed from the cage floor each day to discourage egg laying outside of the eggy. Newly emerged flies were supplemented into the cage as dead flies were removed, extending the period of viable egg

production to around 1 week. After this, a new cage was initiated for injection with young flies of a similar age.

On the day of injection, the eggy was exchanged for a clean device, the cage thoroughly misted with water, and a timer started for egg collection. Cage misting resulted in an immediate cessation of flight activity and prompted flies to drink from the cage netting, delaying any egg laying activity for at least 20-30 minutes. Periodic cage misting was used to synchronising egg collection timings. The injection cage was checked every 30 – 60 minutes for freshly laid egg clutches or active laying into the eggy, with the timer being restarted following each check. If eggs were present, the flies were gently removed from the eggy and the device exchanged with a fresh one, after which the timer was restarted. When checking for signs of laying at frequent intervals, such as 30 minutes, clutches were usually detected while they were still being laid. In this case, eggs could be collected after laying ceased, noting the clutch age at the point of collection. Alternatively, sufficiently large clutches (~600 eggs or more) were collected before laying ceased by gently removing the laying female from the device.

Once collected, the eggy was disassembled, taking care not to disturb the egg clutches. Large clutches were harvested from the eggy using a sharp knife into an egg pot, sealing the pot immediately to retain moisture. The pots were labelled with the parental genotype, the date and specific time of collection, and a unique ID number (i.e. “1” for the first clutch collected) to distinguish between collection pots used for injections on the same day. Injection slides were similarly labelled for comparing injection survival to the specific collection used. Between egg transfers, collection pots were tightly sealed and stored at ambient or injection room temperatures. Pots were shielded from light wherever possible, or placed in an opaque container. Multiple clutches were combined into a single pot, provided they all appeared likely to be fertilised and were collected at the same time. After harvest, the plates of the eggy were thoroughly wiped with paper towel to completely remove traces of eggs, and the device reassembled and exchanged at the next egg collection timepoint.

2.2.5 – Embryo Alignment for Microinjection

While waiting for the first egg collection for injection, the egg alignment and microinjection stations were prepared. For embryo alignment, a SZX7 stereo microscope (Evident Scientific), pre-prepared glass injection slides, and an embryo manipulation tool were used. Using a sterile hypodermic needle (BD Microlance 3 30G, Thermo Fisher Scientific) mounted on a 1 mL syringe (Thermo Fisher Scientific) as an alignment tool, a small portion (~100 eggs) of the clutch was separated and transferred to the coverslip surface of a prepared injection slide. Individual eggs were gently transferred into the coverslip overlap so that they were aligned vertically and in a flat plane on the surface of the bottom cover slip, being pushed up flatly against the side of the top coverslip. It was not necessary to align eggs in the same anterior-posterior orientation before injection. For maximal egg alignment per slide, eggs were aligned head to toe with as small a separation as possible, or with sufficient space if unpractised in alignment. Eggs with evident damage, leakage or that had been touched by a sharp point were not transferred. Traces of nonaligned embryos from the top coverslip were removed or burst using the alignment tool and paper towel.

Alignment proceeded on one slide until the entire coverslip overlap was filled with eggs (between 40-50 eggs per slide depending on egg size and spacing). The top coverslip surface was labelled with a slide ID, the date, number of aligned eggs, injection mix components and egg clutch ID. The slide was either used for injection immediately or placed in a dark container to continue alignment. Depending on alignment speed, up to 5 slides were prepared before proceeding with injection, provided that the embryos were not exposed to non-insectary temperature and humidity for longer than 2 hours in total.

2.2.6 – Embryonic Microinjection Procedure

Embryonic microinjections were performed using a DM IL LED inverted microscope (Leica Microsystems, Wetzlar, DE) with a mechanical stage for manipulating the injection slide and a mounted Flexacam C3 12 MP microscope camera (Leica Microsystems). Microinjections were carried out using a FemtoJet 4i electronic microinjector (Eppendorf, Hamburg, DE) as a pressure supply. Microinjection needle movement was controlled with a TransferMan 4r micromanipulator (Eppendorf). All microinjection equipment was affixed to a STable Active Pneumatic Vibration Isolation Table (Supertech Instruments, London, UK) fitted with a Quiet

Air Compressor system (Supertech Instruments) to reduce lateral needle movement during the injection process.

A slide of aligned embryos was transferred to the inverted microscope, and the microscope stage adjusted so that the embryo at the top of the line was brought into focus with the egg yolk and chorion compartments clearly outlined. Egg viability was then assessed and the posterior pole identified by the appearance of the yolk at both egg poles and their relative degree of retraction from the chorion. When ready to inject, the tip of the mounted injection needle was moved to the region above the first embryo and slowly adjusted until the needle tip was to the side of the egg and in focus with the egg surface.

If required, needles were opened by touching the tip against the glass coverslip. With back pressure applied, the stage was carefully moved so that the needle tip was just touching the side of the top coverslip. The stage or needle manipulator was then moved so that the tip was scraped along the coverslip surface, until drops of liquid were seen forming from the needle tip. The needle was then moved away from the coverslip and the “clean” button on the pressure source was pressed to increase pressure in the needle and expel more liquid and additional air bubbles. If opened correctly, a slow but continuous stream of droplets would appear at the needle surface after purging, with the speed of ejection further adjusted by changing the back pressure.

With an opened needle, the microscope stage was moved to inject the egg at the posterior yolk, injecting laterally where the vitelline space was thinnest. A disturbance in the yolk or entry of the phenol red dye was visible in the egg, indicating injection of mix. The needle was removed from the egg after around half a second, or after mix was seen entering in the egg. Injection was continued down the line of embryos, identifying and injecting the posterior pole in each case. The “clean” button was pressed periodically to eject cytosol backflow from the needle tip and keep the correct flow. Needles often needed periodic reopening for consistent flow. After a full line of eggs was injected, the egg slide was transferred to a dark container, such as a slide holder, in insectary environmental conditions as soon as possible.

2.2.7 – Post-Injection Rearing of G0 Larvae

Following injection, the G0 embryos on the injection slides were incubated in insectary conditions for 3 - 4 days until hatching. Around 72 hours post injection, the eggs were checked for viability by confirming the development of eye spots, and by identifying transient fluorescent marker expression. Prior to hatching, the glued coverslip assembly was gently detached from the glass injection slide, without disturbing the egg line. The coverslip was transferred to a vial of agar diet so that the coverslips were stood vertically above the feed layer. Up to three coverslip assemblies were placed in a single vial, with the egg lines facing the vial walls to avoid damage from coverslips contacting each other.

At around day 6 post injection, the coverslips were removed from the vials and injection survival rate determined by counting the number of unhatched embryos and number of gaps or egg chorion remnants remaining on the coverslip. If eggs were aligned consistently, successful hatching could be identified by the size of the gaps between non-hatching eggs, even if chorion shells were not detected. Hatched G0 neonates were incubated in the agar vials, and grew large enough to be visible without magnification from around 4 days post hatching. Larval transient marker expression was also often observed at this stage. Small densities of G0 larva were reared in the agar vials until the pupal stage, adding more feed as needed. In injection sets with high numbers of surviving G0 larvae (>150), larvae were transferred from the vials to the standard chick crumb diet at 7 days post hatching.

2.2.8 – G0 Cross for Transgenic Line Establishment

Surviving G0 pupa were recovered from spent feed and placed in a pupal container. Upon emergence, adult G0 flies were sexed and transient marker expression observed. A cage for a G0 self-cross was then established containing G0 of both sexes, with wild-type flies supplemented if required to reach a high breeding density. With high numbers of emerging G0 adults, reciprocal G0 cages were set up for each sex and crossed to wild-type flies of the opposing sex. G0 cages were maintained as normal, with regular spraying of water and addition of new G0 adults as they emerged. All egg clutches were collected in their entirety using a brush into individual labelled egg containers.

Egg pots were screened for embryo viability and marker expression, depending on the expected developmental stage of transgenic marker expression. For hatching egg pots

identified as containing transgenic G1 neonates, all marker-expressing neonates were transferred into a vial of agar-feed for rearing. For significant numbers of transgenics, entire egg pots were reared straight away on the standard diet and the transgenic larva isolated when they had grown larger.

Once G1 transgenics were reared to adulthood, reciprocal G1 crosses for each sex were set up with wild-type flies of the opposite sex. The G1 cages were reared as standard and all clutches collected, taking care to identify and collect eggs from clutches laid by a single female to reduce cross contamination with eggs harbouring transgenes at different genomic integration sites. When collecting from single females, each egg clutch was assumed to contain G2 transgenics with an identical site of transgene integration, and were reared as an independent isoline. Transgenics from G2 isolines were then self-crossed to generate homozygous G3, or back crossed with wild-type flies to generate heterozygous isolines.

2.3 – Molecular Biology

2.3.1 - Bacterial Transformation and Culture

Plasmid transformation was performed using the 5-alpha Competent *E. coli* strain (New England Biolabs, NEB, Ipswich, MA, US). Liquid cultures were performed using lysogeny broth (LB) or terrific broth (TB) supplemented with 100 mg / mL ampicillin, 50 µg / mL kanamycin or 25 µg / mL chloramphenicol (Thermo Fisher Scientific), depending on the resistance gene present in the plasmid. Plasmid DNA was recovered from overnight liquid cultures using the QIAprep Spin Miniprep kit (Qiagen, Hilden, DE) or the ZymoPURE II Plasmid Midiprep kit (Zymo Research, Irving, CA, US), depending on the scale required. Bacterial glycerol stocks were prepared by mixing 500 µl of overnight culture containing antibiotics with 500 µl of sterile 50% glycerol (Thermo Fisher Scientific) and freezing in 1 mL cryovials at -80°C.

2.3.2 – Genomic DNA Extraction

Genomic DNA (gDNA) was extracted from BSF tissues using two principal methods, with no clear difference observed in downstream PCR performance when high-quality extractions were used. An organic-phase extraction utilising phenol-chloroform-isoamyl was adapted for BSF tissues from a previously established protocol for gDNA purification in *Ceratitis capitata*[462]*.* Briefly, BSF tissues were immersed and homogenised in high-salt, SDS-based gDNA extraction buffer-1 (7 M urea, 10 mM EDTA, 100 mM Tris-HCl, 350 mM NaCl, 2% SDS) using a sterile pestle (Monarch, NEB). Insoluble debris was pelleted by centrifugation at 13,000 x g for 8 minutes, and the clarified supernatant transferred to a new tube. An equal volume of phenol-chloroform-isoamyl (25:24:1, Thermo Fisher Scientific) was added to the sample and mixed thoroughly by vortexing before phase separation by centrifugation at 13,000 x g for 10 minutes. The upper aqueous phase containing DNA was carefully transferred into a fresh tube, avoiding carryover of the protein-containing organic phase.

Depending on the degree of pigmentation in the recovered aqueous phase, 1-2 additional rounds of phenol-chloroform-isoamyl phase separation were performed, followed by a chloroform-isoamyl (24:1, Sigma-Aldrich) extraction. gDNA in the final recovered aqueous phase was then precipitated by incubation with 0.1 volumes of 3 M ammonium acetate (Sigma-Aldrich) and 2.5 volumes of molecular-grade ethanol (≥99.8%, Sigma-Aldrich). The precipitate was pelleted by centrifugation at 13,000 x g for 30 minutes at 4°C. DNA pellets were washed twice with 80% ethanol, air dried for 5 minutes and resuspended in 30–100 µL of nuclease-free water before measuring DNA concentration and quantity by Nanodrop.

A non-organic gDNA extraction method was also used, significantly adapted from a single fly *D. melanogaster* DNA preparation method to enhance protocol safety and practicality[463]. Although it resulted in higher RNA and carbohydrate carryover, leading to lower purity ratios on the Nanodrop, it provided more consistent gDNA yields without organic solvent contamination, and the extracts performed well in downstream PCR applications.

Briefly, BSF tissues were immersed and homogenized in gDNA extraction buffer-2 (100 mM Tris-HCl, pH 7.6, 100 mM EDTA, 100 mM NaCl, 0.5% SDS) using a sterile pestle (NEB). Homogenates were incubated at 65°C for 15 minutes with occasional vortexing to ensure sufficient cell lysis. An equal volume of 3 M potassium acetate (Sigma-Aldrich) was added to the homogenates, which were vortexed to mix and incubated on ice for 10 minutes. Samples were then centrifuged at 10,000 × g for 5 minutes to pellet insoluble materials, and

the clarified supernatant was transferred to a new tube. To precipitate gDNA, 0.6 volumes of isopropanol (Sigma-Aldrich) were added, with the samples then mixed by vortexing and incubated at room temperature for 5 minutes. The gDNA precipitate was pelleted by centrifugation at 12,000 × g for 10 minutes, and the pellet was washed twice with 1 mL of ice-cold 70% ethanol. Washed pellets were then air-dried, resuspended in 50–100 µL of water, and the DNA concentration and quality were measured by Nanodrop.

2.3.3 – PCR

PCR was employed for the purposes of cloning, screening and sequence analysis with a number of DNA polymerase reagents. Unless otherwise stated, PCR reactions were performed according to the manufacturers' recommendations. For sequence-sensitive applications, including generating defined fragments for plasmid assembly and for analysis by direct sequencing, either Phusion® High-Fidelity DNA Polymerase (New England Biolabs (NEB), Ipswich, MA, US), Phusion® High-Fidelity PCR Master Mix (NEB), or Q5® High-Fidelity DNA Polymerase (NEB) was employed.

50-150 ng of gDNA or 0.1–10 ng of plasmid DNA was included as template in 50 µL PCR reactions, and scaled appropriately with reaction volume. Bovine Serum Albumin (BSA) (Sigma-Aldrich) was included as a PCR additive at a final concentration range of 0.1-1.0 µg / mL when using *H. illucens* organic gDNA extractions as a reaction template. The online tool NEB Tm Calculator (tmcalculator.neb.com) was used to determine the recommended initial reaction annealing temperature (T_A) for each primer pair and polymerase combination.

Red Taq DNA Polymerase Master Mix (VWR International, Radnor, PA, US) was used for colony PCR screening. For colony PCR, single *E. coli* plated colonies were touched lightly at the edge with a pipette tip and the tip transferred to 50 µL of sterile water, with 5 µL of inoculated water being used as DNA template in 15 µL reactions. Platinum Direct PCR Universal Master Mix (Invitrogen, Waltham, MA, US) was used for genotyping of samples processed with Fast Extract DNA solution (VWR). As the DNA concentration of extracts produced by this method could not be determined, clarified lysate was included for template as 1/50th of the total reaction volume (1 µL template / 50 µL reaction).

For routine PCR, primers were designed in accordance with widely-accepted conventions[464]: 18-25 nucleotides (nt) in length, 40-60% total GC content, lacking single base repeats longer than 4 nt and including a 1-2 nt GC clamp at the 3' end. Melting

temperatures (T_M) for each primer set, as determined with the NEB Tm Calculator, were limited to be between 50 – 66°C and to within 5°C of each other. Gradient PCRs spanning a range of ± 4.0°C around the recommended PCR annealing temperature were used to optimise individual reactions when weak, absent or non-specific amplification was encountered.

For DNA assembly and Illumina sequencing, overhanging or adapter sequences were included at the flanks of the produced amplicons by following a 2-step thermocycling method according to the polymerase manufacturers' guidelines. 2-step PCR cycling employed long primers (28-54 nt), with the non-templated adapter sequence included at the 5' end and the target-complementary annealing sequence at the 3' end. Long primer pairs were additionally evaluated for their tendency towards hairpin structure, self-dimer and heterodimer formation with the online tool OligoAnalyzer (Integrated DNA technologies (IDT), Coralville, IA, US). PCR primers and DNA oligos used in cloning were ordered as 0.01-0.05 µmol lyophilized unmodified DNA oligos with salt-free purification from Eurofins Genomics (Ebersberg, DE), unless otherwise stated, and resuspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA).

2.3.4 – Agarose Gel Electrophoresis

DNA and RNA was visualized via gel electrophoresis. Gels were prepared with agarose (Sigma-Aldrich) and Tris-acetate-EDTA buffer (TAE, Thermo Fisher Scientific) to between 0.8 – 2.0 % (w/v) agarose concentration depending on the fragment size being visualised. For RNA degradation analysis, bleach (sodium hypochlorite, 6%, Thermo Fisher Scientific) was added to the gel solution at a final concentration of 1% (v/v) prior to agarose melting to inhibit RNAse activity. SYBR safe DNA gel stain (Thermo Fisher Scientific) was added for nucleic acid visualisation prior to gel casting.

When required, Gel Loading Dye, Purple (NEB) or TriTrack DNA Loading Dye (Thermo Fisher Scientific) was added to samples to aid sample loading and for tracking gel migration. 1 kb Plus DNA Ladder (NEB) was used as a molecular weight marker for all agarose gel experiments, unless other markers are directly mentioned. Gels were electrophoresed in TAE buffer by applying constant voltage across the gel tank between 80-105 V for between 30-60 minutes, with the run settings depending on the gel percentage, expected fragment size and degree of separation required. For RNA degradation analysis, gels and gel tanks

were precooled and electrophoresis performed at 80V for 60 minutes in a 4°C cold room to reduce RNA deterioration during the run.

Gels were imaged under UV using the ChemiDoc XRS+ imaging system and corresponding Image-Lab software (BioRad, Hercules, CA, US). Gel images were processed with colour inversion and brightness and contrast adjustments in Image Lab before exporting as TIF files for annotation.

2.3.5 – Gel Extraction and PCR Cleanup

Gel extraction was used to isolate PCR amplicons of a specific size for cloning or sequencing following separation by gel electrophoresis. Gel plugs containing the desired fragment were excised under a blue light transilluminator with a sterile scalpel, with excess gel being trimmed before weighing the sample. Samples were then processed by addition of 4 µL of Gel Dissolving buffer per mg of gel from the Monarch® DNA Gel Extraction Kit (NEB), and continuing with the associated extraction protocol. PCR cleanup without the gel separation step was achieved using the Monarch® PCR & DNA Cleanup Kit (NEB). With both kits, the centrifugation time following the final ethanol wash was increased from 1:00 minute to 1:30 to further dry the column membrane. DNA was eluted from the column by centrifugation following a one-minute incubation with TE buffer prewarmed to 55°C.

2.3.6 – Nucleic Acid Quantification and Sequencing

Following extraction and column elution, DNA and RNA samples were measured for quantity and relative purity using a Nanodrop One spectrophotometer (Thermo Fisher Scientific). dsDNA and RNA samples with measured A260/A280 values between 1.8-2.0 and 2.0-2.2, respectively, were considered relatively free from protein contamination. For both sample types, A260/A230 values >2.0 were considered free from carry over of inhibitory compounds used in the purification process, including ethanol, organic solvents and guanidine-HCl.

All DNA sequencing was performed by submission of samples to Genewiz (Takeley, UK), according to their stated sample submission guidelines. For clonal DNA samples, routine Sanger sequencing was performed on plasmid and PCR amplicon samples through the Genewiz Sanger Sequencing Service. Amplicon sequencing, or deep sequencing (2x250bp

Illumina, paired end) of heterogeneous amplicons generated with partial Illumina adapter sequences by PCR was performed through the Genewiz Amplicon-EZ Service.

2.3.7 – Plasmid DNA Assembly

All plasmid constructs were produced from DNA sequences that were isolated by PCR from plasmids, genomic DNA extracts, or synthetic gene fragments in the case of *de novo* sequences. Plasmid cloning was mainly performed by the production of specific PCR amplicons encoding ~20 bp of complementary overhanging sequence at the ends of the fragment that correspond to the overhang of the desired flanking fragments, directing the sequence order and orientation. In general, one or more PCR fragments were assembled into a restriction-linearised plasmid backbone with NEBuilder HiFi DNA Assembly Master Mix (NEB), optimising the input DNA concentrations according to the manufacturer's guidelines. Cloning of blunt-ended PCR amplicons into vectors for transformation was performed with the StrataClone Blunt PCR Cloning Kit (Agilent Technologies, Santa Clara, CA, US) according to the manufacturer's specifications. T7 endonuclease-I was used to perform restriction digestion on DNA amplicons during T7 endonuclease-I assays according to the manufacturers' suggested protocol (NEB).

Type IIS cloning was employed when producing the CRISPR RNA constructs through introduction of a 22 bp sequence featuring two inverted BsaI restriction sites encoded at the site of the gRNA protospacer. Briefly, insertion of a spacer sequence for a new target required digestion of the appropriate CRISPR RNA entry plasmid with BsaI-HF (NEB) and removal of the short BsaI cassette by PCR column cleanup. For each CRISPR target, complementary DNA oligonucleotides were designed to encode the spacer sequence as a double stranded DNA fragment when annealed together, with 4 bp 5' and 3' overhangs complementary to the ends of the BsaI-digested vector. The oligonucleotides were 5' phosphorylated with T4 polynucleotide kinase (NEB) and annealed together by boiling at 95°C followed by a slow decrease (5.0°C / min) in temperature to 25°C. The annealed oligo cassette was ligated in excess to the linearised vector using T4 DNA ligase (NEB), followed by plasmid transformation and verification of the correct protospacer insertion by Sanger sequencing.

For cloning long polycistronic RNA cassettes for assembly into crRNA-expressing plasmids, oligos were designed to anneal in the correct orientation, be ligated into a continuous dsDNA

fragment by T4 DNA ligase and then enriched by PCR. The PCR resulted in the addition of flanking BsaI sites to the cassette, requiring cleavage with BsaI-HF before ligation into the linearised entry crRNA vector.

2.3.8 – Construct Cloning Strategy

The plasmids 1142B, 1139A, DV-2.0, V1260 and PUb-GFP/PUb-DsRed, which were used for sourcing gene fragments for construct assembly in this study, were kindly provided by Dr. Angela Meccariello. The *Hr5-IE1* and *Opie2* synthetic promoter fragments were isolated from 1142B (*Hr5-IE1*_GFP) and 1139A (*Opie2*_mCherry) respectively. The 2.0 kb proximal *Actin5C* promoter and 2.0 kb *Polyubiquitin* promoter fragments from *D. melanogaster* were isolated from DV-2.0 (*Dm-Act5C*-dCas9) and V1260 (*Cc*-B2T-Cas9 / *Dm-PUb*-eGFP) respectively. V1260 also served as the source of the *D. melanogaster gypsy* insulator, the eGFP and DsRed markers, human codon-optimised *Streptococcus pyogenes* Cas9, the SV40 terminator, and the *piggyBac* 5' and 3' inverted terminal repeat-flanked plasmid backbone, which formed the base vector for all constructs generated in this study.

The MAD7 nuclease DNA sequence was obtained from Inscripta (Pleasanton, CA, US) and was codon-optimised for expression in *D. melanogaster* using the Codon Optimization Tool (IDT). To complement the modifications present in the Cas9 sequence used in this study, a 3x FLAG tag (DYKDHDGDYKDHDIDYKDDDDK), the SV40 Large T antigen NLS (PKKKRKV) and a short peptide linker (GIHGVPAA) was appended in-frame to the 5' end of the MAD7 coding sequence. The nucleoplasmin NLS sequence KRPAATKKAGQAKKKK was then appended at the 3' end of MAD7, followed by a TAA stop codon. The 3957 bp tagged *Dm*-MAD7 sequence was synthesised as two ~2.0 kb gene fragments (Eurofins Genomics), with each fragment sharing 20 bp of overlapping sequence to facilitate assembly into a single CDS fragment.

While assembling the final plasmid constructs, DNA elements were added stepwise to the *piggyBac* backbone derived from V1260 through a series of cloning intermediaries. First, the *gypsy* insulator element was cloned immediately downstream of the 5' *piggyBac* ITR. DNA assembly with overhanging primers was used to introduce dual attP sites, the recombination signal for phiC31:intergrase, downstream of the *gypsy* insulator and upstream of the 3' *piggyBac* ITR. The 3xP3-DsRed-SV40 and Hr5-IE1-eGFP-SV40 fluorescent marker cassettes were then assembled from their individual elements and separately cloned into

the vector between the attP flanks, producing two *piggyBac* construct lineages encoding different markers.

Approximately 2.0-2.5 kb of promoter and 1.2-1.5 kb of terminator sequence from the BSF *nanos* and *vasa* genes, including 5' and 3' UTR sequences up to the CDS start and stop codons, were isolated by PCR for DNA assembly. These sequences were amplified from a mixed pool of BSF gDNA extracted from male and female adult flies in equal proportions. Similarly, ~450 bp of the U6:1 and U6:2 promoter regions present immediately upstream of the identified U6 snRNA transcript were isolated from a pooled gDNA. The commonly used *D. melanogaster* U6:3 promoter was isolated from the plasmid 1142B.

The *piggyBac Hr5-IE1*-eGPP-SV40 intermediary plasmid was linearised by PacI (NEB) digestion and gel extracted. Using the NEBuilder HiFi DNA Assembly Master Mix, the common backbone was then assembled with different combinations of promoter, nuclease and terminator sequence, yielding each of the final nuclease constructs used in cell culture and BSF transgenesis in this study.

The Cas9-gRNA and MAD7 crRNA scaffold sequences were designed to enable the rapid introduction of new protospacer sequences by Type IIS cloning. Additionally, the crRNA insert was flanked by tRNA sequences at both ends. To promote transcription termination by Pol-III, a 31 bp poly-T terminator sequence found downstream of the BSF U6:2 snRNA was appended to the 3' end of the gRNA scaffold and after the 3' tRNA in the crRNA cassette. The modified CRISPR RNA cassettes were encoded in tandem on a single gBlock gene fragment (IDT), flanked with PacI and PvuI restriction sites at the 5' and 3' ends, respectively.

To construct the CRISPR RNA entry plasmids, the *piggyBac 3xP3*-DsRed-SV40 intermediary plasmid was linearised by with PacI, and the exposed ends dephosphorylated with calf intestinal alkaline phosphatase (NEB) to prevent vector re-ligation. Both gRNA and crRNA cassettes were separately amplified by PCR from the tandem gBlock, gel extracted and digested with PacI and PvuI (NEB). The digested cassettes and dephosphorylated backbone were gel extracted, and each cassette ligated into the vector using T4 DNA ligase. The *3xP3* promoter was later replaced in each CRISPR RNA construct with *Opie2* via DNA assembly, and Type IIS cloning was used to introduce unique spacer sequences to generate the final set of CRISPR targeting plasmids used in this study.

The plasmid *I-hyPBase*, kindly provided by Dr. Ernst Wimmer, encoded the insect-codon optimised version of the hyperactive *piggyBac* transposase variant under the *D.*

melanogaster heatshock protein 70ssb promoter and terminator, and was used directly as the *piggyBac* helper plasmid in all embryo microinjections.

2.3.9 – Guide RNA Synthesis

Specific gRNAs were produced by assembling DNA oligonucleotides into a dsDNA fragment to serve as a template in an *in vitro* RNA transcription reaction. Briefly, a set of complementary DNA oligonucleotides that together encoded a T7 polymerase promoter sequence, a desired gRNA spacer sequence and the gRNA scaffold sequence were annealed into a dsDNA fragment by boiling and slowly cooling the mix of oligos back down to room temperature. The annealed template was then amplified by PCR and recovered by PCR column cleanup. 100 ng of the PCR product was then used as template in a 20 µL *in vitro* transcription reaction with the TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific), following the manufacturer's recommended protocol for generating short RNA transcripts. After incubation, the concentration of the synthesised gRNA was analysed by Nanodrop and adjusted to 1600 ng / µl before 5 µL aliquots were prepared for storage at -80°C.

2.3.10 – Inverse and Splinkerette PCR

The inverse PCR (iPCR) technique was used establish the genomic loci of *piggyBac*-integrated transgenic constructs[465]. Briefly, 150–1000 ng of transgenic BSF gDNA extracts were digested with SauA3I (NEB) and HinP1I (NEB) in separate reactions for 2 hours to overnight at 37°C. The restriction enzymes were heat-inactivated at 60°C for 20 minutes, and the digested gDNA was self-ligated by adding 600 U of T4 DNA ligase (NEB), followed by incubation at 4°C for 16 hours. After ligation, a primary PCR was carried out using Phusion polymerase on 1:10, 1:100 and 1:000 dilutions of the gDNA ligation reactions. A primer set was used that specifically amplify the terminal section of the 5' *piggyBac* arm in SauA3I-digested samples, while a separate set amplified the 3' *piggyBac* arm in HinP1I digestions. The primers were designed to face outwards from each other in the *piggyBac* arms, ensuring successful amplification captured any adjacent genomic sequence at the integration site.

Following the first round of PCR, a 1:100 dilution of the 5' and 3' PCR products was used as the template for a second PCR that employed different sets of internal primers to enrich *piggyBac*-specific amplicons. The second PCR reaction was electrophoresed on an agarose gel, and strong DNA bands were purified by gel extraction. Recovered DNA from the iPCR bands corresponding to either the 5' and 3' genomic integration sites were sent for Sanger sequencing. The non-*piggyBac* aligned sequencing result was then used as input for a nucleotide BLAST search against the BSF genome to identify potential integration sites.

Following mixed success with integration site amplification and multiple attempts to optimise the iPCR steps, a variation of the technique was later employed based on a previously validated integration site identification protocol in BSF[466]. Splinkerette PCR differed from iPCR by utilising the ligation of oligonucleotide linker cassettes to the exposed gDNA ends following digestion, rather than gDNA self-ligation.

As in the iPCR protocol, gDNA was digested with SauA3I or HinP1I. Splinkerette oligonucleotide cassettes complementary to the SauA3I or HinP1I DNA overhangs were then added in excess to the respective digests to promote annealing of the splinkerette with the gDNA. DNA ligation and PCR was performed according to previously described iPCR procedures. The 5' and 3' splinkerette PCR rounds utilised a *piggyBac* arm-binding primer orientated towards the genomic DNA sequence and a primer specific to the terminally ligated splinkerette cassette, facing towards the *piggyBac* arm. A second round of PCR was performed using new sets of internal primers for both the 5' and 3' integration sites. Successful DNA amplicons visualised, recovered and sequenced as previously described.

2.3.11 – Tissue RNA Extraction

Total RNA was extracted for RNA-Seq and cDNA synthesis from *H. illucens* carcass, ovary, testis and embryo tissue isolates. For fly tissues, unmated flies were separated by sex upon emergence and reared in illuminated breeding cages for 2-3 days until sexual maturity. Flies were stunned by cold shock for 1 minute at -20°C, before being euthanized by decapitation and the remove of the wings using clean, RNAse-free dissection scissors (Thermo Fisher Scientific). Abdomens were dissected with two sterile hypodermic needles (BD Microlance 3 30G, Thermo Fisher Scientific) to liberate germline tissues, leaving the germline-free carcass tissue free for collection.

Germline tissues were washed briefly in ice-cold PBS before adding to a 1.5 mL tube and immediately snap-freezing on dry ice. 5 pairs of ovaries and 8 pairs of testes were pooled respectively in single tubes to provide sufficient material for extraction. The dissected carcass and head, but not wing, tissues were large enough for sampling without pooling individuals. For embryo tissues, egg clutches were collected and transferred to an RNAse-free Eppendorf tube within 30 minutes of laying and aged until the required sampling timepoint. A small sample of embryos were reserved and incubated separately to confirm clutch fertilisation and the presence of transgenics later. At the sampling time point, the tube was closed and snap-frozen in dry ice.

Total RNA was extracted from tissue samples with TRIzol Reagent (Invitrogen) by following the extraction procedure recommended by the reagent users guide with some adaptations. With all snap-frozen tissue samples, a volume of TRIzol (0.5 or 1.0 mL) was added to fully submerge the tissue within 8 hours of sampling. Tubes were placed back on dry ice for 10 minutes to allow the TRIzol to freeze, and then on wet ice to partially defrost. Once the defrosting TRIzol had the consistency of a thick slurry, the tissue was thoroughly homogenised with a sterile pestle (NEB) and incubated at room temperature for 10 minutes to allow dissociation of ribonucleic complexes. Samples were centrifuged for 5 minutes at 12,000 x g at 4°C and the clarified supernatant transferred to a fresh tube, at which point the protocol was completed according to the TRIzol reagent users guide.

After airdrying, RNA pellets were resuspended in 30 – 80 µL of room temperature water, depending on the size of the pellet. Secondary structures were denatured by incubating RNA samples at 60°C for 10 minutes before placing on ice for analysis *via* Nanodrop. Aliquots of 1 µg of total RNA were reserved from each extraction for examining RNA deterioration with gel electrophoresis. Extractions that yielded two clear bands corresponding to the denatured rRNA subunits with minimal streaking indicated an intact RNA sample. Pure, intact total RNA samples were either used immediately for further processing or placed at -80°C for long term storage.

2.3.12 – RNA Sequencing

For RNA-Seq, 1.5 µg of each purified total RNA sample was treated with DNAse-I (NEB) to remove potential gDNA contamination, before the RNA was re-purified and concentrated using the Monarch® RNA Cleanup Kit (NEB) according to the manufacturers'

recommendations. RNA quantity and quality was again checked *via* Nanodrop before submitting 600 ng of purified total RNA at 20 ng / µL for each tissue sample to a Novogene facility (Cambridge, UK) for processing. Sample QC, mRNA enrichment, cDNA library prep and Illumina paired-end 150 sequencing was carried out by Novogene, in addition to data QC and filtering, read alignment and mapping to the reference genome post sequencing.

Raw sequencing data was returned as BAM file and fastq format. All DEG, KEGG and GO analyses were performed using the raw read count and FKBP values for each gene in each tissue sample. DEG analysis was performed with a custom R script with the DESeq2 package.

2.3.13 – cDNA Synthesis and RT-PCR

For each BSF tissue extract, 2.5 µg of total RNA was treated with DNAse-I (NEB) to reduce gDNA contamination in subsequent library preparations. DNAse-treated RNA was purified and concentrated with the Monarch RNA Cleanup Kit (NEB) before re-quantification *via* Nanodrop. cDNA synthesis was performed on 1.5 µg of RNA from each sample in a 20 µL volume at 42°C for 60 minutes using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). Oligo(dT) primer was included in all reactions for the enrichment of mRNA, and random hexamers included in a 1:1 ratio with Oligo(dT) in samples to be analysed for non-coding RNA expression, namely those expressing gRNA or crRNA. No reverse transcriptase control (NORTs) reactions were performed in tandem for each sample to confirm a lack of gDNA contamination prior to cDNA synthesis. After heat-inactivation of reverse transcriptase, cDNA was diluted 1:5 with DNAse-free water and stored at -20°C.

Transgene expression was confirmed in the samples by standard RT-PCR using cDNA as templates and Phusion polymerase. PCR was performed for 32 cycles with the primer sets developed for qPCR and the appropriate T_A for Phusion polymerase. Short amplicons corresponding to housekeeping genes and transgenes were visualised by gel electrophoresis with 2% agarose gels.

2.3.14 – Immunoblotting

Following tissue collection, BSF egg and ovary tissues were snap-frozen in 250 µL passive lysis buffer (PLB, Promega, Madison, WI, US) on dry ice. Samples were then placed on ice to partially defrost before tissues were thoroughly homogenised with a sterile pestle (NEB). The homogenates were centrifuged at 16,000 x g for 5 minutes at 4°C to pellet debris before the clarified supernatants were transferred to new tubes on ice.

All remaining immunoblot steps, including sample denaturing, SDS-PAGE, blotting and image acquisition were kindly performed by Dr. Moeez Khan of Imperial College London in August 2024 according to a protocol developed previously for *Anopheles stephensi* tissue samples[467]. The total protein concentration of the BSF tissue samples were first estimated with Pierce Coomassie Plus Assay Reagent (Thermo Fisher Scientific). 10.0 µg of total protein was then diluted in 1X Laemmli sample buffer (BioRad) supplemented with 5% 2-Mercaptoethanol (Sigma-Aldrich) and the samples denatured by boiling at 95°C for 10 minutes. Samples were then loaded into a 4-15% Mini-PROTEAN TGX precast protein gel (BioRad) and electrophoresed in 1X Tris-glycine-SDS electrophoresis buffer (Thermo Fisher Scientific) at 200V for 40 minutes. 400 ng of NLS-Cas9-NLS Nuclease (GenScript) was loaded as a positive Cas9 protein condition, and the Precision Plus Protein Dual Color standard (BioRad) was loaded as a molecular weight marker.

Protein was transferred from the gel to a sheet of 0.45 µm nitrocellulose membrane (BioRad) using the Trans-Blot Turbo Transfer System (BioRad). The membrane was blocked in 3% (w/v) skim milk powder in Tris Buffered Saline with Tween 20 (TBST, Thermo Fisher Scientific) for 1 hour at room temperature with gentle rocking. The membrane was then cut in two horizontally at the 75 kDA marker using a scalpel, allowing for the separate Cas9 (~158 kDA) and endogenous beta-actin (~42 kDA) containing membranes to be incubated separately with their respective primary antibodies. The two membranes were then incubated in 10 mL of blocking solution containing their respective primary antibody for 1 hour. The ≥75 kDA Cas9-containing membrane slice was probed with a 1:1000 dilution of the GenCRISPR SpCas9 Antibody (4A1), (Mouse) (GenScript, A01935-40), and the <75 kDA beta-actin membrane was probed with a 1:1000 dilution of THE Anti-β-Actin [HRP] Monoclonal Antibody (Mouse) (GenScript, A00730).

After antibody binding, membranes were rinsed twice in TBST, then washed by incubating in TBST while rocking for five minutes, repeating the incubation twice with fresh TBST. As the Cas9 primary antibody was not HRP-conjugated, the Cas9 membrane was incubated in a secondary antibody Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, HRP (Thermo Fisher Scientific, G-21040). Following a second round of washing, both membranes were prepared for imaging by incubating in Amersham ECL Select Western Blotting Detection Reagent (Thermo Fisher Scientific) for five minutes. Excess reagent was then drained and the membranes imaged with a 30 second exposure time using a ChemiDoc XRS+ and Image-Lab software (BioRad).

2.4 – Insect Cell Culture

2.4.1 – S2 Cell Maintenance

The semi-adherent insect cell line *Drosophila* S2 (Invitrogen) was cultured as a monolayer in an incubator at 28°C in ambient conditions. Unless otherwise stated, all culture handling and centrifugation occurred at room temperature and all cell incubations occurred in the 28°C climate chamber. For routine culture, cells were maintained in T75 flasks containing 15 mL of complete media, consisting of *Schnider's* growth medium (Gibco, Waltham, MA, US) supplemented with 10% heat-inactivated fetal bovine serum (Gibco) and 1X Antibiotic/Antimycotic (Gibco). Cultures were passaged weekly by repeated aspiration and expulsion of culture media to dislodge adherent cells from the flask and break up clumps for a single cell suspension.

Cell density and viability was assessed using a BRAND Neubauer haemocytometer (Thermo Fisher Scientific). 100 µL of single cell suspension was diluted in 0.4% Trypan Blue Solution (Thermo Fisher Scientific), and 10 µL of the dilute suspension was added to the counting chamber between the haemocytomer surface and top glass coverslip. Live cells were distinguishable from dead cells at 10X magnification by trypan blue exclusion, allowing for estimation of live : dead cells as a percentage. Non-clumping live cells within each of the four corner sets of 16 squares were counted and the average cell count determined across the four corner sets. The average was multiplied by the trypan blue dilution factor, then multiplied by 10,000 to determine the live cell count / mL of original cell suspension. Upon

reaching cell densities of 10-15 x 10^6 cells / mL, cells were passaged by splitting at a 1:5 dilution in fresh complete media to a final density of 2–3 x 10^6 cells / mL.

S2 cells were retained for a maximum of 5 consecutive passages before initiating a new flask with a frozen stock from an earlier passage. S2 cryovials were removed from liquid nitrogen storage and partially thawed in a 30°C water bath before diluting the defrosted suspension in 5 mL of complete media. Cells were pelleted at 1000 x g for 4 minutes, the cryopreservative-containing suspension removed, and the pellet resuspended in 5 mL fresh media. Freshly initiated cells were allowed to recover in T25 flasks before seeding a T75 flask on the next passage.

S2 stocks were prepared in 1 mL cryovials at low passage numbers. Cells from a passage with >95% viability were washed once in PBS and resuspended to a density of 11 x 10^6 cells / mL in 1 mL of freezing medium, consisting of 1:1 ratio of fresh and conditioned complete media with 10% DMSO supplemented as a cryoprotectant. Freezing rate was controlled by placing cryovials in a Mr Frosty freezing container (Thermo Fisher Scientific) in a -80°C freezer for 24 hours, before transferring stocks to liquid nitrogen for long term storage.

2.4.2 – S2 Cell Transfection

A calcium phosphate-mediated transient transfection method was adapted to scale with culture volumes used in 24-well plates. The day before transfection, cells from a low passage number were resuspended to a density of 1.0 x 10^6 cells / mL in complete media without antibiotics. For each transfection condition, 600 µL of diluted culture was seeded per well of a 24 well plate for a concentration of 0.6 x 10^6 cells / well. Plates were sealed with parafilm to avoid evaporation and incubated for 16–24 hours to allow cells to adhere and reach a density of around 1.2 x 10^6 cells / well.

On the day of transfection, up to 3.6 µg of sterile-filtered plasmid DNA (1.8 µg each for the co-transfection of two plasmids) was diluted in sterile water to a final volume of 52.8 µL for each well. 7.2 µL of 2 M $CaCl_2$ (Thermo Fisher Scientific) was added to bring the solution up to 60 µL and 240 mM $CaCl_2$. A second tube per well was prepared containing 60 µL of 2X HEPES (Sigma-Aldrich). With continuous vortexing of the HEPES-containing tube, the entire DNA-$CaCl_2$ solution was slowly added dropwise to the second tube over two minutes to ensure the production of a fine precipitate. Transfection mixes were incubated at room temperature for 40 minutes before adding the whole solution dropwise to cell layer with

gentle swirling to disperse the mix. After transfecting each well, plates were resealed with parafilm before incubating for 16-24 hours.

The day after transfection, 400 µL of complete media (without antibiotics) was added to the transfected cells. Cells were detached from the plate by repeated pipetting and pelleted by centrifugation at 1000 x g for 3 minutes. The supernatant was removed, and the cells resuspended in 0.6 mL of fresh complete media (without antibiotics) before replating in a new 24-well plate. Plates were resealed with parafilm and incubated for an additional 48 hours to allow for an increased period of protein expression.

Transfected cells were imaged and harvested 72 hours post-transfection. Cells were imaged at 10X magnification in brightfield, GFP and TX Red fluorescent channels using an EVOS M5000 Imaging System (Invitrogen). The distribution of GFP and DsRed marker co-expression was used to calculate transfection efficiency across multiple fields per well.

2.4.3 – Fluorescence Activated Cell Sorting

After imaging, positively co-transfected cells from each cell population were enriched for later sequence analysis by fluorescence activated cell sorting (FACS). Cells were detached from culture plates, washed in PBS and resuspended in 400 µL of PBS supplemented with 2% fetal bovine serum, to a concentration of around 5.0 x 10^6 cells / mL. Immediately prior to sorting, DAPI stain was added to the cells at a concentration of 10 µg / mL to allow for live : dead cell exclusion.

FACS was performed using a BD FACSAria III cell sorter and BD FACSDiva software (BD Biosciences, Franklin Lakes, NJ, US). Gating was manually set based on forward scatter (FSC) and side scatter (SSC) properties to exclude cellular debris with the use of a heat-killed cell sample. Compensation for GFP and DsRed was performed using mono-transfected control populations expressing the single fluorophores. Detection for each fluorophore was achieved using the following channel settings:

DAPI – 405 nm laser, 450/50 nm bandpass filter

eGFP – 488 nm laser, 530/30 nm bandpass filter

DsRed – 651 nm laser, 582/15 nm bandpass filter

Co-transfected populations were sorted on the basis of detection of both GFP and DsRed. For each successfully sorted population, between 30,000 and 56,000 GFP/DsRed co-expressing cells were sorted into a corresponding Eppendorf tube containing PBS with 2% fetal bovine serum. Cells were pelleted by centrifugation at 1000 x g for 15 minutes before the supernatant was removed and the cells resuspended in Fast extract DNA solution to a concentration equal to ~1000 cells / µL. DNA extraction was performed according to the Fast extract protocol and the DNA preparations subjected to mutant profile analysis by PCR and DNA sequencing.

2.5 – Software and Bioinformatics

Raster graphics were compiled, edited and annotated using Paint.net (version 5.0). Vector graphics were produced using Inkscape (version 1.3). Graphical data representations were produced using both GraphPad Prism (version 10.1.1) (Dotmatics, Boston, MA, US) and the R package ggplot2 (version 3.5.1).

All gene sequence information was identified using the NCBI gene database and through protein and nucleotide alignments performed with the NCBI BLAST tool. BLAST was also used to identify conserved regions shared between orthologous genes. All *Drosophila melanogaster* sequence information were derived from the NCBI RefSeq assembly GCF_000001215.4 (dm6), and *Hermetia illucens* sequences from the assembly GCF_905115235.1 (iHerIll2.2.curated.20191125).

All sequence mapping and analysis, including plasmid cloning designs and Sanger sequence alignment, was performed using the Geneious Prime® software (Biomatters, Auckland, NZ). All nucleotide and amino acid sequence alignments were performed in Geneious Prime® using the inbuilt Clustal Omega tool. Selection of CRISPR target site candidates was aided by use of the web tool CHOPCHOP (version 3)[468]. Indel detection in Sanger sequencing data was performed using the web tool DECODR (version 3), and analysis of gene editing outcomes in amplicon sequencing data was performed using CRISPResso2 (version 2.3)[469,470].

2.6 – Statistics

All statistical analyses were performed using base R packages. The G-Test of Goodness of Fit was conducted to assess whether the observed rates of transgene inheritance deviated from the expected Mendelian ratios. An a priori power analysis was performed to determine the appropriate sample size for screening to detect potential inheritance bias. Chi-square tests were used to compare neonate to adult survival rates across transgenic genotypes. Fisher's exact test was employed to assess whether transgene inheritance was biased by the sex of the offspring.

Chapter 3: **Developing an Endogenous CRISPR Toolkit for Gene Editing and Transgenesis in Black Soldier Fly**

3.1 – Introduction

The primary aims of the current study involved the characterisation of the CRISPR gene editing (GE) system for creating targeted mutations when expressed endogenously in BSF. To control the initiation of gene editing, the CRISPR nuclease and CRISPR RNA components would be designed to be maintained as transgenes in separate engineered BSF lines. Upon mating of these lines, a fraction of the progeny should inherit a copy of each transgene, following which GE activity may be possible from as early as the embryonic stage.

To produce a segregated molecular toolkit for GE in BSF, two types of DNA transgenesis construct would be required: one set encoding the CRISPR nuclease genes, and another encoding the CRISPR RNA expression cassettes corresponding to different target genes. Based on numerous instances of successful employment for transgenesis in non-model insect species, the *piggyBac* (*pBac*) transposon system was selected as the most likely method for completing integration of the transgenesis constructs in BSF[416]. The segregated CRISPR component designs, therefore, were all produced to fit within a common *pBac* vector backbone, allowing non-site specific integration when co-injected into BSF embryos with a source of *pBac* transposase.

With a suitable transgenesis method outlined, the contents and intended function of the different types of constructs used for transgenesis should be considered. One potential problem with the described CRISPR crossing strategy revolves around the potential toxicity associated with nuclease expression, and Cas9 more specifically. In multiple previous cases, Cas9 has been associated with endowing a low fitness when overexpressed endogenously, with a potential combination of protein burden and off-target mutagenic activity explaining high observed nuclease toxicity[471,472]. In addition to this, a non-toxic off-target activity is also encountered with constitutive Cas9 expression, potentially

disrupting genetic stability throughout the organism and hampering efforts to link causal mutations in a GOI with a phenotypic change[473].

To address this, the CDS of the CRISPR nuclease component can be placed under tight conditional regulation using a germline-specific promoter and terminator flanking sequences. Nuclease expression will then, ideally, follow the pattern of expression for the germline-specific gene in question, being highly enriched in gametes and embryonic cells but restricted in all other tissues. This allows GE activity to be kept to its maximum in germline cells, thus increasing the likelihood of edits being passed on to following generations. At the same time, GE activity is dampened in other somatic tissues and developmental stages, reducing both protein burden and the risk of off-target activity causing accumulation of mutations and limiting fitness.

In several insect species, the regulatory regions for a selection of germline-specific genes have been mined, typically referring to a promoter and a terminator element. Rather than attempt to unravel the sequence and location of transcriptional regulatory motifs, a commonly employed strategy is to simply isolate a region (~2.0 kb) immediately upstream of the start codon of the endogenous germline gene for use as a promoter sequence[474]. The same strategy can be used to isolate the terminator, selecting a similarly sized region immediately downstream of the gene's stop codon.

Using both of these elements to flank the CDS of a transgene can roughly approximate the same expression level and context as seen for the chosen germline gene, although differences between the two genes in terms of distal promoter elements, genomic loci and epigenetic status are likely to influence their relative expression levels[475]. In insect studies, the regulatory regions of the development genes *nanos* and *vasa* are commonly isolated as effective sources of endogenous germline-regulatory control[385,476,477].

In *D. melanogaster, nanos* is an essential zinc-finger RNA binding protein that represses the translation of specific mRNAs, including *hunchback*, to regulate embryonic development and define the anterior-posterior axis[424]. The concentration of *nanos* at the posterior pole of the embryo is one of the prerequisite signals for cells in this region to develop into primordial germ cells (PGCs), the precursors of ovaries and testes in the adult germline[478]. In addition to its role in body patterning, the presence of *nanos* in the adult germ tissues is required for the continual maintenance of germline stem cells (GSCs), preventing premature differentiation.

Vasa is an RNA helicase in the DEAD-box family with a similar embryonic and adult germline distribution as *nanos*[425]. Located at the posterior pole in embryo development, *vasa* regulates the translation of *nanos* and other maternal genes by unwinding the secondary structures of their mRNAs localised in this region, leaving them open for active protein expression. It also plays a role in directing the migration of PGCs in the posterior of the embryo, and supports both the maintenance of GCSs and differentiation of the gametes in adult germline tissues[479]. With both *nanos* and *vasa* being highly expressed in PGCs in the early embryo, as well as in GSCs in the adult tissues responsible for production of gametes, a CRISPR nuclease placed under a similar pattern of expression is likely to be highly efficient for mediating GE in cells destined to seed the following generation.

While germline-restricted regulatory elements have been well-characterised for improving endogenous CRISPR systems, there are fewer options for controlling CRISPR RNA expression. As small and non-coding RNAs, CRISPR RNAs are commonly placed under the control of Pol III promoters in eukaryotes[480]. With tight regulatory control governing nuclease expression, it has been assumed that having a high, constitutive and global expression of the CRISPR RNA is more beneficial in terms of GE efficiency. Pol III promoters, which mediate the exclusive transcription of small non-coding RNAs like rRNA subunits and tRNAs, are well suited to this role. A Pol III promoter that has seen successful employment in a number of species for efficient gRNA expression is found in the U6 small nuclear RNA (snRNA) gene. The U6 transcript forms a central part of the spliceosome machinery in the processing of pre-mRNA, and as such is generally expressed at a consistent level in most cell types[481]. With most organisms encoding multiple copies of the U6 gene, the U6 promoter has been employed for expressing gRNAs in many species since some of the earliest experiments regarding endogenous CRISPR transgenesis[482].

The final consideration for the BSF CRISPR system is the selection of appropriate mutagenesis targets. Ideally, genes should be chosen that allow for easy identification of the mutant phenotype at multiple developmental stages, while not imparting any significant fitness cost. Conserved pigmentation markers, such as the *white* and *yellow* genes, often represent the easiest target for creation of clear mutant phenotypes when testing the efficacy of GE technologies such as CRISPR in a new species of insect[483,484]. In addition to genes with a clear readout, several GOI can be identified in BSF that may have value for enriching beneficial economic traits in an engineered strain, or line.

Although several desirable genetic traits in BSF have been described in this study, the traits with the most obvious and immediate value are final larval mass, or body weight at the point of harvest, and larval development time. In insects, these two traits are closely linked, with the overall decision to proceed to pupal stage depending on reaching a sufficient larval weight, or critical weight, as well as waiting for a specific period of time to elapse[485]. In the absence of any wide genetic resources for the identification of potential yield-associated QTL in BSF, candidate targets for increasing final larval mass and decreasing development time will instead be mined from the literature, primarily from evidence of orthologous mutants with the desired traits in related insect species.

With the general requirements for the transgenic CRISPR toolkit outlined, the first step to their production is to select an assortment of endogenous regulatory elements and targets for mutagenesis in BSF. This includes the identification and isolation of germline-regulatory promoter and terminator elements for controlling CRISPR nuclease expression, as well as Pol III promoters for the constitutive expression of CRISPR RNAs. With these in place, the toolkit can be assembled across a series of *pBac* plasmids for preliminary assessment of GE efficiency in S2 cell culture, before being taken forward for transgenesis in BSF.

3.2 – Results and Discussion

3.2.1 – Selecting Pigmentation Target Genes for Mutagenesis

As the gene deemed most likely to produce a clear, unambiguous phenotype when knocked out with the CRISPR toolkit, a gene and corresponding protein sequence orthologous to the *D. melanogaster* eye pigment precursor transport gene *white (w)* was first identified *via* protein BLAST in the BSF genome (*Hi_w*, *LOC119658012*). Multiple protein sequence alignment was then performed to compare the potential BSF *Hi_w* primary protein sequence to the *w* orthologues of three insect species that display a prominent *w* mutant phenotype (**Figure 6**). The identified *H. illucens* protein sequence shared a high homology with the corresponding gene in *D. melanogaster* (65.0%), *A. gambiae* (57.8%) and *Ceratitis capitata* (82.7%), suggesting that the identified BSF gene *LOC119658012* was the functional equivalent of *w* in the species.

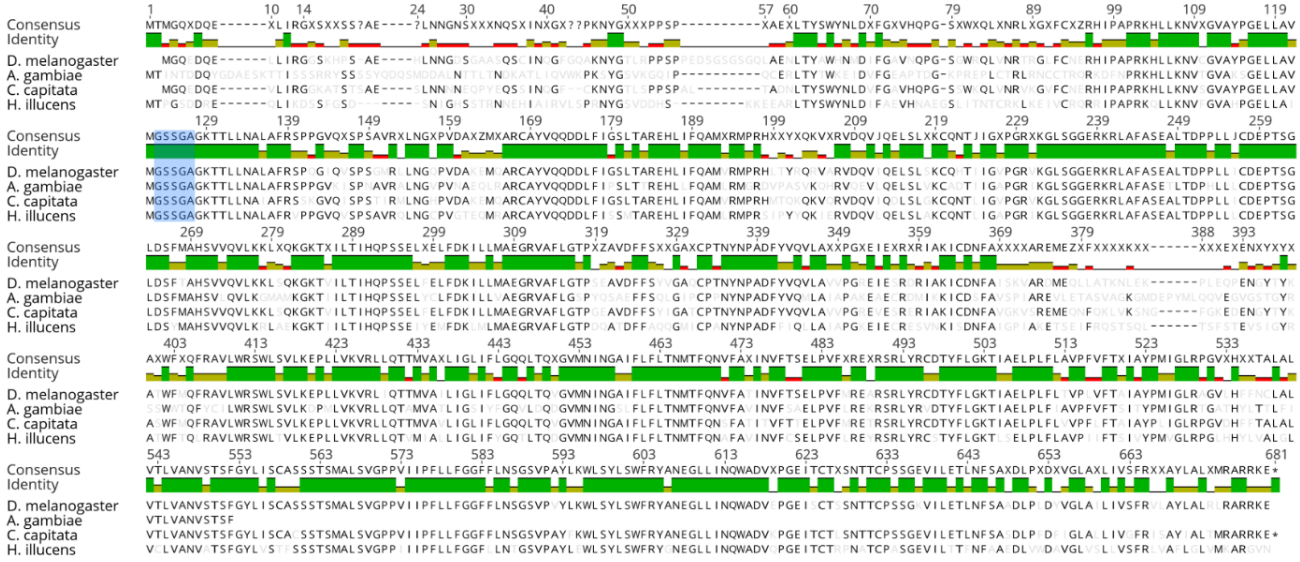


Figure 6: Multiple protein alignment for the identification of the *H. illucens white* gene. The *white* protein sequence from *D. melanogaster* was used as a query in protein BLAST for the identification of orthologous genes in the BSF (*H. illucens*) reference genome. A high-confidence prediction was made for BSF *w* being encoded by the gene *LOC119658012*, with the proteins of both species sharing an identity of 65.0%. Multiple alignment with the *D. melanogaster, A. gambiae* and *C. capitata* genes confirmed high protein homology and sequence conservation, with the conserved region identified for later CRISPR target selection in *H. illucens* highlighted in blue.

Located on the 5th chromosome, *Hi_w,* hereby referred to as *w,* encodes a single mRNA of 8 exons and 2647 bp in length, with the *w* CDS encoding a single protein of 679 AA in length (**Figure 7A**). To compare the *w* genomic sequence identified from the current reference genome (iHerIll2.2.curated.20191125) with that of the line intended for later rearing in the BSF insect colony, the T4-Jan-3 ("T4") *wild-type*, DNA was first extracted from the legs of a pool of three male and three female T4 flies by organic extraction-based methods. Despite several silent polymorphisms, the *w* protein coding sequence and reading frame in the T4 line was identical to that found in the reference genome.

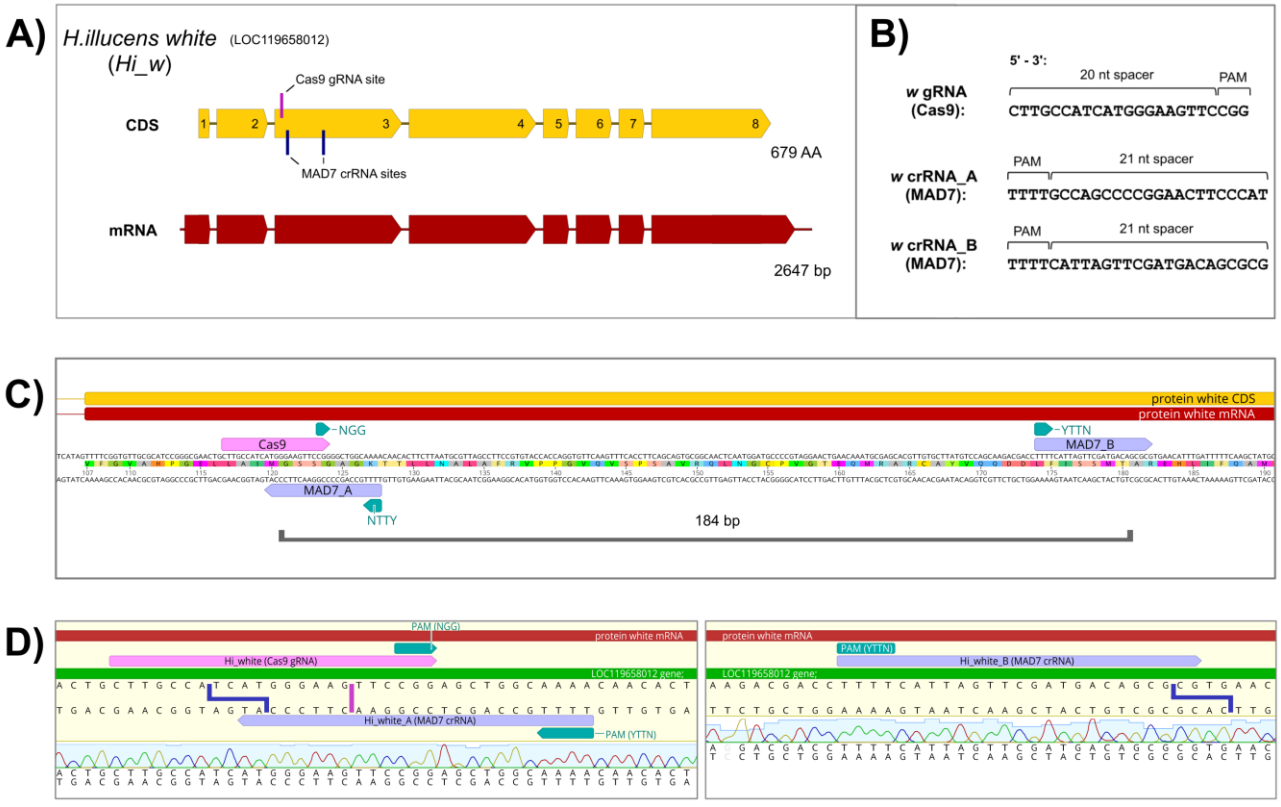


Figure 7: Gene structure analysis and target selection in *H. illucens white*. A) A structural overview of the *Hi_w* gene, with the relative identified CRISPR target sites in exon 3 indicated by coloured lines. **B)** Sequences corresponding to the selected *w* target sites for activity with Cas9 or MAD7, including the genomic PAM sequences not included in the CRISPR constructs. **C)** A sequence overview of the conserved region in exon 3 chosen for target identification, with the relative locations of the Cas9 gRNA and the two MAD7 crRNA sites. **D)** Both Cas9 gRNA and crRNA w_B are oriented to the sense DNA strand (5'-3'), while crRNA *w*_A shares homology to the inverse strand, ensuring that the crRNA *w*_A and gRNA-w targets overlap around the cut site. The genome sequence of this section of the *w* gene in the T4 *wild-type* line was confirmed *via* sequencing.

With the *w* sequence confirmed in the T4 genome, which would represent the *wild-type* BSF line for the purpose of later GE and transgenesis experiments, potential CRISPR target sites were investigated for generating a *w* knock out. First, a conserved region that indicated high protein homology shared between the *w* orthologues of the previously aligned insect species was identified (**Fig. 6**). Conserved protein regions shared between species generally suggests that these residues are critical for protein function. Therefore, targeting conserved protein coding regions with CRISPR increases the likelihood that a non-frameshifting amino acid change following indel still inactivates the activity of the protein, creating an effective null mutation.

To avoid any complications from possible alternate start codons, a long section of conserved protein sequence near the start of the 3rd exon was selected for identifying potential target sites for Cas9 and MAD7 (**Fig. 7A**). The online tool CHOP-CHOP was then employed to identify a list of Cas9-gRNA and MAD7-crRNA variable sequences in this region, following the constraints of either a 20 bp sequence upstream of an NGG PAM site (5'-3') or a 21 bp sequence downstream of a YTTN PAM site (5'-3') for Cas9 and MAD7 respectively. One Cas9 gRNA *w* sequence and two MAD7 crRNA *w* sequences were then selected for subsequent cloning (**Fig. 7B**).

As a multiplex-capable nuclease, the decision was made to generate a *w* KO with MAD7 by targeting DNA cleavage at two distinct targets 184 bp apart (**Fig. 7C**). The expected cleavage site of the first MAD7 target, crRNA *w*_A, is closely located to the cleavage site expected for the Cas9 gRNA, allowing for the GE activities of the two nucleases to be more closely compared at an almost identical target locus. In addition to characterising the multiplex capability of the MAD7 nuclease and the polycistronic precursor CRISPR RNA developed later, the simultaneous cleavage of two sites in close proximity on the chromosome is expected to be more likely to produce a KO, with the NHEJ mechanism being likely to ligate the two larger chromosome fragments directly and omit the small inter-target DNA fragment[253].

Although no commercial source of purified MAD7 protein could be acquired, the cleavage activity of recombinant Cas9 protein at the *w* locus with the selected gRNA sequence was assessed *in vitro* (**Figure 8**). Using a DNA template produced from a gRNA *w*-specific primer, an *in vitro* transcription (IVT) reaction was performed to produce a stock of transcribed *w*-gRNA. From the T4 *wild-type* genome extraction, a 455 bp *w* test amplicon that spanned the

area around the Cas9 target site was also produced as a Cas9 DNA cleavage substrate. Purified Cas9 protein and the IVT *w*-gRNA were combined and incubated to create active Cas9-RNP complexes, which were then incubated with the *w* test amplicon in an *in vitro* Cas9-cleavage assay. With cleavage products of the expected size apparent on an agarose gel, the chosen *w*-gRNA sequence showed a sufficient activity for directing Cas9 cleavage in the *w* locus *in vitro* (**Fig 8**).

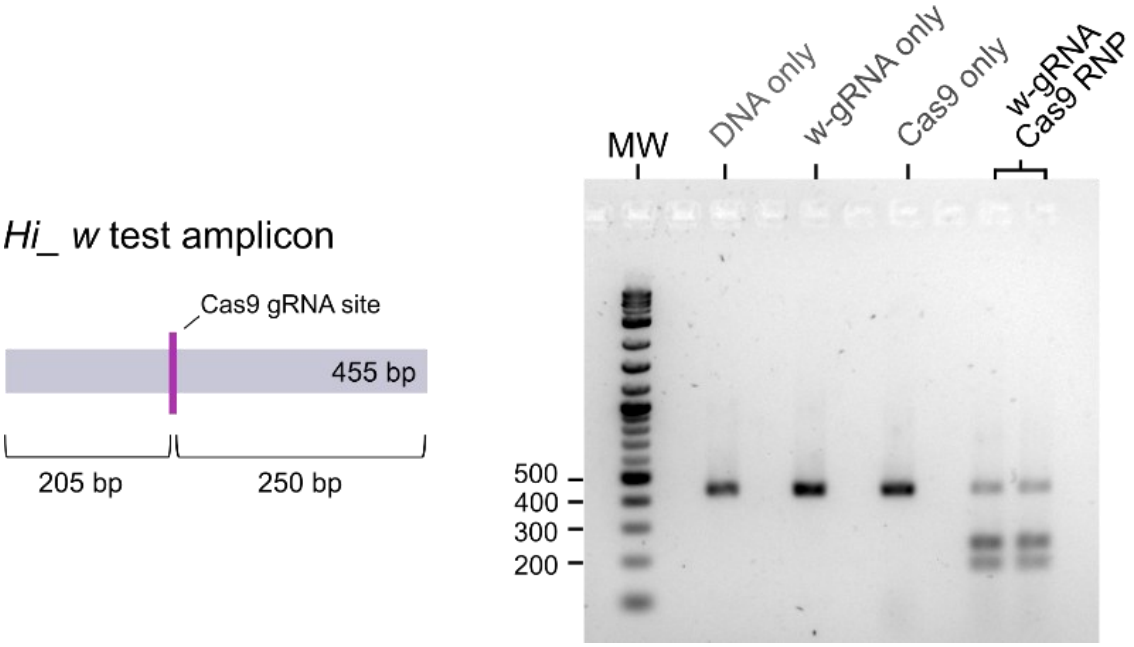


Figure 8: Confirmation of Cas9 gRNA-w activity *via in vitro* Cas9 cleavage assay. A 455 bp amplicon was produced from T4 *wild-type* genomic DNA that encompassed the region surrounding the *w* gRNA target site. The DNA test amplicon was incubated for 30 minutes at 37°C alone, and with Cas9, gRNA or preassembled Cas9-gRNA-w RNP complexes. Strong single DNA bands at the expected amplicon size confirm no cleavage activity in the non-RNP conditions, with the presence of two cleavage products of the expected size in the RNP condition suggesting efficient cleavage at the expected target site. 1 kb Plus DNA Ladder (NEB) was used as a MW marker.

Although ultimately untested, an additional four potential pigmentation genes were identified as alternative targets in BSF, as a contingency should *w* KO not associate with a clear phenotype. Using the same workflow as for identification of *w*, the *H. illucens ebony* (*e*), *yellow* (*y*), *sepia* (*se*) and *white pupa* (*wp*) genes were determined from multiple protein alignment (**Figure 9**). In *D. melanogaster, e* and *y* are involved in the synthesis of melanin in the larval and adult cuticles, with their eponymous mutant phenotypes representing a deficiency in pigment deposition leading to a clear change in colour. Mutations in *se* leads to a deficiency in the transport of red pigment granules in the eye, resulting in a dark brown

colouration. *wp* mutants, as the name suggests, specifically affects cuticle pigmentation at the pupal stage. Not present in *D. melanogaster,* the *wp* mutant is commonly employed in tephritid flies as a genetic sexing marker for applications in biocontrol, with the BSF orthologue identified by protein alignment with *C. capitata, Bactrocera dorsalis* and *Zeugodacus cucurbitae*[486].

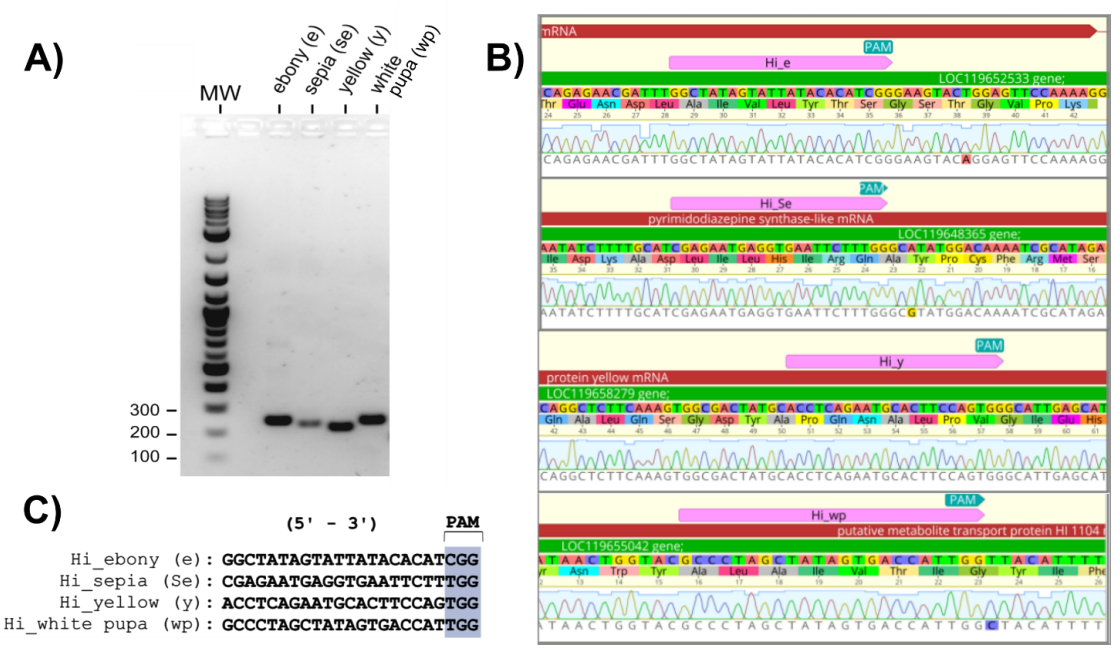


Figure 9: Alternative pigmentation gene targets for mutagenesis in BSF. A) For the pigmentation genes *ebony (e), sepia (se), yellow (y)* and *white pupa* (*wp),* a genomic region of around 250 bp surrounding the chosen Cas9 gRNA target site was amplified from T4 *wild-type* genomic DNA for sequence analysis. **B)** Following sequencing, several silent polymorphisms were found for the selected pigmentation genes in the T4 sequence trace (bottom) when compared with the reference genome (T). Regions of conserved protein sequence were identified using multiple protein alignment and selection of potential gRNA sites avoiding the T4 polymorphic regions was performed using the online tool CHOPCHOP. **C)** gRNA sequences and corresponding genomic PAM sites identified for the four alternative pigmentation genes.

With each of the BSF pigment orthologues sharing a high protein identity with genes from the compared species during alignment (56.4% - 73.9%), putative Cas9 gRNA targets were identified and the surrounding genome sequence confirmed in the T4 genotype using the same process as with *w.* Although validating these mutant phenotypes ultimately fell out of the scope of this project, the gRNA-expression constructs for each gene were successfully produced and sequence-verified, with the potential for later testing. Data regarding the pigmentation genes analysed in this study is presented in **Table 1**.

Gene ID *(H. illucens)*	Gene symbol	Gene description	Gene location	Transcript length (nt)	Protein length (aa)	Target residues (aa, 5'-3')	Target [PAM] and spacer sequence (5'-3')	Gene ID *(D. melanogaster)*	Reference (DOI):

119658012	***white (w)***	Eye pigment precursor family transporter *white*	Chr-5, NC_051853.1 (69556823..69614667)	2647	679	(*Cas9*)
						117 – 124 :
						(*MAD7_A*)
						120 – 128 :
						(*MAD7_B*)
						174 – 182 :
							`CTTGCCATCATGGGAAGTTC[CGG]`
							`[TTTT]GCCAGCCCCGGAACTTCCCAT`
							`[TTTT]CATTAGTTCGATGACAGCGCG`
								31271	10.1007/s00438-023-02088-0
									10.1016/S0005-2736(99)00064-4
									10.1111/j.1365-2583.1995.tb00027
									10.1126/science.270.5244.2005

119652533	***ebony (e)***	Nonribosomal peptide synthetase ebony	Chr-3, NC_051851.1 (92107679..92138561, complement)	3221	869	182 - 190	`GGCTATAGTATTATACACAT[CGG]`	42521	10.1007/BF00326537 10.1534/genetics.108.096388
119648365	***sepia (se)***	Pyrimidodiazepine synthase-like	Chr-2, NC_051850.1 (147585815.14759297 complement)	1279	256	33 - 41	`CGAGAATGAGGTGAATTCTT[TGG]`	38973	10.1042/BJ20060424 10.1016/j.beproc.2005.06.004
119658279	***yellow (y)***	L-dopachrome tautomerase yellow	Chr-5, NC_051853.1 (88373889..88415321)	1768	505	54 - 62	`ACCTCAGAATGCACTTCCAG[TGG]`	30980	10.1111/1744-7917.13371 10.1073/pnas.82.21.7369 10.1534/genetics.108.096388 10.1038/s41598-017-04686-6
119655042	***white pupa (wp)***	Putative metabolite transport protein HI_1104	Chr-4, NC_051852.1 (2764869..2839535, complement)	2645	527	16 - 23	`GCCCTAGCTATAGTGACCAT[TGG]`	101451947 (*C. capitata*)	10.1038/s41467-020-20680-5 10.1101/2020.05.08.076158

Table 1: Pigmentation gene candidates identified for mutagenesis during this study. All BSF genes were identified *via* protein multiple sequence alignment against orthologous genes from *D. melanogaster* or *C. capitata*. Gene information, descriptions, locations and comparative Gene IDs were identified from the NCBI database. gRNA and crRNA target sequences were identified using the online tool CHOPCHOP[468].

3.2.2 – Identification of Larval Growth Regulatory Genes for Multiplexed Editing

In addition to pigmentation genes, several genes with predicted function in the BSFL growth and developmental pathway were also identified for mutation with the CRISPR toolkit. In the developed CRISPR cross strategy, several GOI can be targeted simultaneously *via* multiplexed GE, with the transheterozygote progeny of the cross having variable repair outcomes at each target site. In an ideal scenario, this could be employed to generate variation in a genomic locus predicted to have an influence over a desirable trait, with the most beneficial phenotype isolated from a heterogeneous pool of edited genotypes.

However, in the absence of specific information on loci associated with phenotypic variation in BSF, candidate genes were instead mined from literature sources that described comparative mutations in *D. melanogaster*. The FlyBase bioinformatics database was also used to locate studies on desirable larval phenotypes, specifically those associated with increased final larval mass and/or reduced larval development time. Guided by descriptions of mutant phenotypes in *D. melanogaster*, orthologous BSF genes were identified using protein BLAST and, whenever possible, compared with orthologs from other species to increase confidence in their selection. Based on their effects when mutated in *D. melanogaster*, these BSF genes could be reasonably expected to influence larval growth parameters similarly when mutated independently .

From a list of potential targets, six candidate *D. melanogaster* genes with orthologs in BSF were selected for possible application with the CRISPR gene editing toolkit: *Maf1*, *ImpL2*, *CRBN*, *Sirt6*, *TBCd17*, and *Piezo* (**Figure 10**). These genes were associated with various larval growth phenotypes when ablated in *D. melanogaster*, each with the potential to yield similar beneficial effects as comparative mutants in BSF. In each case, the criteria for inclusion were that *D. melanogaster* lacking the gene activity remained viable and fertile, could be feasibly maintained over several generations, and demonstrated either a modest increase in mass or size at the final larval instar or pupal stage, or a clear reduction in larval development time (defined as time to pupation) without a corresponding reduction in mass.

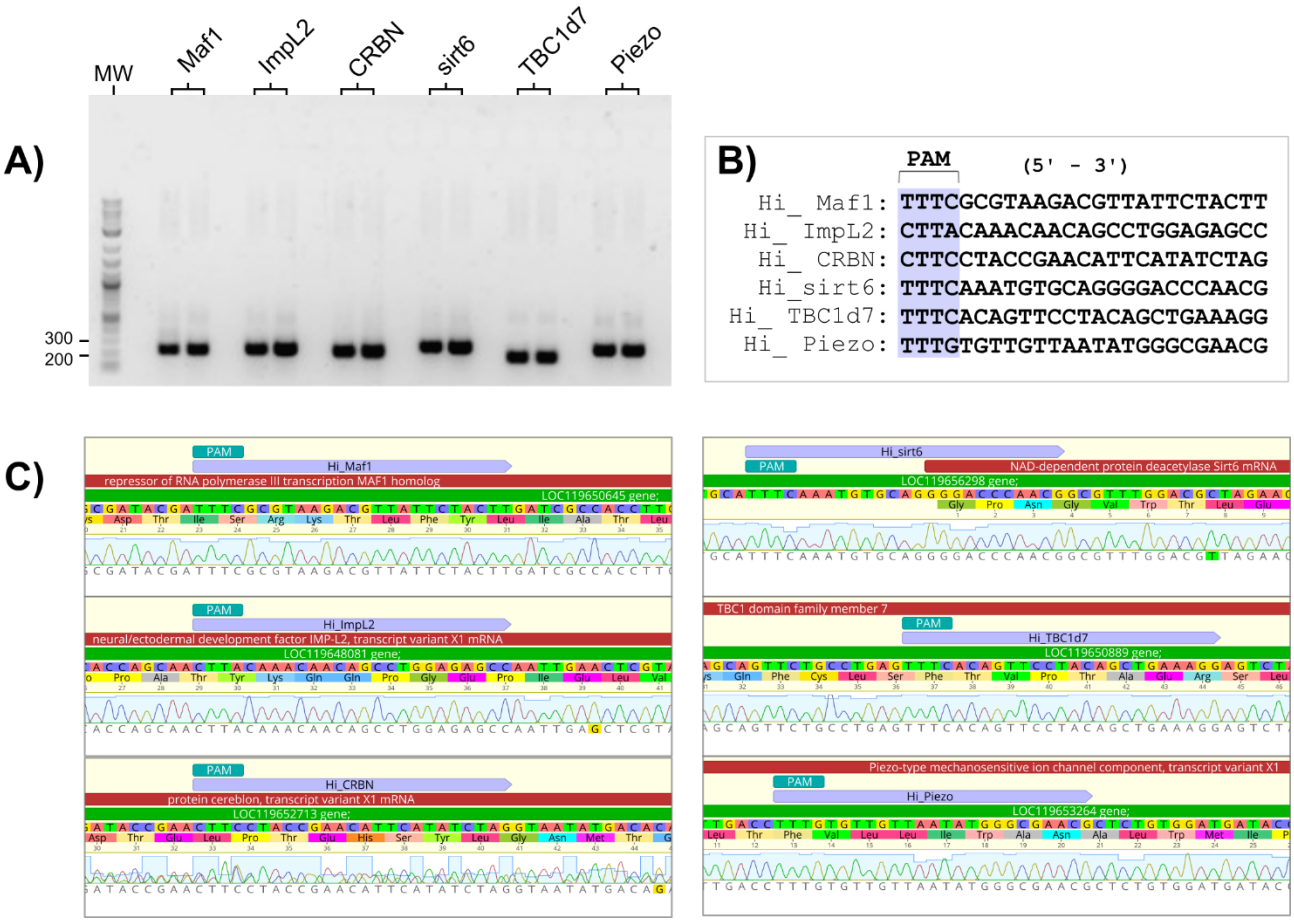


Figure 10: Candidate growth regulatory gene (GRG) targets for increased larval mass or reduced larval development time. A) A ~250 bp region flanking the identified MAD7 crRNA target site for each of the 6 GRG genes was amplified from the T4 *wild-type* line for sequence analysis. **B-C)** No polymorphism was detected directly in the selected crRNA target sequences, with the MAD7 PAM sites highlighted in blue.

In *Drosophila, Maf1* acts as a repressor of Pol III transcription, and in particular is associated with inhibiting the synthesis of the initiator methionine tRNA ($tRNA_i^{Met}$) as a strategy to regulate body size and developmental timing[487]. The protein is responsive to the nutrient status and oxidative stress in the cell, with a low nutrient status being transduced *via Maf1* to a suppression of $tRNA_i^{Met}$ expression. This in turn limits overall protein synthesis and inhibits cellular growth and development in conditions of nutritional stress, helping to maintain energy balance and conservation of resources.

Researchers found that inhibiting *Maf1* production in *D. melanogaster via* ubiquitous expression of an RNAi transgene was found to stimulate high rates of tRNA synthesis in the larval fat body, regardless of nutritional status. This was accompanied by a systemic increase in insulin signalling and growth, with around a 23% increase in pupal volume in *Maf1*-deficient insects. Interestingly, a significant acceleration was also seen in the *Maf1*

suppressed larva population, with over 60% entering the pupal stage within 120 hours AEL compared to less than 20% of the non-inhibited flies. Following, the authors suggested that the systematic response seen in *Drosophila* was similar to that seen when inhibiting the production of ecdysone in the fat body, similar to the mechanism that was effectively targeted in the generation of the high-weight BSF *HiPtth* mutant[488,253].

With similar growth inhibitory functions, *Imaginal morphogenesis protein-Late 2* (*ImpL2*) is a secreted protein that functions like an insulin-like growth factor binding protein in insects, inhibiting the interaction of insulin-like peptides with insulin receptors to inhibit overall growth. Whereas *ImpL2* overexpression impairs overall growth in *Drosophila*, homozygous null mutants generated by P-element insertion have been found to be associated with a significant increase in body weight, with male weight increasing by 24% and females by 64%[489]. A later study also found that mutant larva had a two-fold increase in circulating insulin level and a reduced development time, with the pupation time brought forward by 10 hours compared with non-mutants[490]. The KO of *ImpL2* appeared to phenocopy the overexpression of Insulin-like peptide 2 (ILP2), reducing larval starvation resistance at the expense of accelerated growth.

Cereblon (CRBN), a conserved receptor in the E3 ubiquitin ligase complex, also shows a regulatory activity in insulin signalling-mediated growth[491]. In *Drosophila*, homozygous *CRBN* mutants produced *via* CRISPR-Cas9 GE have been associated with elevated insulin signalling, leading to increased body weight and organ size[492]. Although viable and fertile, around 30% of homozygous mutants had a larval developmental day of around 12 hours, with the remaining larva showing a similar development time to controls. Despite no increase in food intake, the mutant larva saw an 8.4% increase in pupal length on average, as well as a general increase in global insulin signalling. KO of CRBN was also associated with a reduction in ImpL2 expression, raising the potential for synergistic effects when both genes are targeted simultaneously.

The NAD+-dependent deacetylase *Sirtuin 6 (Sirt6)* plays a important function in regulating glucose homeostasis, lipid metabolism and fat storage, along as having several other roles in coupling nutritional status to overall growth[493]. *Sirt6* mutant larva, generated with CRISPR, were found to be significantly larger and heavier than *wild-type* controls from 48 hours AEL, with disparity in growth becoming even more apparent at later ages[494]. With no apparent change in critical weight, or the body weight developmental checkpoint past which larva

begin to transition to the pupal stage, larval development time was greatly accelerated in tandem with increased larval mass, reaching the pupal stage over 24 hours faster than non-edited larva. However, adult longevity and resilience to starvation was reduced in *Sirt6* mutants, with adequate larval nutrition being essential for fitness. Interestingly, an overexpression of *Sirt6* was found to increase life span and resistance to oxidative stress in adult *Drosophila*, at the expense of a reduced rate of protein synthesis and a lower body weight[493].

Central to regulated growth in *Drosophila*, the mechanistic target of rapamycin complex 1 (mTORC1) promotes overall cell growth in response to nutrient availability and insulin signaling[495]. The Rab GTPase *TBC1d7* is a core member of the tuberous sclerosis complex (TSC), a tumour suppressor that inhibits the activity of mTORC1 to dampen cell proliferation and survival[496]. Using CRISPR-Cas9, a homozygous null TBC1d7 mutant was viable and had a roughly 25% increase in body size at the pupal stage[497]. Independent from TCS activity, the mutant was later found to be associated with a significant increase in ILP2 secretion from insulin producing cells in the brain, possibly breaking a link between low-nutrition sensing and the suppression of insulin signalling and global growth.

Finally, *Piezo* is mechanosensory ion channel found in neurons in the gastrointestinal tract, with a role in volume sensation contributing to a decrease in appetite in response to large meals[498]. With overexpression effectively causing flies to fast, null *Piezo* flies displayed a markedly increased appetite, with food intake, feeding duration and body weight increased significantly when compared with non-mutant *Drosophila*[499]. Interestingly, mutants had an otherwise normal lifespan and enteric anatomy, although a higher defecation rate, and eventually stopped feeding after significant crop distention, suggesting the presence of compensatory satiety pathways aside from *Piezo*-mediated gut mechanosensation.

With the exception of the feeding volume regulator *Piezo*, the other genes identified as candidate mutants in *Drosophila* were central mediators in nutrient sensing, insulin signalling and overall energy metabolism, with each having varying levels of pleiotropy. Although several *Drosophila* mutants met the criteria for enhanced growth, many exhibited associated fitness costs, most commonly reduced adult longevity and increased sensitivity to starvation or oxidative stress. However, these performance trade-offs were considered acceptable given the intended use of comparable BSFL mutants for mass rearing with excess feed and the goal of harvesting larval biomass. While some evidence suggests that these genes may

exhibit synergistic effects when mutated together in BSF, the consequences of replicating the *Drosophila* mutant phenotypes in this species remain entirely unknown. Potential outcomes of multiplexing could range from enhanced growth profiles to limited growth, the accumulation of fitness costs, or even no observable change in phenotype.

With no single target appearing more promising than the others, and limited resources to edit all six genes in separate lines, it was decided to target each growth-regulatory candidate simultaneously using MAD7 by encoding them on a polycistronic MAD7 crRNA array. The genes included in the larval growth regulatory gene (GRG) array were identified through protein BLAST analysis of orthologs in *D. melanogaster* and other insect species where orthologues could be found. For each gene, the previous steps of multiple protein alignment, identification of conserved protein regions, and selection of suitable target sites using CHOPCHOP were repeated. Details of the BSF orthologs of the larval GRG candidates, inferred from studies in *Drosophila*, are presented in **Table 2.**

3.2.3 – Selection of Pol III RNA Promoters for Guide RNA Expression

Following selection of targets for GE, the *pBac* CRISPR RNA expression plasmids could be designed. As a non-model insect of only recent attention, no transgenic expression constructs had been described in BSF at the time of experimentation, with no indication regarding which regulatory promoter elements would be suitable in the species. Previous studies in *A. gambiae* and *C. capitata* have successfully employed the *D. melanogaster* U6-3 (U6:3) promoter for ubiquitous and constitutive expression of CRISPR gRNAs, and so was selected as a promising candidate for inclusion with CRISPR toolkit plasmids in this study[474,477].

In addition to U6:3, the BSF reference genome was mined for potential endogenous U6 snRNA genes for promoter isolation, with the rationale that endogenous regulatory elements were more likely to recapture endogenous behaviours and lead to more stable and efficient expression[500]. Using the *D. melanogaster* U6 snRNA transcript sequence as a query, two identical U6 transcripts were identified in the BSF reference genome, although 8 other annotated U6 genes were also found by gene description. Rather than isolate each gene, the two U6 snRNA transcripts that shared perfect homology to the *D. melanogaster* U6:3 transcript were selected for promoter isolation. The two BSF-endogenous U6 genes were

Gene ID *(H. illucens)*	Gene symbol	Gene description	Gene location	Transcript length (nt)	Protein length (aa)	Target residues (aa, 5'-3')	Target [PAM] and spacer sequence (5'-3')	Gene ID *(D. melanogaster)*	Reference (DOI):
119650645	***Maf1***	Repressor of RNA polymerase III transcription Maf1	Chr-3, NC_051851.1 (137502757..137525694, complement)	3876	234	92 - 100	**[TTTC]GCGTAAGACGTTATTCTACTT**	3354925	10.1073/pnas.1113311109
119648081	***ImpL2***	Ecdysone-inducible gene L2	Chr-2, NC_051850.1 (172834269..172893195)	768 - 789	255 - 263	62 - 70	**[TTTC]GCGTAAGACGTTATTCTACTT**	38513	10.1186/jbiol72
119652713	***Cereblon (CRBN)***	Protein cereblon	Chr-3, NC_051851.1 (44745511..44758124, complement)	1695 - 1757	465 - 480	86 - 94	**[CTTC]CTACCGAACATTCATATCTAG**	41230	10.1074/jbc.M116.757823
119656298	***Sirtuin 6 (sirt6)***	NAD-dependent protein deacetylase Sirt6	Chr-4, NC_051852.1 (45382300..45391125, complement)	1346	351	65 - 69	**[TTTC]AAATGTGCAGGGGACCCAACG**	41254	10.1111/acel.13576
119650889	***TBC1d7***	TBC1 domain family member 7	Chr-3, NC_051851.1 (14019648..14028396, complement)	1045	293	46 - 54	**[TTTC]ACAGTTCCTACAGCTGAAAGG**	42869	10.1083/jcb.201706027
119653264	***Piezo***	Piezo-type mechanosensitive ion channel component	Chr-3, NC_051851.1 (152349757..152461421)	7134 - 7896	2378 - 2632	417 - 425	**[TTTG]TGTTGTTAATATGGGCGAACG**	34112	10.7554/eLife.63049

Table 2: Larval growth regulatory gene (GRG) candidates identified for mutagenesis during this study. All BSF genes were identified *via* protein multiple sequence alignment against orthologous genes from *D. melanogaster* or *C. capitata*. Gene information, descriptions, locations and comparative Gene IDs were identified from the NCBI database. crRNA target sequences were identified using the online tool CHOPCHOP[468].

from then defined as U6:1 (*LOC119650510,* chromosome 2) and U6:2 (*LOC119650510,* chromosome 6).

As Pol III promoters are generally less elaborate than Pol II promoters, previous studies have indicated that a region of 300~400 bp immediately upstream of the U6 snRNA transcribed sequence is generally sufficient for capturing a high level of gRNA expression and activity[500]. A region of 450 bp immediately upstream of the U6 snRNA transcript was amplified from the genome of the T4 line, with sequence verification confirming the presence of both U6:1 and U6:2 promoters. The nucleotide sequences of the suspected promoter regions of U6:1, U6:2 and the *D. melanogaster* U6:3 were then aligned for comparison and identification of possible regulatory sequences.

In the time since this analysis was first performed, however, a study has been released that describes a similar identification of endogenous U6 promoters in BSF, in that case used for the transgenic expression of shRNA for applications in RNAi[422]. Kou and colleagues (2022), working from an earlier genome reference than the one employed in this study, identified 3 BSF U6 snRNA transcripts with identical homology to the *D. melanogaster* transcript, calling them U6:1, U6:2 and U6:3; to avoid any confusion, these will be referred to as U6:1 (Kou *et al*.), U6:2 (Kou *et al*.) and U6:3 (Kou *et al*.) for the rest of this section, while the genes identified in the current study will be referred to by their gene or species names, U6:1 (*LOC119650510*), U6:2 (*LOC119660634*) and *Dm*_U6:3.

Multiple nucleotide alignment was later performed for the three U6 promoters utilised in the current study and the three identified by Kou and colleagues (**Figure 11**). Despite not being able to find the sequences described by Kou and colleagues in the current reference genome, alignment suggests that the U6:1 and U6:2 identified in the current study are likely analogous, with differences in polymorphisms explaining the difference between the two reference sequences. In the study performed by Kou and colleagues, the authors noted a high, but essentially identical, level of RNAi activity when using the U6:1 (Kou *et al*.) and U6:2 (Kou *et al*.) promoter isolates, and a present but significantly lower activity when using U6:3 (Kou *et al*.). From multiple sequence alignment, the two promoter regions of U6:1 (Kou *et al*.) and U6:2 (Kou *et al*.) shared a high identity with each other (83.0%), while the U6:1 (*LOC119650510*) promoter had a very high identity to both: 94.1% identity to U6:1 (Kou *et al*.), and 94.3% identity to U6:2 (Kou *et al*.).

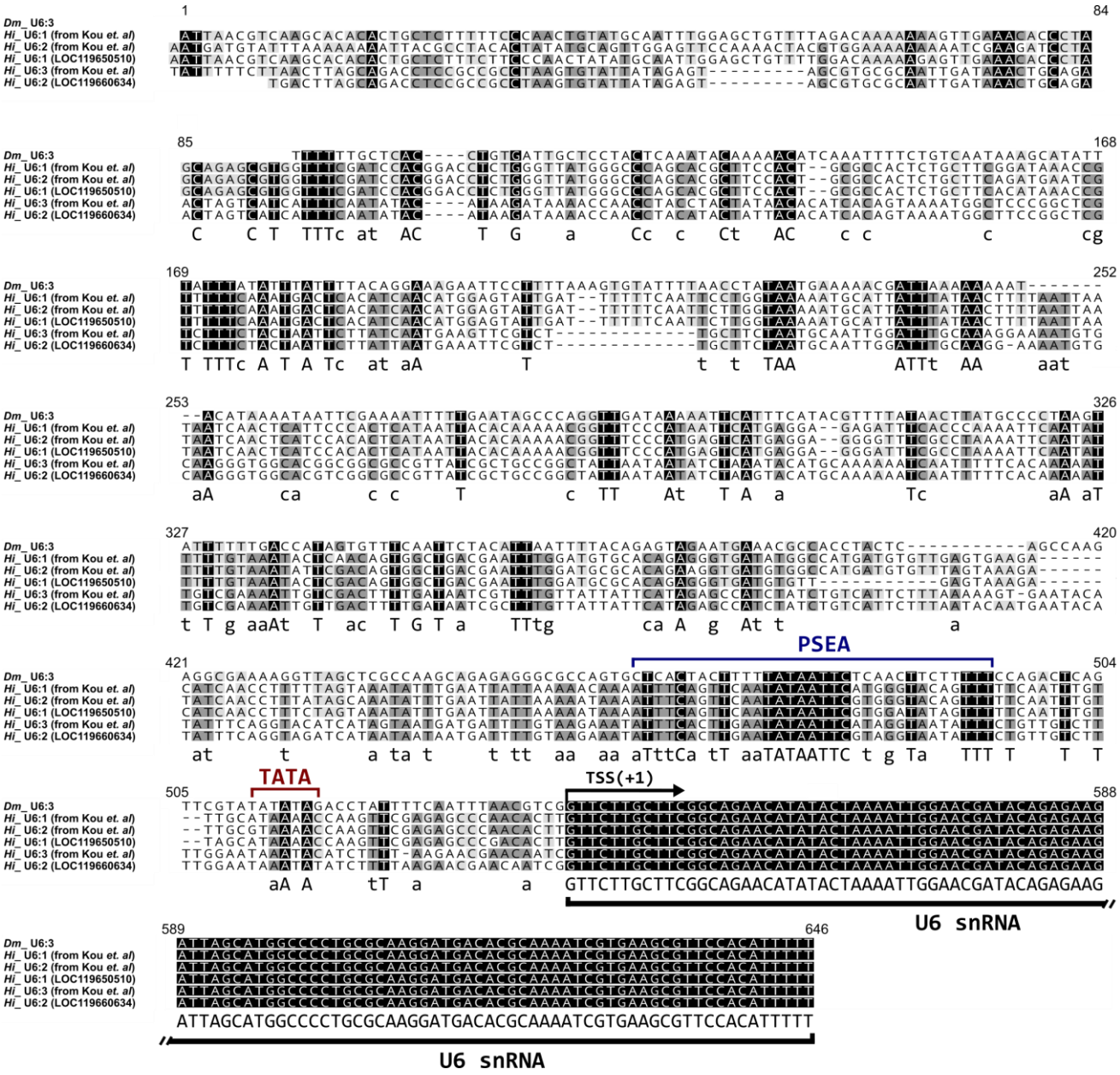


Figure 11: Multiple nucleotide alignment of *Dm*_U6:3, the two BSF-endogenous U6 promoters identified in this study, and the three BSF U6 promoters identified in Kou *et al.* (2022). High conservation is seen in all BSF U6 promoter isolates, particularly in the suspected proximal sequence element A (PSEA) and TATA box core promoter regions. Note that while *Dm*_U6:3 encodes a classical TATA box motif (TATAWAW), none of the BSF-endogenous promoters encode this exact motif.

Incidentally, U6:3 (Kou *et al*.) had comparatively low identities to both U6:1 (Kou *et al*.) (46.6%) and U6:2 (Kou *et al*.) (45.0%), perhaps going to explaining the lower activity seen with this promoter when used in the RNAi experiments. However, the U6:2 (*LOC119660634*) identified in the current study also has a very high identity to U6:3 (Kou *et al*.) at 94.8%. Taken together, it would seem likely that U6:1 (*LOC119650510*) identified here is functionally

the same gene as the U6:1 (Kou *et al*.) and U6:2 (Kou *et al*.) identified previously, while U6:2 (*LOC119650510*) is very likely the equivalent of U6:3 (Kou *et al*.). Although empirical data would be required to verify this assumption, it seemed logical that U6:1 (*LOC119650510*) might produce a high expression of gRNA when applied in BSF, while U6:2 (*LOC119660634*) would theoretically lead to a lower activity.

Interestingly, the 5 BSF endogenous U6 promoters all appear to differ significantly in terms of identity to *Dm*_U6:3, while sharing high identity to each other in several conserved regions across the promoter. Kou and colleagues performed a similar alignment in their study, although only including their own identified BSF-endogenous promoters, in an attempt to identify regions of conservation between the three genes. In addition to smaller regions found throughout the promoter, two highly conserved regions were identified; the proximal sequence element A (PSEA), around +78 bp upstream from the U6 snRNA transcription start site (TSS), and a TATA box motif found +28 bp upstream of the TSS.

When comparing all 6 U6 genes described in this study, the BSF endogenous isolates have close to an identical sequence at both the PSEA and TATA box regions, providing further evidence that these are the critical core promoter elements for initiating U6 expression. Interestingly, while *Dm*_U6:3 has the classical TATA box sequence motif (TATAWAW), none of the BSF promoters encode this motif at the expected TATA box region, despite it being highly conserved in insects. Recognised in Kou *et al.,* the authors comment that the TATA box motif may not be essential for a functional U6 promoter in this species, having demonstrated activity of the TATA-less U6 promoters in earlier RNAi experiments.

3.2.4 – Guide RNA Expression Construct Assembly

Following the identification of the U6:1, U6:2 and U6:3 promoters, the rest of the base CRISPR RNA *pBac* transgenesis construct was designed. This consisted of a base *pBac* vector containing two *pBac* inverted terminal repeat (ITR) sequences flanking a central cassette, in which a *gypsy* insulator element and two inverted *attP* sites were also present. The gypsy insulator was included to act as a chromatin domain boundary and a barrier against enhancer-promoter interaction, effectively neutralising much of the positional effect on transgene expression influenced by the site of genome integration[501]. Although not specifically employed in this study, the *attP* sites were added to allow for exchange of the

central transgene cassette with other sequences following integration, as part of the phiC31 integrase system[502].

Into this base *pBac* vector, both fluorescent marker genes and a CRISPR RNA expression genes were subsequently cloned to yield several “entry” vectors (**Figure 12A**). Firstly, the CRISPR RNA vectors were constructed with a fluorescent marker gene comprising of the eye-specific constitutive marker *3x3P*, the DsRed protein and a short SV40 terminator[477]. The *3x3P* promoter was later exchanged for *Opie2*. Overhanging primers and a 2-step PCR protocol was then used to isolate 450 bp of sequence immediately upstream of the U6 snRNA transcript in the U6:1 and U6:2 genes (**Fig. 12C**). The U6:3 promoter was isolated from a previously constructed *pBac* vector used for transgenesis in *C. capitata*[477]. Using different primer sets, the U6 promoter fragments were generated with ends corresponding to the Cas9 gRNA cassette (**Fig. 12B**). The assorted fragments were combined *via* DNA assembly to yield the three entry Cas9-compatible gRNA vectors, differentiated by the U6 promoter used (**Fig. 12A**).

To facilitate the easy and economical of new gRNA-encoding plasmids, the entry vectors were designed with a placeholder spacer region encoding two inverted BsaI restriction sites, allowing for type IIS or “scarless” cloning[503]. Essentially, the entry plasmid was first digested with BsaI to expose overhanging sequence leading from the end of the U6 promoter to the start of the gRNA scaffold, with different variable sequences able to be ligated into the space as short, annealed oligonucleotides with overhangs complementary to the vector (**Fig. 12D**). While ultimately the testing of multiple different gRNA sequences fell out of scope for this study, the type IIS entry vector system was highly efficient and could be easily employed for the rapid construction of gRNA expression plasmids in later BSF studies.

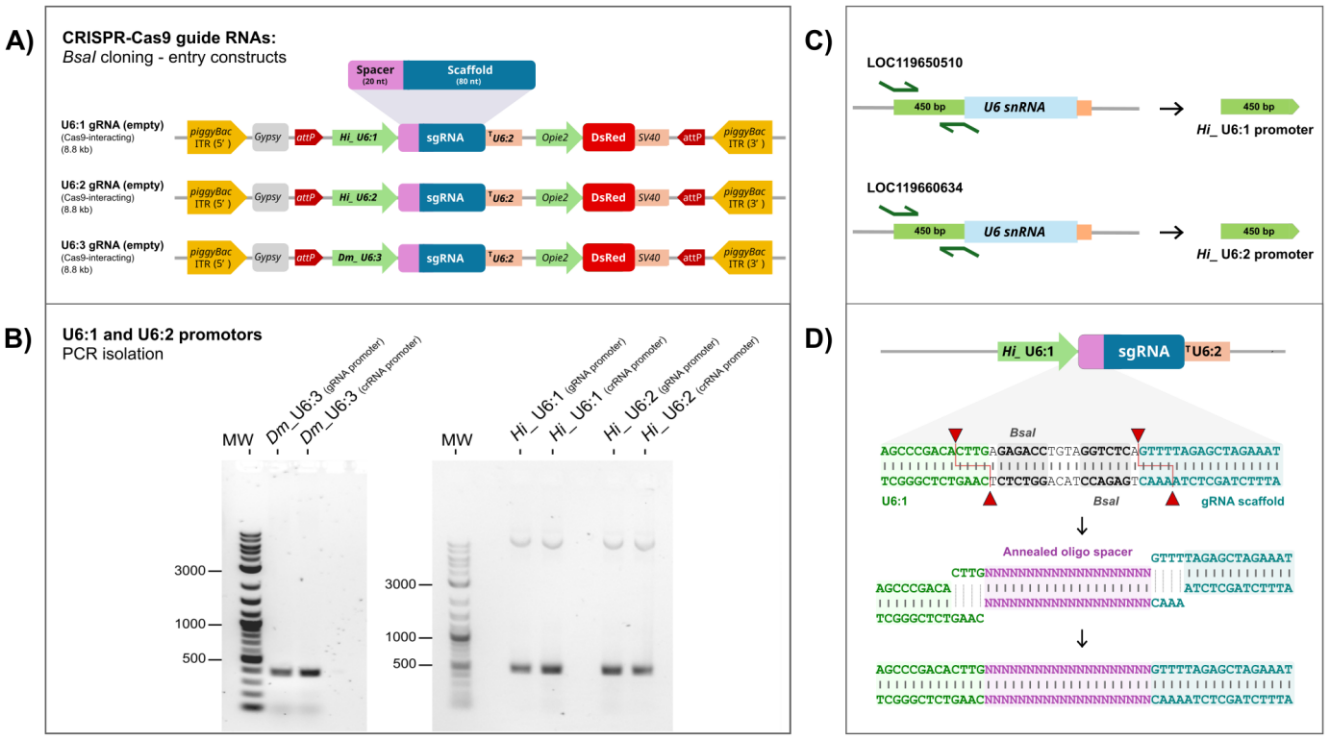


Figure 12: Entry gRNA expression constructs and type IIS cloning. A) Entry *piggyBac* donor constructs were designed for the rapid cloning of target-specific gRNA spacer sequences into a gRNA expression cassette under the control of the *H. illucens* endogenous U6:1 and U6:2 promoters, and the *D. melanogaster*-derived U6:3 promoter. The gRNA constructs all employed an additional *Opie2*-DsRed fluorescent marker gene, replacing the 3xP3 promoter included in the original design following doubts regarding its activity in BSF. **B)** PCR was used to isolate the 450 bp *Dm* U6:3 promoter fragment from a plasmid source, while the 450 bp U6:1 and U6:2 promoters were both amplified from a BSF gDNA extract. The 1 kb Plus DNA Ladder (NEB) was used as a MW marker. **C)** The endogenous BSF promoters were identified as 450 bp immediately upstream of the start of the U6 snRNA sequence in each gene. Overhanging primers were used to facilitate the cloning of the promoter fragments into the gRNA or crRNA entry vectors. **D)** Type IIs cloning was utilised to facilitate rapid cloning of gRNAs specific to different target sequences. The three gRNA BsaI entry vectors each featured dual BsaI restriction sites in place of the targeting gRNA spacer sequence. Upon restriction digestion, the short BsaI spacer fragment was removed during DNA column cleanup to leave the exposed ends of the RNA promoter and the start of the gRNA scaffold sequence, with each end having non-complementary 4 nt overhangs. A new targeting spacer sequence, consisting of two complementary 24 nt oligonucleotides annealed together, was ligated in-frame and in the correct orientation via a pair of 4 nt overhangs complementary to the exposed promoter and gRNA scaffold ends in the BsaI-digested vector.

While the Cas9-gRNA expression constructs were simple to construct, the MAD7 multiplexing crRNA constructs required more careful design. The three entry crRNA plasmids, each encoding a different U6 promoter, are identical to the Cas9 gRNA plasmids aside from the central array section (**Figure 13A**). Due to concerns that long polycistronic arrays would be unstable during Pol III transcription, or that they might form complex

secondary structures, tRNA sequences were encoded to flank both ends of the crRNA transcript; this technique was found to increase the stability and activity of long crRNA transcripts in studies with Cas12a[403]. Following transcription of the tRNA-flanked crRNA, host nucleases and MAD7 are expected to jointly process the transcript into the individual MAD7 gRNAs for GE (**Fig. 13C**).

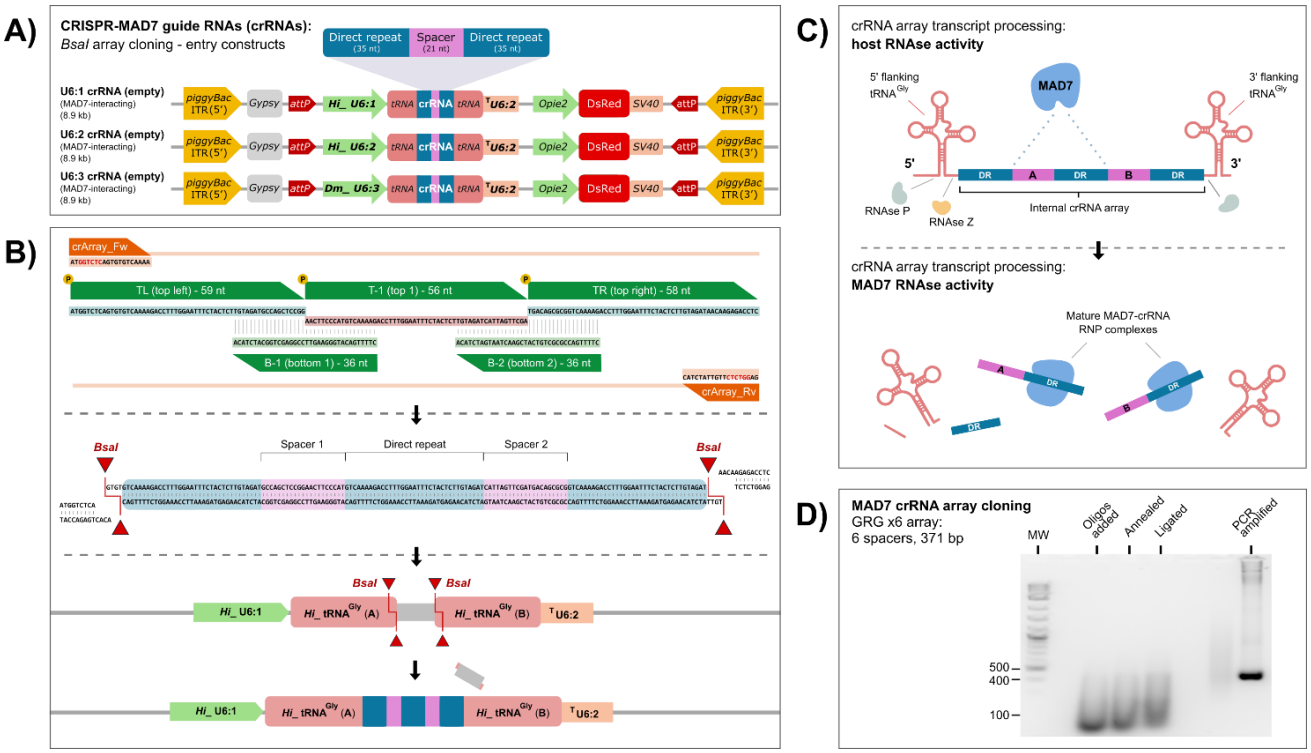


Figure 13: Entry crRNA expression constructs and type IIS cloning. A) Similar to the Cas9-specific gRNA entry constructs, the MAD7-specific crRNA entry constructs were designed using the three Pol III promoter fragments previously isolated by PCR. The promoters were cloned upstream of an RNA expression cassette featuring two tRNA sequences flanking a central sequence of two MAD7 direct repeats, themselves separated by dual BsaI restriction sites to facilitate type IIs cloning of new spacer sequences. All crRNA constructs featured the *Opie2*-DsRed fluorescent marker gene. **B)** Overview of the type IIS assembly strategy for generating polycistronic MAD7 crRNAs with multiple spacer sequences. Long oligonucleotides encoding the central polycistronic crRNA array sequence were phosphorylated and annealed to form a partial double-stranded fragment, and were then ligated together. Outer PCR primers were used to amplify the ligated cassette and introduce inverted BsaI sites at the DNA ends for type IIS cloning. The amplified crRNA cassette and the entry crRNA vector were both digested with BsaI, with the cassette then ligated between the two encoded tRNA sequences in the crRNA gene. **C)** A combination of host RNases and MAD7 process the tRNA-flanked crRNA array transcript, removing the terminal tRNAs and liberating the individual gRNAs for MAD7 targeting. **D)** The sequential reactions for cloning the GRG array using the oligo cloning protocol described in B), showing the process of annealing, ligating and PCR amplification of the central array structure prior to BsaI digestion.

Early on this study, the decision was made to encode the six GRG genes on a single crRNA array for multiplexing, along with other crRNA designed to encode multiple targets against the *w* gene in BSF and *D. melanogaster*. To mediate rapid cloning, a cheap and rapid crRNA oligo / type IIS cloning method was designed and adapted from two previously described methods (**Fig. 13B**)[403,504]. This protocol, which could be completed in less than a day, was employed throughout the study for the efficient cloning of complex MAD7 arrays, and similarly would be suitable in the design of future multiplexing experiments.

Following construction of the 2 types of entry plasmid, oligo and type IIS cloning methods were subsequently employed for the insertion of the Cas9 and MAD7 variable sequences for the mutagenesis candidate genes identified earlier. At this point, it was determined that the long-life cycle of BSF and unpredictability in the ability to generate transgenics, the CRISPR toolkits should be subjected to a preliminary analysis for activity in insect S2 cell culture. This would provide an opportunity to assess the relative strengths of the U6 promoters in a GE context, and confirm if MAD7 has activity with the designed tRNA-flanked crRNA constructs.

As the S2 cell line is originally derived from *D. melanogaster,* the *w* gene was selected as a suitable target for analysis of GE activity in cell culture, with editing needing to be confirmed on the sequence level in this system rather than *via* a visible phenotype[505]*.* The structure of the *D. melanogaster w* gene was analysed, with the sequence of *w* being confirmed in genomic DNA (gDNA) extraction taken from a passage of S2 cells (**Figure 14**). In a similar strategy as employed for BSF *w*, Cas9 and MAD7 targets were designed to overlap at one locus, while two other MAD7 crRNA arrays was also produced that encoded two and three target sites in total. As neither the *3xP3* nor *Opie2* insect tissue-specific promoters currently employed for marker expression in the CRISPR RNA-expressing constructs would allow for detectable fluorescence in cell culture, the marker cassette was replaced with a *Hr5IE1*-DsRed marker in the CRISPR RNA S2 transfection plasmids. Ultimately, three gRNA-expressing targets were produced for the same Cas9 target, encoding either the U6:1, U6:2 or U6:3 promoter, while the three produced MAD7 constructs all encoded the U6:3 promoter and differed instead by the number of targets.

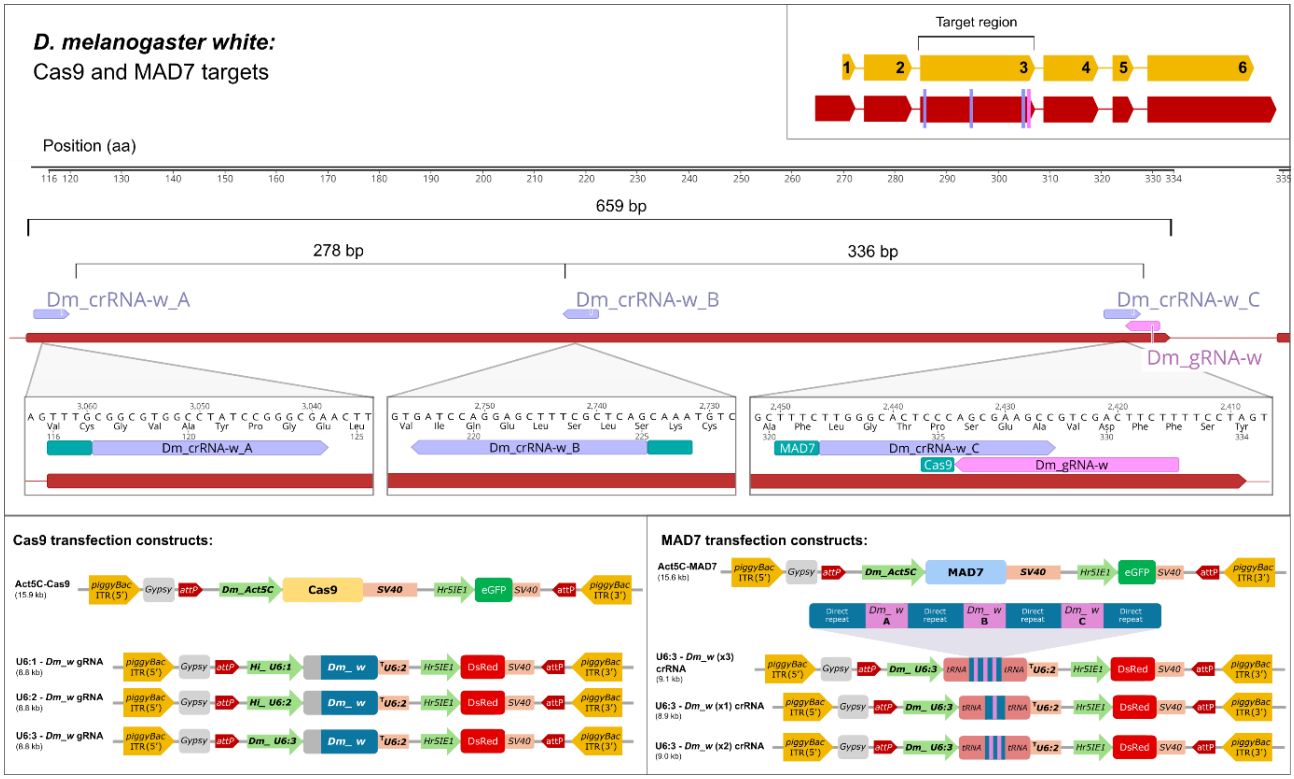


Figure 14: The *D. melanogaster w* gene structure, Cas9 and MAD7 target sites and CRISPR RNA expression constructs produced for S2 cell transfection. Three target sites (A-C) were identified across the third exon of the *white* coding sequence for cleavage with MAD7. A single Cas9 target site overlapping with the terminal MAD7 crRNA target sequence was also identified, allowing for comparison of enzyme activity at an adjacent genomic locus. As neither the *3xP3* nor *Opie2* fluorescent marker promoters were expected to work effectively in cell culture, the *Hr5IE1* promoter was cloned upstream of the DsRed marker in the gRNA and crRNA entry constructs for visualisation of transfected cells. Type IIS cloning was then employed with the entry constructs to produce three Cas9-specific gRNA constructs, each targeting the single identified gRNA target site and differing only by their Pol III RNA promoter. MAD7 crRNA constructs were also produced for each of the three identified single target sites, with each construct utilising the same U6:3 promoter. The Act5C-Cas9 and Act5C-MAD7 nuclease expression constructs were paired with their respective RNA expression constructs to complete the CRISPR toolkits used in all S2 cell culture experiments.

In total, 11 CRISPR RNA constructs were employed in this study (**Figure 15**). In addition to the six produced for the S2 transfection experiments, three Cas9 gRNA-expressing constructs were produced against the *w* target in BSF, utilising the three U6 promoters. Both under the U6:3 promoter, one crRNA was produced to target the two *w* MAD7 targets in BSF, and another encoded the array for the six GRG targets.

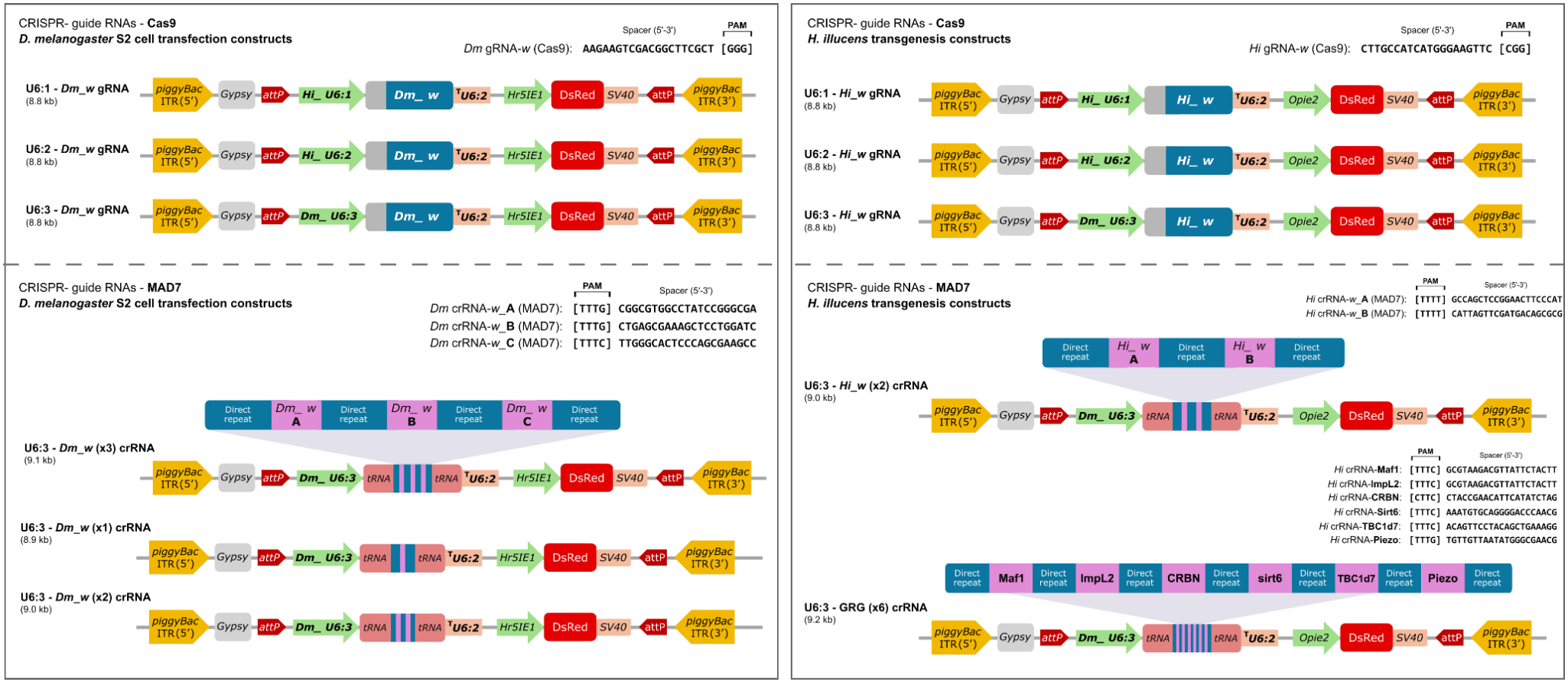


Figure 15: Construct schematics for the CRISPR RNA expression plasmids used in S2 cell culture and BSF transgenesis experiments. The general plasmid structure, Pol III promoter, spacer RNA sequence and context for all CRISPR RNA expression constructs used in this study are outlined in this figure. The central section of alternating spacer and direct repeat domains for the MAD7 multiplexing constructs has been enlarged relative to the rest of the construct to illustrate their order within the RNA transcript. All RNA expression constructs destined for use in *D. melanogaster* cell culture featured the *Hr5IE1* promoter for DsRed marker expression, while all designed for use in BSF transgenesis retained the *Opie2* promoter.

3.2.5 – Identification of Germline-Specific Regulatory Elements for CRISPR Nuclease Expression

The CRISPR nucleases Cas9 and MAD7 were intended for germline-restricted expression, which could be best achieved through identifying endogenous regulatory elements that could be employed for the CRISPR nuclease constructs. The ideal gene should show significantly enriched levels of expression in the germline tissues of BSF, and comparatively lower levels in the remaining somatic tissues. A series of candidate genes that might be suitable for controlling germline expression in BSF was first identified from literature sources of similarly-regulated CRISPR systems employed for techniques such as gene drive in insects: *beta 2 tubulin (b2-tub), Bicaudal-C (BicC), Bicaudal-D (BicD), Exuperantia (exu), nanos (nos), suppressor of defective silencing 3 (sds3), shut-down (shu), vasa (vas)* and *zero-population-growth (zpg)*[474,477,500].

Potential BSF homologues were then identified for each gene *via* protein BLAST against *D. melanogaster, A. gambiae* and *C. capitata* variants. Interestingly, two genes in BSF were identified as containing a *nanos* RNA binding domain, although both were unannotated, and so they were annotated as *nos_1* and *nos_2*. Following identification of potential germline-gene candidates, RNA sequencing was employed to compare their relative expression levels in germline and carcass tissues (**Figure 16**). Male testes and female ovary tissues were dissected from mature adult BSF, with germlines from several adults pooled to gather sufficient tissue for total RNA extraction. Following removal of the germline tissues, the remaining adult carcass tissues were also harvested to serve as the non-germline comparison tissue. Following RNA extraction, total RNA from testes, ovaries, male carcasses and female carcasses were submitted for mRNA-enrichment and RNA sequencing.

In male flies, the number of genes in testes tissue that were upregulated, downregulated or not significantly different were roughly equal, with 7 of the 10 candidate germline genes being significantly upregulated in the male germline (**Fig. 16A**). Although germline-upregulated genes represented a much lower proportion of the significant differentially expressed genes in females, 8 of the expected germline genes were found to be above the threshold for enrichment in ovaries (**Fig. 16B**). Only *exu, nos_1*, *nos_2*, *vas* and *zpg* were significantly upregulated in the germline of both sexes.

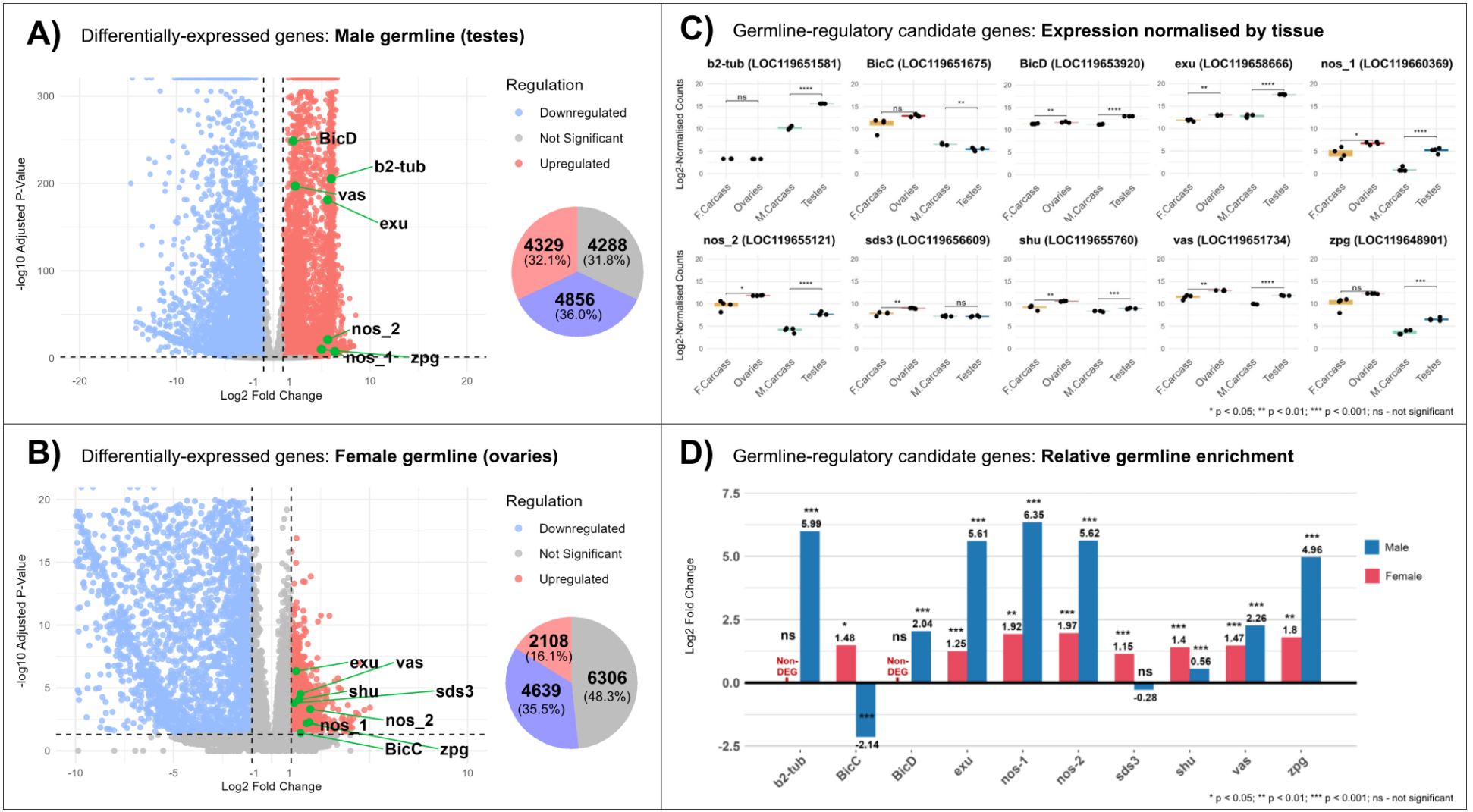


Figure 16: Germline tissue enrichment and differential gene expression of endogenous germline regulatory candidate genes in BSF. to determine the differentially expression of germline regulatory genes. A-B) Volcano plots of differentially-expressed genes (DEGs) for male and female germline tissue extracts when assessed against non germline tissues of the same sex. Germline tissues (testes or ovaries) were harvested for RNA extraction from unmated adult BSF three days post-pupal emergence to allow sexual organs to develop to maturity. Adult testes were isolated in pairs from BSF abdomens, with the accessory glands, external genitalia and fat body removed. Tissues from multiple adults of the same sex were pooled (testes from 7 males pooled, ovaries from 5 females pooled) to increase the amount of starting material available for RNA extraction. Simultaneously, samples of adult carcass tissue, consisting of the adult body

remaining after removal of the wings and traces of germline tissue, were harvested for each sex as a non-germline tissue control. Total RNA extraction, cDNA library preparation, RNA-seq, and DEG analysis were then performed. Ten candidate genes expected to show a pattern of germline-enrichment based on literature review were assessed for their differential expression in BSF germlines, with significantly upregulated genes marked in green. **C)** Normalised read counts of the identified germline regulatory candidate genes for germline and non-germline tissues in each sex. Higher Log2 normalised counts indicate a higher relative gene expression across all samples, with pairwise comparisons made between germline and non germline tissues of the same sex indicating a significant difference in expression level. **D)** The degree of germline regulatory candidate gene enrichment in germline tissues, depicted as fold increase in expression over non germline tissues for both sexes. The significance value indicates the degree to which the germline is enriched for expression over non-germline tissues.

While each of these genes could potentially be employed for the germline-regulation of CRISPR nucleases in BSF, ultimately *nanos* (*nos_1*) and *vasa* (*vas*) were selected for continuation in the first set of nuclease constructs. *nos_1* was selected over *nos_2* due to sharing a slightly higher identity to the *D. melanogaster nanos* protein (30.8% vs 24.4%), although both contain the *nanos* RNA binding domain and are otherwise uncharacterised. Both *nanos* and *vasa* were chosen due to their proven efficiency and more extensive characterisation in multiple other species when compared with the remaining candidates, although this should not preclude the potential use of *exu, nos_2* and *zpg* in BSF transgenesis constructs in the future.

With the germline genes identified, the *nanos* and *vasa* sequence promoter and terminator fragments were amplified from the gDNA of the T4 *wild-type* line (**Figure 17**). To recapture the regulation of the endogenous germline genes, the promoter sequence included the entire 5'UTR sequence up to the start codon, while the terminator included the 3'UTR sequence immediately following the stop codon. Other studies have found a 2 kb fragment immediately preceding the start codon, and a 1.5 kb fragment immediately proceeding the stop codon for the promoter and terminator regions respectively suitable for reconstituting germline regulated expression[474]. As the *nanos* gene encoded three different mRNA variants, the entire unspliced 5'UTR sequence was included as part of the *nanos* promoter sequence, allowing for the expression of any of the transcripts.

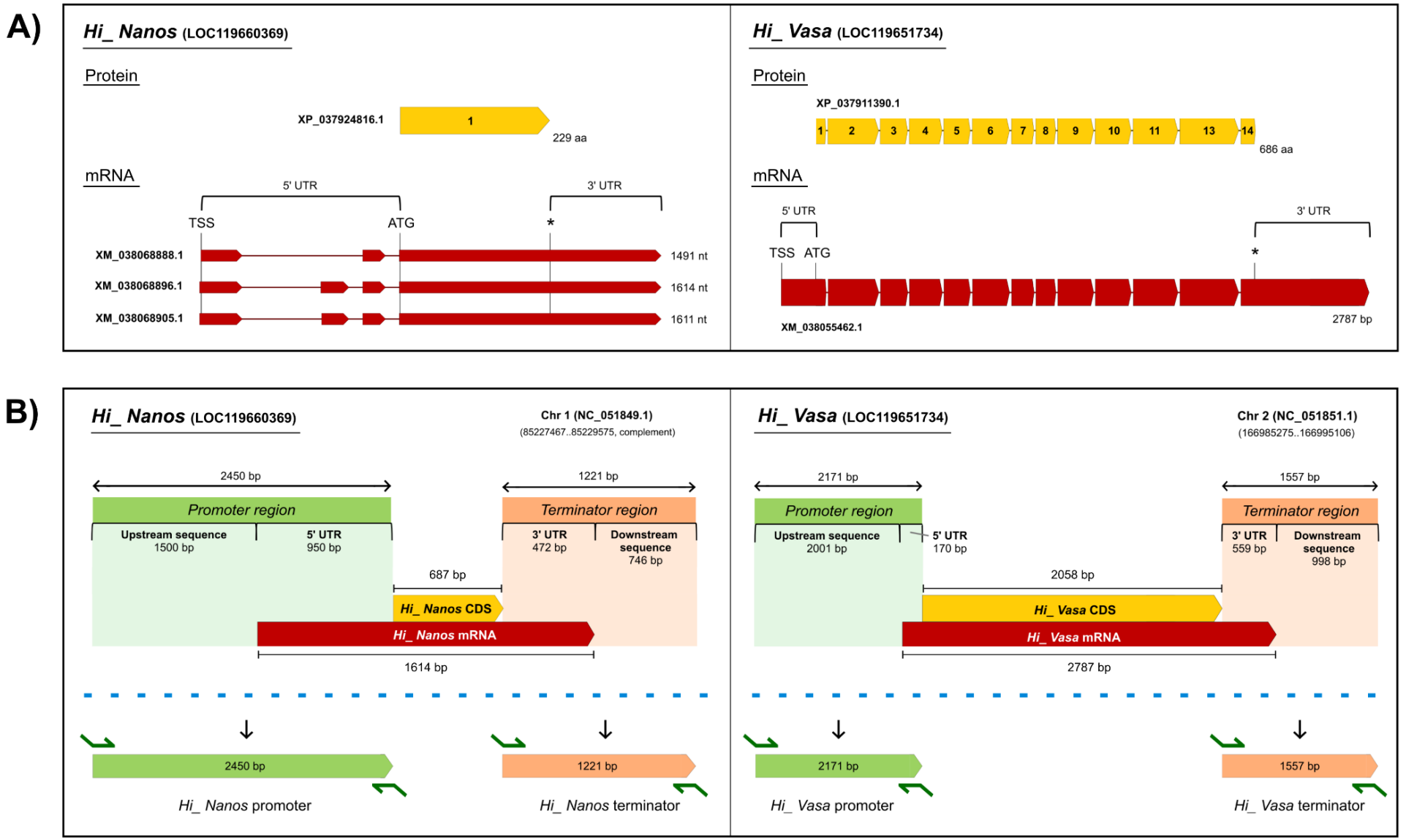


Figure 17: The protein, mRNA and gene structure of the Nanos and Vasa genes and regulatory regions identified for the cloning of the germline restricted nuclease expression constructs. A) Both Nanos and Vasa genes encode a single protein sequence, with the translated exons aligned with their relative mRNA transcripts in this depiction. While Vasa is encoded by one mRNA transcript, Nanos encodes three possible mRNA isoforms, differing only by the non-translated exons included in the 5'UTR for each transcript. The predicted transcription start site (TSS), translation start codon (ATG) and stop codon (*) for

each gene are annotated, as well as their 5' and 3' UTR sequences. Relative sequence lengths and alignments are not depicted to scale and are intended for illustrative purposes only. **B)** A structural overview and genomic loci of the Nanos and Vasa genes. The gene regulatory sequence identified as the promoter region consisted of the entire genomic DNA between the mRNA TSS and ATG, as well as ~1500-2001 bp of non transcribed sequence immediately upstream of the TSS that is assumed to harbour transcriptional enhancers and transcription factor binding sites387. To avoid biasing the endogenous alternative splicing mechanism of the Nanos promoter, the entire unspliced 5' mRNA sequence was included rather than attempting to remove intronic sequence from the 5' UTR. The regulatory sequence identified as the terminator region included the mRNA sequence from the translation stop codon to the predicted transcription termination site, as well as ~746-998 bp of downstream AT rich sequence that may enhancing termination by destabilising the binding of RNA polymerase506. The promoter and terminator sequences for each gene were amplified from gDNA by PCR, indicated by the blue line, using primer sets that included overhanging terminal sequences to facilitate cloning by DNA assembly of each component into the nuclease encoding plasmid constructs. Relative sequence lengths and alignments are not depicted to scale and are intended for illustrative purposes only.

3.2.6 – CRISPR Nuclease Construct Assembly

The CRISPR nuclease constructs were constructed along a similar basis as the CRISPR RNA constructs, with the replacement of the *3xP3*-DsRed cassette with a *Hr5IE1*-GFP marker. Both Cas9 and MAD7 were placed under the control of the *nanos* and *vasa* promoter and terminator fragments for a source of germline-regulated expression. In addition, the constitutive and ubiquitous *Act5C* promoter, isolated from *D. melanogaster*, was added into separate Cas9 and MAD7 constructs to serve as a source of constitutive nuclease expression suited for S2 cell culture. The SV40 terminator was then assembled to complete the *Act5C*-Cas9 and *Act5C*-MAD7 constructs. In total, 6 CRISPR nuclease constructs were produced: *nanos*-Cas9, *nanos*-MAD7, *vasa*-Cas9, *vasa*-MAD7, *Act5C*-Cas9 and *Act5C*-MAD7 (**Figure 18**).

In addition to the nuclease constructs, the [l]hyPBase plasmid was employed as a source of *pBac* transposase, encoding an insect codon-optimised hyperactive *piggyBac* variant expressed from the *D. melanogaster hsp70* heat-shock promoter[462]. Finally, a *D. melanogaster* polyubiquitin (*PUb)* promoter-laden construct for the dual expression of GFP and DsRed was sourced for use in BSF transgenesis[477].

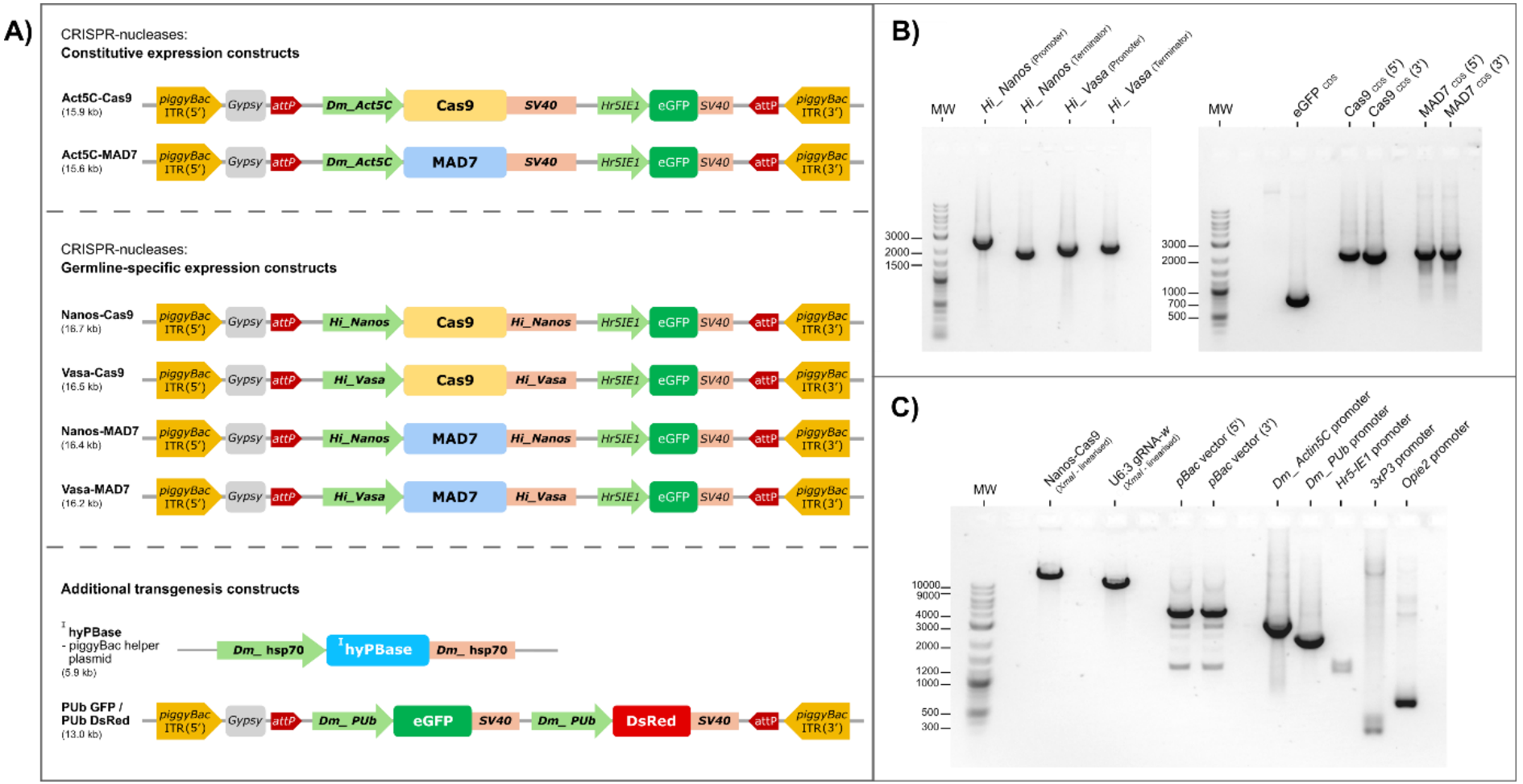


Figure 18: Construct schematic for the CRISPR nuclease constructs used in S2 cell culture and BSF transgenesis experiments. A) The Cas9 and MAD7 CRISPR nucleases were placed under the control of the *D. melanogaster* Actin5C promoter and SV40 terminator for high, constitutive expression in cell culture experiments. For BSF transgenesis, the *Nanos* and *Vasa* promoter and terminator elements were used to restrict the expression of Cas9 or MAD7 to germline tissues. All nuclease constructs encoded the *Hr5IE1*-eGFP fluorescent marker. Additionally, the *D. melanogaster* PUb-eGFP and PUb-DsRed dual marker construct was used for BSF transgenesis for validation of PUb promoter activity. A IhyPBase *piggyBac* plasmid helper construct, encoding a hyperactive *piggyBac* transposase variant codon-optimised for expression in insects and expressed under the *D. melanogaster hsp70* promoter, was co-injected with all *piggyBac* donor constructs for facilitating BSF transgenesis. **B-C)** The endogenous nuclease and promoter elements used to assemble the nuclease and RNA constructs were acquired *via* PCR amplification from BSF gDNA. Non-BSF endogenous regulatory, plasmid vector and protein coding sequences were PCR amplified from various plasmid sources to complete assembly of the final plasmid constructs used in cell culture and transgenesis.

3.2.7 – Preliminary Assessment of CRISPR Toolkit Activity in Insect Cell Culture

Prior to using the developed CRISPR constructs for transgenesis in BSF, a preliminary assessment of their relative activities was undertaken in S2 cell culture. Several constructs were produced to cause Cas9 and MAD7 directed mutagenesis in the *D. melanogaster w* gene. Following co-transfection of nuclease and CRISPR RNA-expressing plasmids, transduced S2 cells were recovered for genome isolation and sequence analysis, with Cas9 or MAD7 mutagenic activity being analysed at the *w* target sites.

Briefly, S2 cells were transiently transfected in triplicate with *Act5C*-Cas9 and either U6:1 gRNA-*w*, U6:2 gRNA-*w* or U6:3 gRNA-*w* expression plasmids (**Figure 19A**). In parallel, cells were also transfected with *Act5C*-MAD7 and either U6:3 crRNA-*wx*1, U6:3 crRNA-*wx*2 or U6:3 crRNA-*wx*3, with each crRNA plasmid encoding a CRISPR RNA precursor against 1, 2 or 3 targets in the *w* gene respectively. After 72 hours of incubation, GFP and DsRed fluorescence was detected in the transfected cell populations, indicating construct activity and a suitable transfection efficiency (**Fig. 19B**). 96 hours post-transfection, the GFP+ / DsRed+ cell population was isolated from each transfection condition by FACS, after which the enriched co-transfectant cells were harvested for gDNA isolation (**Fig. 19C-D**).

For analysis of Cas9 activity in cells transfected with the three U6 promoter constructs, the gDNA from the three triplicate populations was pooled and used as template for production of a 1.0 kb region surrounding the Cas9 target site for analysis of indel formation using the T7-endonuclease I assay (**Figure 20A**). Despite the appearance of smearing in the T7-positive conditions, no appreciable amplicon cleavage products were noted during gel electrophoresis for any of the Cas9 transfected populations, indicating a homogenous DNA pool and suggesting that minimal GE had taken place during transfection. To confirm this, the *w* amplicons were subjected to Sanger DNA sequencing, after which the recovered sequencing traces were analysed using the trace deconvolution tool DECODR, with an amplicon produced from a non-transfected S2 cell population serving for comparison (**Fig. 20B**).

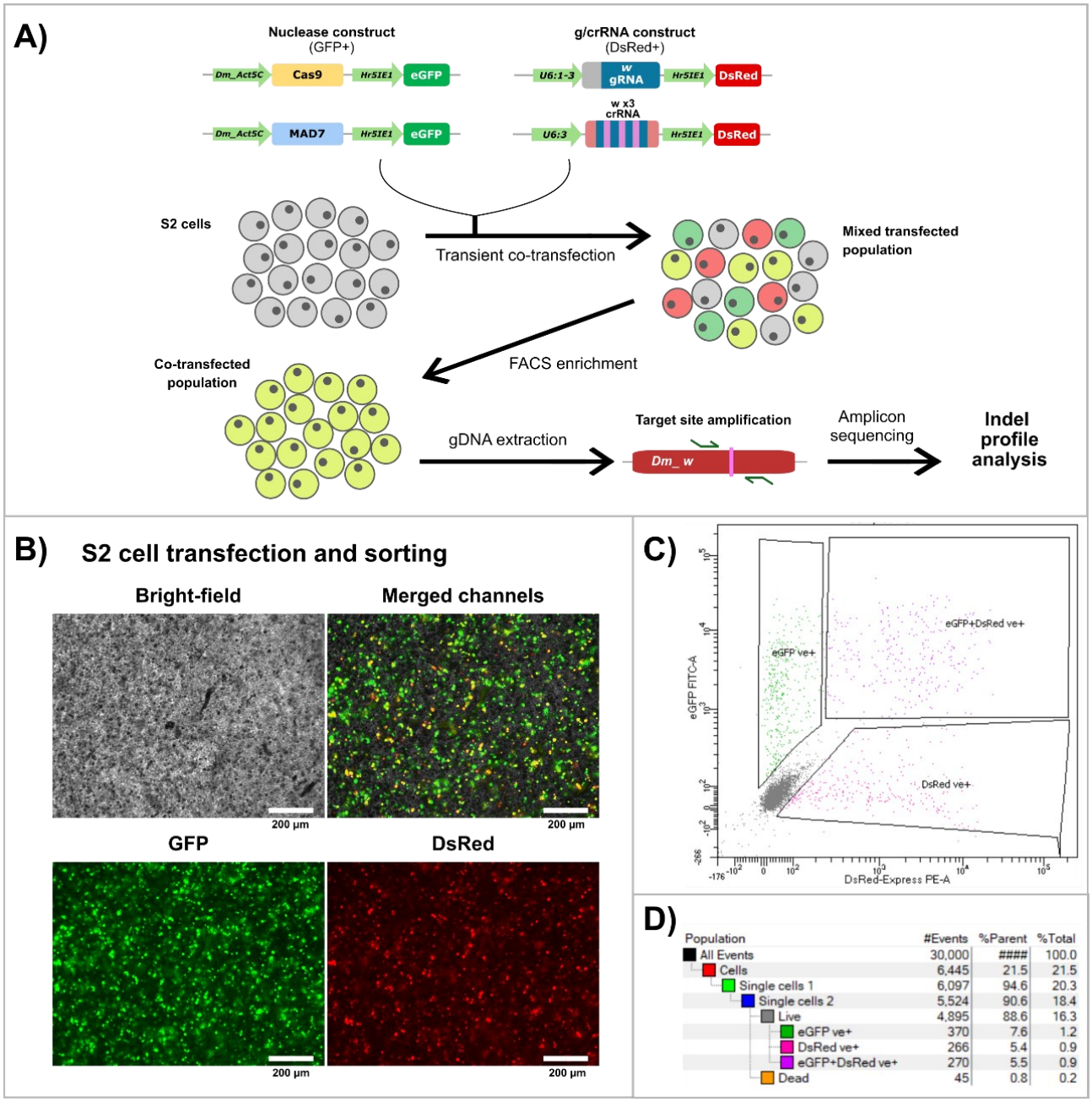


Figure 19: CRISPR toolkit S2 transfection and analysis workflow. A) Workflow of S2 transient co-transfection with CRISPR nuclease and RNA-expressing plasmids, followed by FACS enrichment of co-transfectants and amplification of the *w* target site. **B)** Representative field of view of co-transfected S2 cells (*Act5C*-Cas9 / U6:2 gRNA-*w*) with a high rate of transfection (>50%), imaged ~72 hours post-transfection. **C-D)** FACS gating settings and sorting output used to identify S2 cells expressing GFP, DsRed or both fluorophores.

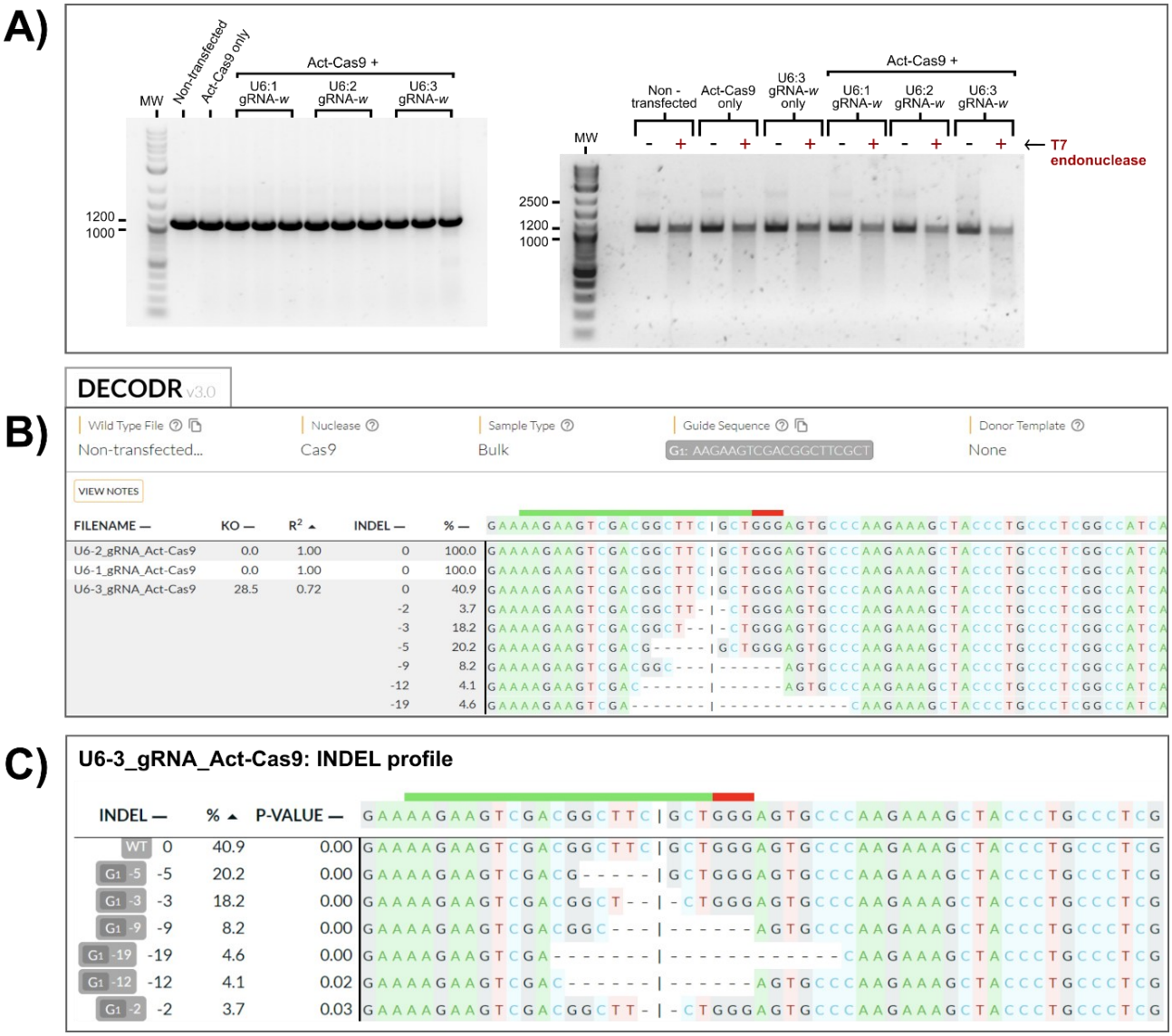


Figure 20: Cas9-transfected S2 amplicon analysis *via* T7-endonuclease I assay and trace deconvolution: A) A 1.0 kb amplicon surrounding the *w* Cas9 target was amplified from pooled gDNA extracts taken from transfected cell populations. The 1 kb Plus DNA Ladder (NEB) was used as a MW marker for each gel presented. **B-C)** DECODR sequence deconvolution indicates significant indel formation in the U6:3 gRNA-*w* transfected condition.

Interestingly, while the U6:1 gRNA-*w* and U6:2 gRNA-*w* transfected populations indicated no evidence of indel accumulation at the Cas9 target site, a significant proportion (28.5%) of the U6:3 gRNA-*w* amplicon content was indicated as edited sequence leading to gene KO, contrary to data seen in the qualitative T7-endonuclease I assay. On inspection of the indel profile sourced from the U6:3 gRNA-*w* transfected cells, only 40.9% of the amplicon was identified as non-edited *wild-type* sequence, suggesting a high rate of genome editing following co-transfection of *Act5C*-Cas9 and U6:3 gRNA-*w.* To compare the indel profiles at higher resolution, the *w* target amplicons were re-amplified from the pooled gDNA for each of the Cas9 transfection conditions and submitted for next generation sequencing using the AmpliconEZ (MiSeq 2 x 250 bp Illumina) platform (**Figure 21**).

The data in the returned FASTQ files was visualised using the CRISPResso2 pipeline[470]. The U6:1 gRNA-*w* and U6:2 gRNA-*w* transfected populations had a low baseline of polymorphism evenly distributed across the amplicon, with no obvious enrichment around the target site and a WT allele frequency (U6:1 = 98.4%, U6:2 = 97.6%) comparable to the non-transfected amplicon (WT = 97.9%) (**Supplementary figures 1-2**). This strongly suggests that the GE activity mediated by gRNA expression from these promoters is weak, if not fully absent in *D. melanogaster*. In comparison, the U6:3 gRNA-*w* transfected cell amplicon sequencing data indicated a low proportion of WT alleles in the total population (14.9%), with a clear pattern in the frequency of deletion with close proximity to the Cas9 cleavage site. Deletion was the most common indel profile in the amplicon (78.5%), followed by substitution (15.8%) and insertion (5.7%), with the second most common allele (13.1%) behind the WT allele encoding a deletion of 30 bp or more immediately upstream of the Cas9 cut site.

With U6:3 being endogenous to *D. melanogaster,* the apparent inability for the BSF-endogenous U6:1 and U6:2 to lead to any appreciable gene editing activity with the same gRNA sequence provides strong evidence for these promoters having minimal transcription activity in *D. melanogaster,* and possibly in other insect systems. Conversely, U6:3 gRNA-*w* transfected cells encountered a high activity of gene editing, indicating that the expression cassettes for *Act5C*-Cas9, gRNA and both marker genes were functional in the current construct design.

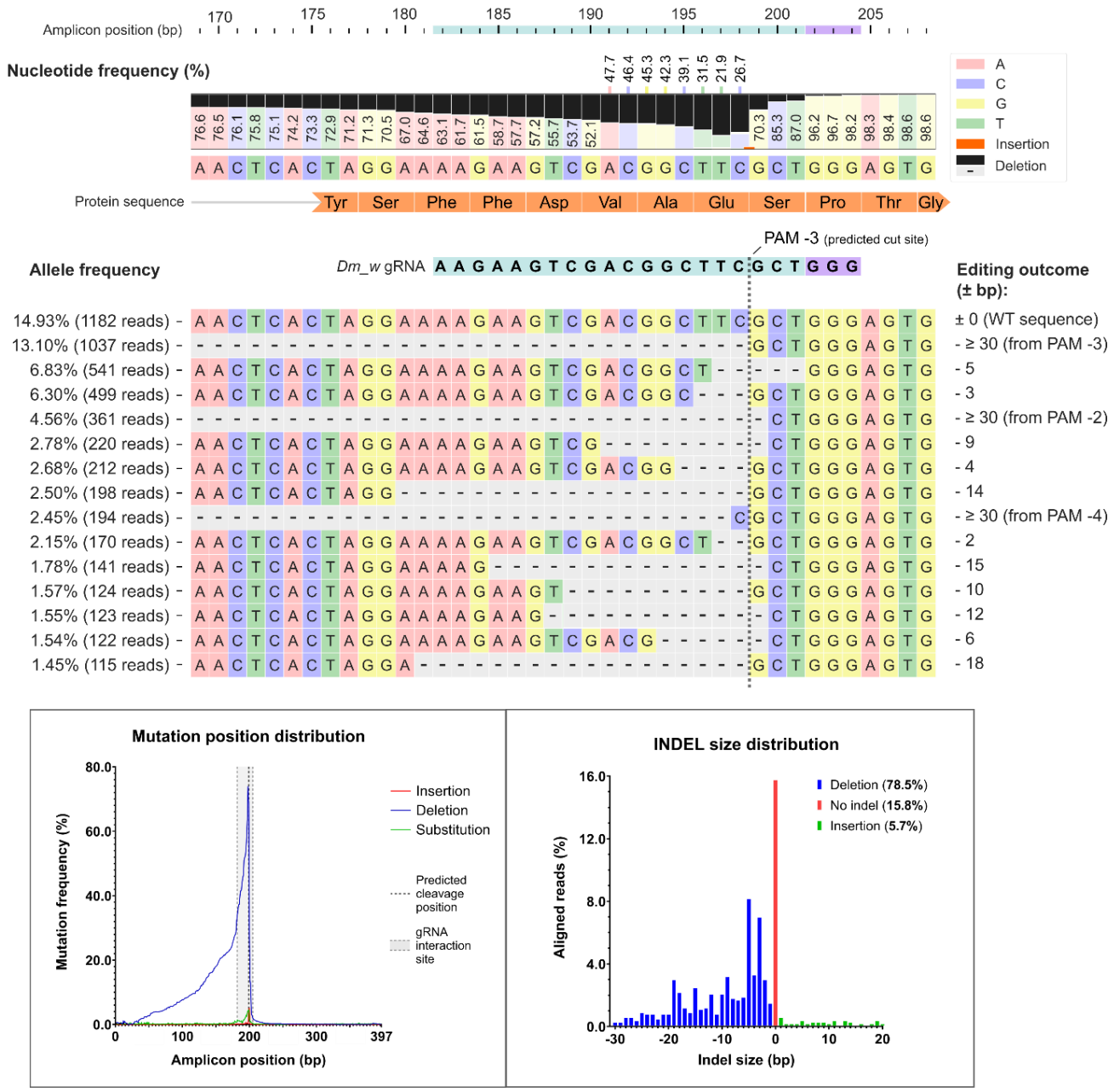


Figure 21: Amplicon sequencing analysis of *Act5C*-Cas9 and U6:3 gRNA-*w* transfected S2 cells. Mutation position distribution indicates a steady increase in deletion frequency at positions closer to the Cas9 target site, mirrored in the proportion of deleted bases in the nucleotide frequency close to the target site. The majority of alleles are associated with deletions downstream of the predicted cut site (PAM – 3).

Following the analysis of Cas9 and the U6 promoters, the S2 cells transfected with *Act5C*-MAD7 and either U6:3 crRNA-*wx*1, U6:3 crRNA-*wx*2 or U6:3 crRNA-*wx*3 were investigated for signs of MAD7 single target and multi-target GE activity (**Figure 22A**). Following production of the *w* amplicon from S2 gDNA extracts, a similar result in the T7-endonuclease I assay was noted for each MAD7 condition as seen in the experiments with Cas9, with no cleavage products detected. Occurring in non-transfected, single plasmid transfected and all co-transfected conditions, the smearing encountered in the T7-endonuclease I+ condition likely stems from the protein binding to the DNA and affecting migration in the agarose gel[506]. The amplicons were again submitted for Sanger sequencing and trace deconvolution with DECODR, alongside amplicons produced from small populations of cells previously transfected with *Act5C*-MAD7 and U6:1 crRNA-*wx*1 or U6:2 crRNA-*wx*1, although these populations were not enriched by FACS and would be expected to be enriched in the WT allele. However, none of the MAD7 transfection conditions, despite several of them encoding the U6:3 promoter, indicated anything below a 100% WT allele rate, suggesting that an issue lay with either the design of the crRNA expression constructs, or the MAD7 protein in the S2 cell system (**Fig. 22B**).

While considering the re-design of the tRNA-flanked crRNA polycistronic array sequence insert, the *w* amplicons from the MAD7-transfected conditions were cloned into a blunt vector and transformed into *E. coli* for a fine-grained analysis of potential large deletions that might be masked in by the larger sequences in the amplicon pool. With each *E. coli* colony transformed with a single amplicon copy following ligation into a plasmid vector, colony PCR was performed to identify any colonies harbouring large *w* deletions. From the plate transformed with the 1.0 kb amplicon from U6:3 crRNA-*wx*3 transfected cells, 2 out of the 24 colonies sampled appeared to show a truncated *w* sequence (~400 bp) corresponding to the expected product length seen following deletion of the entire sequence (~600 bp) between the three MAD7 sites (**Fig. 22C**).

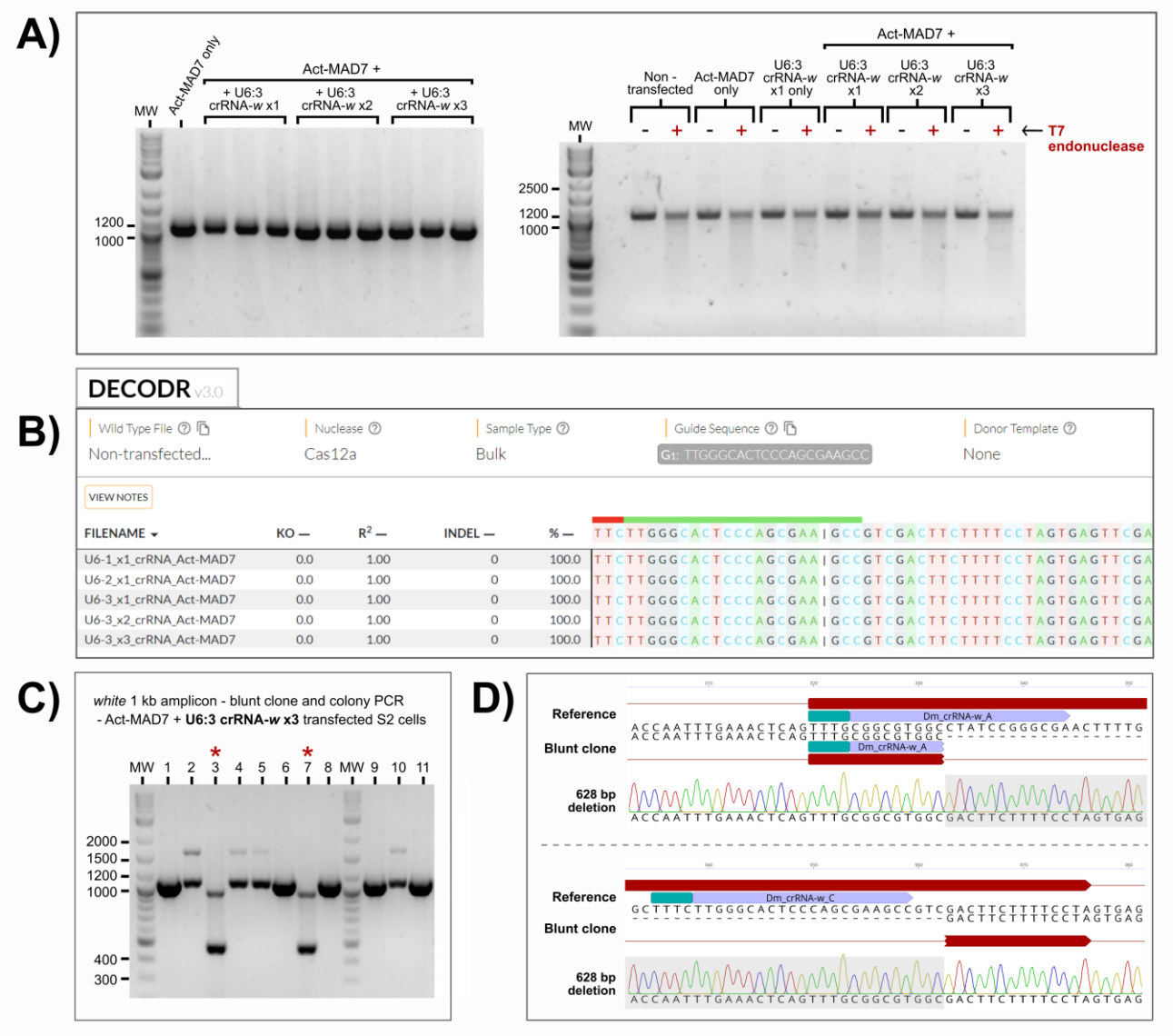


Figure 22: MAD7-transfected S2 amplicon analysis *via* T7-endonuclease I assay, trace deconvolution and blunt-clone colony PCR: A) A 1.0 kb amplicon surrounding the *w* MAD7 target was amplified from pooled gDNA extracts taken from transfected cell populations. **B)** DECODR sequence deconvolution indicates no clear indel formation in any condition. **C)** Colony PCR of *E. coli* transformed with blunt-cloned U6:3 crRNA-wx3 amplicon indicates two truncation products. Colonies producing bands larger than the 1.0 kb *w* sequence could not be identified during sequencing, possibly arising due to non-specific activity. **D)** View of the two junctions in the 628 bp MAD7 deletion mutant.

Following plasmid miniprep and sequencing, both plasmids encoded a truncated *w* sequence with an identical 628 bp deletion spanning the outermost MAD7 target sequences (**Fig. 22D**). Although the amplicon blunt cloning technique described does not allow for any reasonable quantitative estimation regarding GE efficiency, this qualitative assessment of a single MAD7 editing outcome indicates that the nuclease has some activity within the

presented toolkit, with the observed truncation spanning across two target sites being unlikely to be the result of random polymorphism or DNA polymerase error.

Although several attempts to repeat the MAD7 transfection experiments for additional data, the *Act5C*-MAD7 plasmid construct proved surprisingly difficult to purify at high yield and purity. Despite efforts to optimise and expand the culture conditions, plasmid miniprep and maxiprep yields remained much lower for this construct (~2 µg plasmid / 100 mL of culture) than all other produced *pBac* vectors, including *Act5C*-Cas9 (>270 µg / 100 mL). With re-transformation into fresh *E. coli* stocks not fixing the issue, one can only assume that *Act5C*-MAD7 is somehow a toxic insert when encoded in *E. coli,* with all plasmid preps eventually being pooled for preferential experiments in BSF transgenesis.

3.3 – Conclusion

Following identification of target pigmentation and larval growth regulatory targets, a plasmid-encoded CRISPR transgenesis toolkit was produced with promoters and terminators derived from the BSF genome. The CRISPR nucleases Cas9 and MAD7 are expected to show a germline pattern and distribution of expression, following the isolation of the *nanos* and *vasa* promoters identified *via* germline RNA sequencing. The gRNA and crRNA elements were designed to be expressed under the BSF endogenous U6:1 or U6:2 promoters, or the *D. melanogaster* U6:3 promoter. Although the BSF U6 promoter isolates showed minimal activity in S2 cell culture, the Cas9 and MAD7 constructs were understood to be functional with the U6:3 promoter in *D. melanogaster*. With both nucleases showing varying levels of activity in cell culture, the developed CRISPR toolkit appears developed enough to be used for BSF transgenesis

Chapter 4: **Establishing an Efficient Platform for Genetic Transduction and Transgenesis in Black Soldier Fly**

4.1 – Introduction

With preliminary characterisation of the CRISPR constructs indicating their functionality in insect cell culture, the study could progress to attempting to integrate the CRISPR components as stable transgenes in BSF. Earlier fly dissection experiments had been performed either off-site or with a limited sample of imported BSF pupa, which were reared to adulthood in a small environmental chamber for the experiments but not maintained further. To be able to perform BSF transgenesis and the later transgenic CRISPR fly crosses, a functional insect colony would need to be designed for the continuous rearing of each development stage of the species. As a novel farmed insect, BSF are produced globally in many contexts and for various different applications. As such, a huge diversity in rearing methods are employed to suit the scale, feedstock requirements or intended function of the desired colony. Whereas scaling-up poses a challenge when rearing in most species, effectively scaling-down in colony size proves to be a difficult problem in BSF, with larval feeding and adult mating being most effective with larger groups and more difficult to achieve with lower numbers[356]. The goals of the current study required the production and simultaneous rearing of up to 12 transgenic lines, with crosses between them producing additional genotypes for maintenance. A simple and standardised rearing method for a small insectary space that could be maintained by a single individual would therefore need to be established.

The success of the current study hinged on the ability to produce transgenic BSF lines that encoded the designed CRISPR constructs. Although growing rapidly, with significant developments being reported elsewhere throughout the short duration of this project, methodologies for effective transgenesis in this species remain relatively unestablished and a nascent area of research[422,440,441]. Although several reports of successful microinjection for CRISPR-Cas9 knockout and germline transformation in BSF have been made in the

recent literature, the methods used to achieve this are not consistent between different cases, and are often described with a lack sufficient detail for wide replication[253,430,435]. The clear commercial promise of edited and transgenic BSF understandably means that much of the driving innovation behind the development of these enabling technologies comes from the private sector and BSF startup companies. As such, several of the few known BSF transgenic lines have been reported with no reference to how they were produced, or even what transgenes they encode.

Although alternative techniques do exist, including electroporation and the more recently ReMOT method, the gold standard for insect transgenesis remains embryonic microinjection[507–509]. Still, being technically challenging, time-intensive and generally low efficacy, microinjection can be difficult to perform reliably, and remains fairly non-standardised across different species[508]. In addition to specialised equipment, successful microinjection requires practice and experience with the methodology, as well as a good understand of the rearing and development of the species in question. Troubleshooting, adaptation and trial-and-error may be required at each stage of the process, which can result in a method optimised for a specific setup with poor generalisability across different systems.

Despite some variation in the specifics, successful embryonic microinjections with the aim of transgenesis have been previously achieved following a similar order of operations in multiple insect species[508]. To take advantage of the syncytium stage of the developing embryo for highest chance of transformation, the injection proper is set to be completed before cellularisation occurs, within a time window empirically determined for each species. This can range from as low as 20 minutes in *D. melanogaster* to several hours in *Anastrepha ludens*[510,511]*.* In most cases, embryos are aged until a later stage prior to cellularisation before injection; this is beneficial for more accurate replication when collecting a pool of eggs from multiple females laying within a defined collection window, and can allow for the development of a thicker chorion that better protects the egg during manipulation[512].

Embryos are often washed with deionised water, with some species requiring surface dechorination to permit easy passage of the injection needle into the syncytium compartment. Dechorination can be achieved by physical or chemical methods; In *Drosophila* and *Ceratitis,* the internal egg yoke can be removed from the tougher outer chorion skin by gentle manipulation with an aligning tool or through the aid of double-sided

tape[513]. Alternatively, embryos can be briefly immersed in a dilute solution of bleach before washing, which reduces the chorion to a thinner, more easily punctured layer[514].

After pretreatments, embryos are arranged in a flat plane with a manipulation tool onto a surface for injection, usually being immobilised with an adhesive or a support on one side. This provides some lateral resistance when pushed up against with the needle, allowing the tip to enter the egg and not push it away. The egg layer can be covered with a thin layer of oil to prevent eggs from drying out during incubation. If possible, aligned eggs can be staged based on morphological cues and the anterior-posterior axis identified[514]. The egg line is then sequentially injected with small volumes of injection mix, with injections localised to the posterior pole where germ cells are expected to form and proliferate following cellularisation. Injected eggs (often termed generation 0, G0) are then incubated in favourable climate conditions until their expected hatch point. The numbers of hatched larva and eggs that ultimately failed to develop are counted to establish the survival rate following injection, and the surviving larva manually transferred to a food source for rearing to the adult stage. When inserting transgenes, monitoring for transient expression of fluorescent markers in injected survivors can be indicative of successful integration within those tissues, and can inform the setup of later G0 crosses[508].

Despite the promising applications for BSF transgenesis, there still exists a significant gap in the literature regarding how this can be best achieved. With the main purpose of this study remaining the development and characterisation of the CRISPR-encoding BSF lines, an additional goal was to develop, refine and describe a reproducible microinjection and rearing method for generating transgenics with high efficiency and generalisability across different injection setups. Details were gathered from several published descriptions in both academic literature and patents to optimise the described method, with several novel innovations being incorporated to simplify and streamline the transgenesis processes.

Additionally, few components crucial for building an effective transgenesis toolkit, such as constitutive promoters for the reliable expression of fluorescent markers, have been tested in BSF. With the production of several different CRISPR-encoding lines, the suitability of the *piggyBac* transposon system for transgenesis in the species could be explored, as well as the compatibility of several candidate constitutive promoter elements for further expansion of the BSF genetic toolkit.

4.2 – Results and Discussion

4.2.1 – Establishing a Method for BSF Maintenance

For an initial BSF stock, flies of the T4 genotype utilised in the earlier DNA and RNA sequencing experiments were imported into the lab. Going forward, the T4 line would be considered as the *wild-type* strain from which all transgenic stocks would be made and compared against. A protocol for insect maintenance, detailed below, was developed according to a general small-scale rearing guideline provided by Beta-Bugs Ltd., with adaptations being made regarding the types of containers used.

All development stages were reared in the same enclosed climate chamber with a set temperature of 28°C (± 0.5°C) and relative humidity (RH%) of 70% (± 5%). Additionally, flies in breeding cages were exposed to a 14:10 hour light : dark cycle, while the egg, larva, pupa and non-breeding fly stages were kept in the dark as much as possible. The initial batch of imported T4 flies were used to initiate a breeding cross in a 30 cm^3 mesh cage. Recently-emerged flies (<24 hours old) were sexed and transferred from their hatching container to the cage, with a target sex ratio of 50:50 for each sex, and no lower than 40:60. The optimal cage density for stimulating breeding and egg production in BSF was understood to be 6500 flies / m^3, translating to ~176 flies for the 30 cm^3 cages employed here[330]. As such, cages were set up with between 60 – 100 of each sex.

BSF do not feed in the adult fly stage, only taking liquids[515]. As such, the only other objects placed in the cage was an attractant pot and an egg collection device, or “eggy”. As BSF are attracted to laying eggs near a potential source of food for hatching larvae, the attractant pot was supplemented with an organic material, usually pupal casings or spent feed, suspended in water[516]. The pot was sealed, with holes in the lid allowing the attractant odour to escape.

The flies were monitored for signs of their characteristic mating behaviours in the cage, after which gravid females would deposit a clutch of eggs in a provided eggy (**Figure 23A**). Female BSF tend to lay egg clutches in small spaces or crevices to protect them from the environment[517]. With an attractant present, the majority of oviposition in the cage was directed to the thin gap between the plates of the eggy, allowing easy egg collection (**Fig. 23B**). The specific device used, produced by Beta-Bugs Ltd, consisted of plastic plates on

a central bolt assembly, with a ~2 mm spacing between the plates created using flat washers (**Fig. 23C**).

Once several females had laid, the device was removed from the cage and disassembled, allowing for the transfer of individual egg clutches with a thin knife point into a small, closed container (**Fig. 23C-D**). Eggs in the container, referred to as an egg pot, were checked for signs of embryo development over the following days with a hand lens, with the development of red eye-spots becoming apparent from around 2.5 - 3 days after egg laying (AEL)[134]. Between 3.5 and 4 days AEL, neonate larva began to hatch in the egg pot, with their motility making the hatching egg pots clear (**Fig. 23E**). Once hatching was evident, the lid of the egg pot was removed and the container placed in a larger pot filled with larval feed. Neonates were left to feed in these “meal 1” pots for around 6 days, or until the feed was processed into a dry soil-like consistency (**Fig. 23F**).

All feed pots were covered with a breathable mesh that allowed gas exchange and maintenance of the proper moisture level. At this stage, larva were easily visible and were deemed large enough to transfer to a larger container of feed, or meal 2 pot (**Figure 24A**). Neonates were not directly added to these containers, as the large amount of supplied feed would quickly foul or dry out. Larger larvae were needed to efficiently turn over and condition the feed placed in this container, requiring the meal 1 pot with a small amount of feed to be employed as a rearing starter. With a small number of neonates, meal 1 was also susceptible to drying out, so the feed was checked regularly during the days following addition of the egg pot, spraying with tap water and stirring if required.

At the point of transfer, meal 1 pots were either overturned or placed directly into feed in the meal 2 pot, depending on the feed consistency in meal 1 (**Fig. 24B**). Meal 2 pots were then left unattended for several days, with the larva being large enough to condition and turn over the feed without the need for additional stirring or watering. Depending on the number of larva present, another dose of feed was added to the container (meal 3), usually the same amount as added as provided in meal 2. The need for additional feed was determined by the consistency of the feed; a fully processed meal 2 appeared dark and dry, with the larva having not reached the full expected size (**Fig. 24C**).

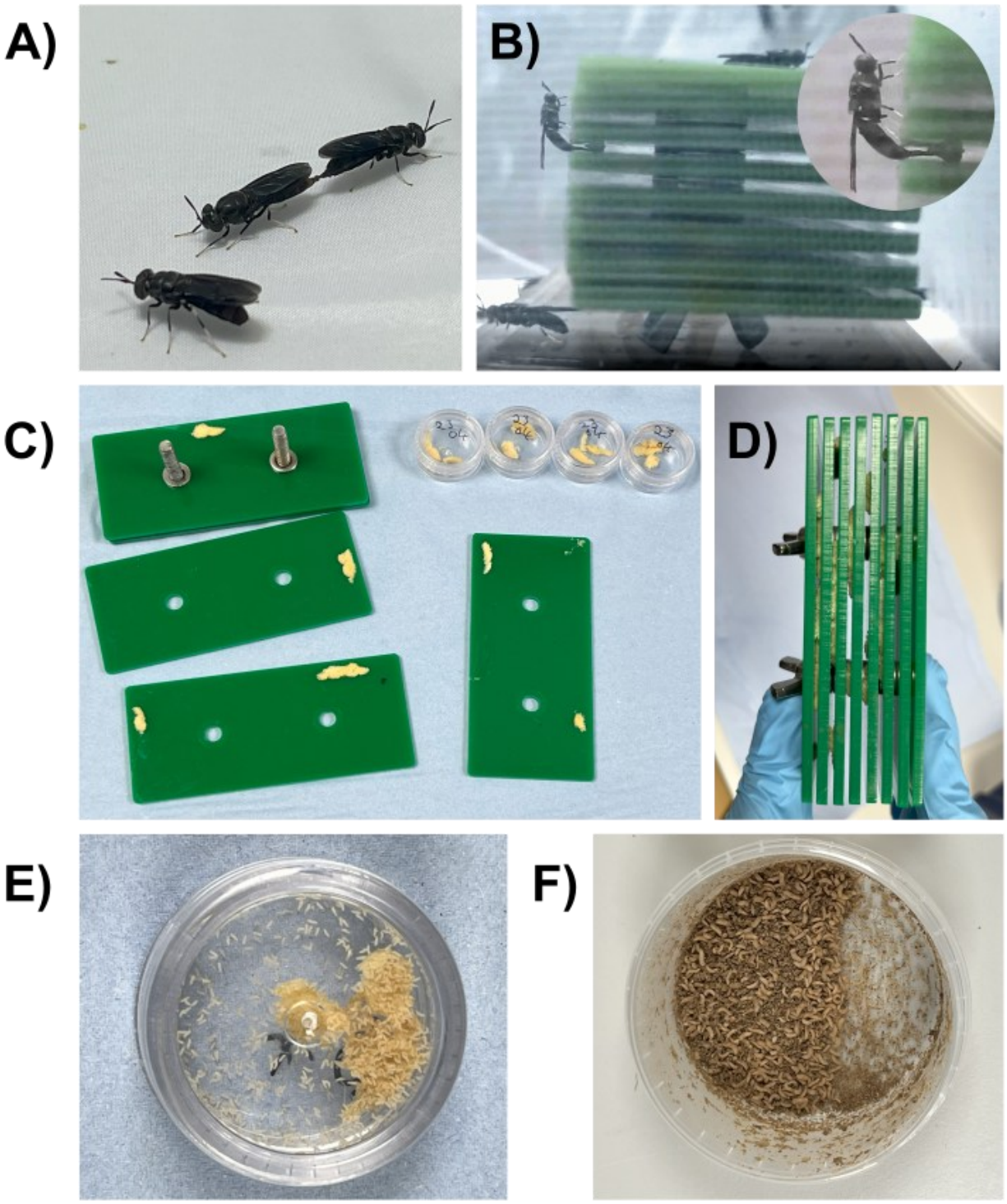


Figure 23: Black soldier fly mating, egg laying, egg collection and neonate rearing: A) Characteristic adult mating behaviour can be observed in cages to confirm cage productivity. **B)** BSF females lay eggs in crevices near a source of odour, exploited in artificial rearing systems through an egg collection device, or “eggy”, above an attractant pot. **C-D)** Egg clutches from individual females are simple to distinguish from each other in the disassembled eggy. **E)** Hatching neonates become evident in clear egg collection pots. **F)** Neonates are reared in small meal 1 containers until large enough to transfer to a new food source.

Once approaching full size, any additional feeding was stopped. Above a certain size, prepupa would cease eating and become highly mobile, attempting to crawl away from the feed and find a dark and dry area to pupate[134]. This behaviour was be exploited to allow fully-grown prepupae to self harvest from the remaining feed. For this, the meal 2 pot was uncovered and placed inside a larger secondary container. Initially, plastic ramps were added to help facilitate self-harvest from the meal 2 pot to the secondary container (**Fig. 24D**). This step was later determined to be unnecessary and removed, with the meal 2 pots appearing shallow enough that prepupa could efficiently climb over the container edge without the use of the ramps. After a sufficient number of larvae had self transferred into the secondary container, or pupal box, the meal 2 pot was removed, with any larva remaining in the feed being transferred manually (**Fig. 24E**). With high numbers of prepupa, paper towels were added to the box to encourage a dry atmosphere for pupation.

Self-harvested prepupa would at first appear highly mobile once in the pupal box. Over the course of several days, mobility would decrease, with a significant darkening of the cuticle from yellow-beige to dark brown indicating a transition into the pupa stage (**Fig. 24F**). Once the population had fully transitioned, the immobile pupa was transferred from the box into separate containers for handling and storage (**Fig. 24G**). The pupal stage lasted between one and two weeks, and could be extended as long as 6 weeks when pupa were stored in a 21°C incubator.

As more pupae were generally produced per egg clutch than was required for initiating a new breeding cage, the population was portioned by weight into individual containers for simpler breeding cage set up (**Fig. 24H**). With the average weight of a single pupa considered to be ~270 mg, 60-65 g of pupae (approximately 220 – 240) were weighed out into tall, clear containers (hatching box) close the expected period of adult emergence[518]. With these, emergent flies were easily identified as they migrated up the walls to the top of the container. Flies were then anesthetised using a CO_2 pad through the lid of the hatching box and removed for screening, sexing and cage set up (**Fig. 24I**). Fly cages were maintained for a maximum of two weeks, or until multiple egg clutches had been collected and significant mortality was seen in the fly population.



Figure 24: Larval, pupal and adult stages in BSF. A-B) Larva in meal 1 pots were ready for transfer into a larger container of feed once the feed appeared dry and mobile larvae were clearly visible without magnification. **C-D)** Larvae were reared in larger containers of feed until feed was exhausted, after which plastic ramps were added to aid movement of the non-feeding larval stage into a dry outer container for pupation. **E-G)** After migration into the pupation container, non-feeding larvae decreased in movement and became darker, with dark brown, immobile insects classified as having entered the pupal stage. **H-I)** Pupa were portioned into clear containers to allow the monitoring of emerging adult flies. Following emergence, adult flies were anaesthetised with CO_2 before sexing, screening and transfer into breeding cages.

Transfer from hatching boxes was only performed quickly after emergence to reduce fly crowding, with separation of flies by sex in hatching boxes not considered essential. In several fly species, rapid isolation and separation of newly-emergent virgin females is considered essential to avoid unintended mating with emerging males, potentially leading to cross contamination. However, in this study, and in contained BSF rearing more generally, breeding behaviour was only found to occur given sufficient lighting and with enough space available for adequate flight[519,520]. Mate formation appears to rely on an aerial courtship in a brightly-lit area, followed by the in-air collision and initial coupling between male and female flies[134,433]. As limited lighting and space are often implicated in a failure to stimulate intentional breeding in the species, hatching boxes were stored in dark conditions, with enough room present in the box for flies to safely eclose and find a perch for wing expansion[521]. With too little space for flight, however, no mating behaviour was observed in hatching boxes and no clutch laying reported.

For the rearing method described, only general timescales are provided. This is because the duration of each stage was found to be highly plastic based on the number of individual larvae reared in the same feed containers. Although the initial input number of neonates for rearing in each meal 1 pot was originally standardized at 400, this was found to be unnecessary for the purposes of generating transgenic lines and general maintenance.

Instead, entire egg clutches were reared in excess in meal 1 pots, with multiple egg pots per meal 1 pot occasionally used. With larva numbers easier to approximate when larger, an estimated ~500 larva from a meal one pot were poured into meal 2 pots, with overdosed meal 2 pots being supplemented with more feed when their feed was exhausted. Unless otherwise mentioned, all the neonates from a clutch were reared together in the later cross experiments, with food added as needed, rather than attempting to reduce the amount being reared prior to feeding.

4.2.2 – Optimising Breeding Cage Lighting for Viable Egg Production

Although the rearing of neonates to the adult stage was achieved with these methods without issue, difficulties were encountered at the fly breeding stage. Although egg clutches were observed being laid in the days following cage initiation, none of the collected clutches were recorded hatching. These clutches were characteristically smaller than expected, and

signs of embryo development were not evident in the days following their oviposition; these eggs were suspected of being unfertilised (**Figure 25**). Although not appearing significantly different from eggs later confirmed to be fertilised immediately upon being laid, clear morphological differences could be seen within the first 2 hours after egg laying (AEL). Unfertilised embryos had a patchy appearance, opposed to uniformity in the dense egg cytoplasm content in fertilised eggs. While clear gaps were seen at both embryo poles in fertile eggs was observed within 4 hours AEL, the patchiness seen in embryo content in unfertilised eggs was exacerbated, and further changes in these eggs at this were not observed. While the appearance of red eye-spots confirmed egg fertility status at around 3 days AEL, unfertilised egg clutches began to develop an orange-brown film on the surface, a significantly darker hue than normally seen in as the clutch matures. While poor egg viability has been mentioned in BSF research, no specific references could be found regarding their alternate morphology to eggs from viable clutches.

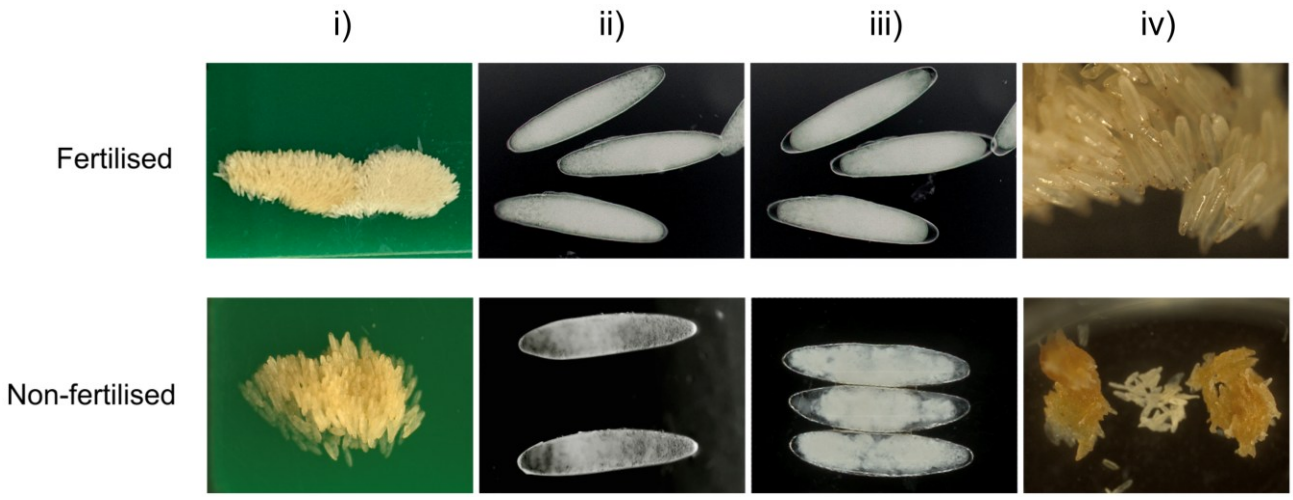


Figure 25: Morphological differences between fertilised and non-fertilised BSF egg clutches. Clear differences in embryo appearance were observed between fertilised and non-fertilised egg collections from as early as 2 hours AEL, both at the clutch level **(i)** and in individual eggs **(ii)**. The difference in egg development became more obvious from 4 hours AEL, where defined void spaces at both egg poles were apparent in fertilised eggs, and a more dispersed and non-uniform pattern of egg yolk density was clear in non-fertilised eggs **(iii)**. At 3 days AEL, a difference in egg clutch colouration was apparent, with non-fertilised egg clutches having an orange-brown hue compared to the yellow-beige found with fertilised eggs **(iv)**.

The inability to produce viable egg clutches from fly crosses posed a challenge to maintaining the colony, requiring additional imports of T4 pupa until the issue could be resolved. It was noted that no breeding activity with caged flies was observed, either through aerial activity or the characteristic linkage of oppositely-oriented male and female on surfaces during mate coupling (**Fig. 23A**). Despite the climate chamber currently being employed for cage maintenance possessing strong LED light strips, with light intensity measured over the 9,000-lux recommended for initiating breeding, consideration was given to the type of wavelengths emitted by this LED setup[518].

While several groups have reported no issues with stimulating breeding under standard LED strip lights, the type of light used is acknowledged to play a role for effective BSF breeding in captivity[519]. Artificial breeding lights that produce wavelengths corresponding to the UVA-spectrum (315-400 nm) as part of their emission profile have been linked to significantly increasing the incidence of mating behaviour and production of viable clutches in BSF breeding setups[522]. A BSF-specific LED breeding lamp with a tailored UVA-emission spectrum was obtained for the established colony.

The light was first fixed above the cage in the environmental chamber, and the effect of the LED lighting measured for the following 3 cage generations for comparison to the prior three failed generations recorded using the standard visible-spectrum LEDs. The addition of the breeding lamp made a clear difference to total number of clutches produced, with the majority of collected clutches being fertilised in the UVA-LED lighting condition (**Figure 26**). Breeding activity was also observed in these cages, further supporting the conclusion that inappropriate lighting led to the production of the unfertilised clutches.

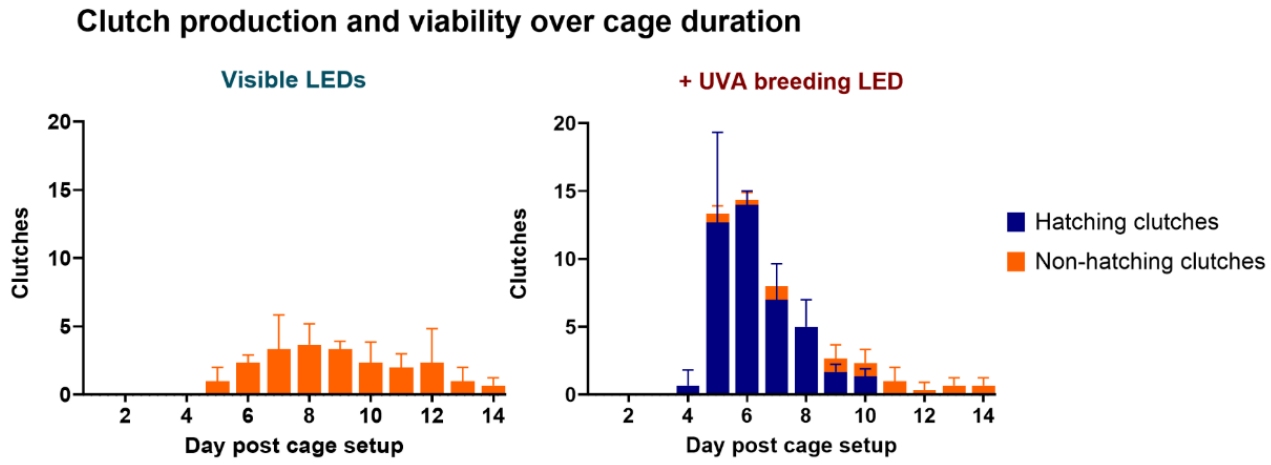


Figure 26: Egg clutch production and viability was greatly enhanced when employing UVA-component LED lighting as opposed to using visible-spectrum LEDs alone. While a small number of egg clutches were laid in cages under visible-spectrum LEDs only, none of these clutches were determined to be fertilised following 4 days of incubation. The addition of a BSF-specific breeding lamp containing UVA-component LEDs significantly increased both the number of clutches laid as well as the proportion of clutches that gave hatching larvae. Both lights were hung 20 cm directly above the cage, and the light intensity adjusted until a lux value of 9,000 was measured at the floor of the cage.

4.2.3 – Expanding the BSF Insectary Rearing Space

With the production and rearing of multiple clutches addressed, it was apparent that no more than a single breeding cage could be reared in the small climate chamber at a time. A design for a small, self-contained space for rearing all BSF life stages was developed with input from Beta Bugs Ltd. The basic structure featured a standing shelf in a hydroponics style tent measuring 1.5 x 1.5 m in diameter (**Figure 27**). This acted as a closed system with the door flap sealed, allowing for maintenance of the various BSF development stages in containers on the shelf.

Several more BSF breeding LED lamps were acquired and arranged on the roof section, with the BSF breeding cages being organised directly below on the top shelf. Lighting duration was controlled using a mechanical timer switch. The shelf directly below was reserved as either a working space or place for additional pupal boxes or breeding cages, with an additional LED lamp housed above and controlled by a switch. The lower two shelves were reserved for egg, larva and pupa rearing, with egg and meal 1 containers preferentially stored on the middle shelf for better shielding from heat sources.

Temperature was provided by an oil heat radiator positioned below the bottom shelf, which was continuously powered and provided most of the heating required during periods when the tent door was sealed. To quickly return the environment to the set temperature when opened, a reactive fan heater was positioned on the floor facing upwards. This heater was powered intermittently by an external temperature controller, which responded to a drop in temperature from the set point of 28°C *via* a temperature probe positioned at cage-height.

A consistent humidity was similarly maintained using an ultrasonic humidifier mounted to the middle shelf. This was fed directly with deionised water from a 20 L stationary external tank, which in turn could be refilled periodically *via* a connected hose from a removable 10 L tank. While the humidifier had a built-in output adjustment, the component was also attached to an external humidity controller, with a connected probe taking humidity measurements at cage height and powering the unit accordingly. A standing fan was mounted directly below the humidifier exhaust to help direct the humidified air towards the breeding cages, where maintenance of the environment was most critical, and generally encourage airflow across the space. An extraction fan connected to a carbon filter was fitted to remove accumulating odours and stale air, which was directed with flexible ducting out of the tent to an exhaust point on the roof of the room. Finally, the main opening was fitted with a magnetic fly screen to reduce fly escapees, and UV-resistant film was applied to the window section on the door flap so the internal space could be observed from outside without additional eye protection.

Despite being a cheap (<£250) and rudimentary setup, the tent environment allowed remarkably flexibility in arrangement of the different BSF containers and cages, and was a significant improvement in terms of space than any reasonably affordable climate chamber. Although climate was controlled simply through cycling the power to the fan heater and humidifier when temperature or humidity dropped, the climate maintained was remarkably stable; humidity fluctuated by less that 5% from the set point, which proved more accurate than a commercial environmental chamber in use in the same room. Temperature fluctuations were more significant, with around a 4°C measured increase during daylight hours than when lights were in operation and adding heat to the space. However, the higher temperature did not seem to affect breeding activity in the cages.

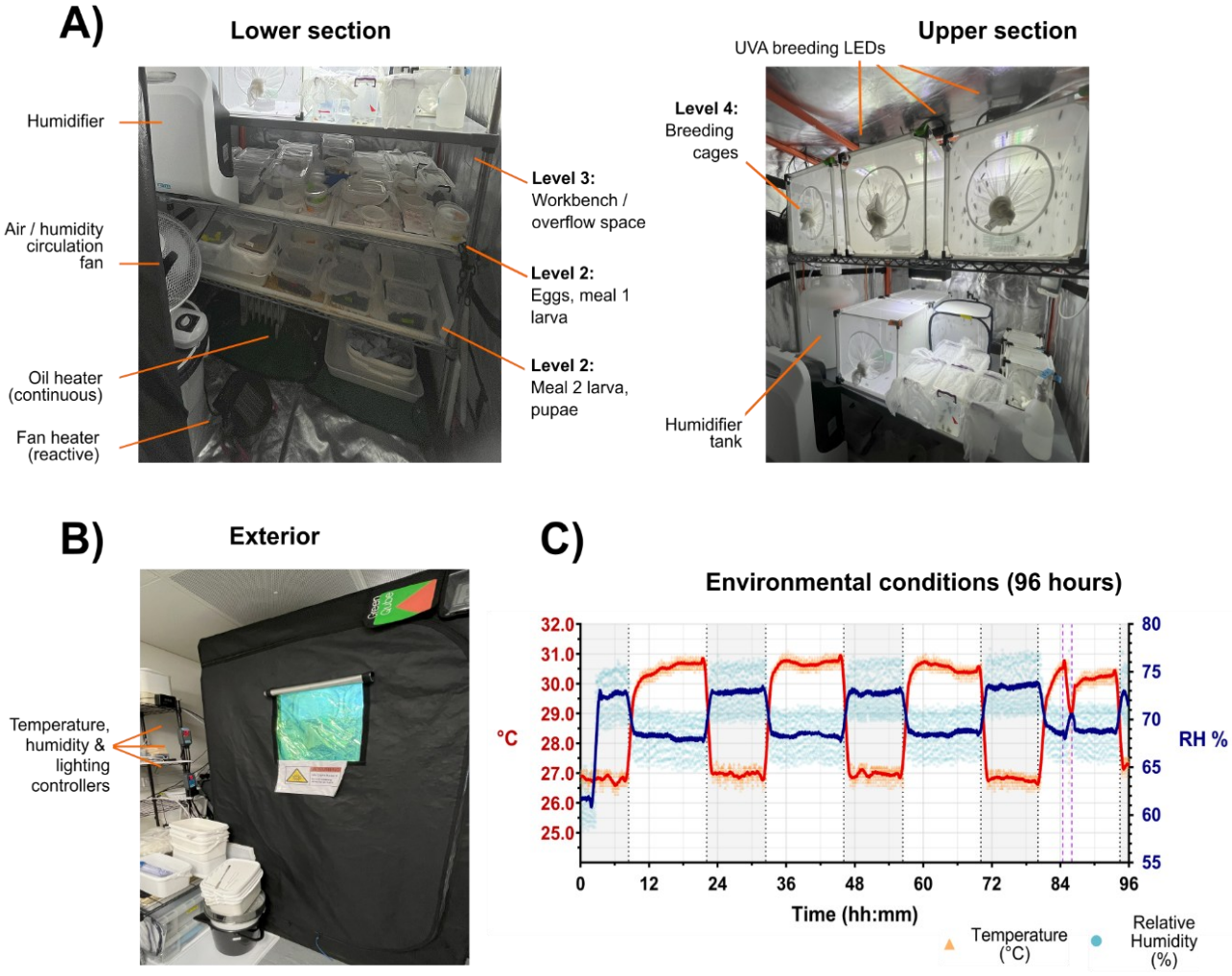


Figure 27: A self-contained BSF rearing space with environmental feedback and controls. A) A 1.5 x 1.5 m sealed hydroponics tent was used to create an environmental chamber to retain temperature, humidity and odour. All BSF developmental stages were kept inside the tent in containers or cages on a metal shelving unit, with the two dark lower shelves housing the egg, larval and pupal stages and the two upper shelves functioning as a space for breeding cages and an occasional workbench. A row of 50W breeding LEDs were fixed to tent poles on the insectary ceiling and the underside of the top shelf, providing illumination set to 9,000 lux for cages placed between 10 – 30 cm below. A 20 L tank served as a deionised water reservoir that, through a float valve, supplied a gravity-fed humidifier affixed to the shelving unit. The humidifier drainage and overflow outlet pipes were directly connected to a drainage pipe outside of the tent to facilitate easier cleaning of the humidifier, as well as to reduce the risk of the tank overfilling and flooding. The main water tank was periodically refilled through a detachable hose connected to a removable 10 L tank placed on the top shelf. A stationary oscillating fan was placed directly below the humidifier vent so that humidified air was blown upwards and across the tent environment, allowing for a consistent humidity and movement of air across the space. To ventilate stagnant air and odours, an in-line extractor fan connected to a carbon filter was suspended from the tent ceiling, with air intake passing through the carbon filter and out of the environment through the tent wall *via* ducting (not visible in image). Heating was achieved through an upwards-facing 50W fan heater placed on the tent floor, as well as a 20W oil-filled radiator placed directly under the shelving unit. The radiator was active continuously for a consistent output of heat, while the fan heater was powered intermittently by a reactive temperature control unit once the temperature dropped below the set point. This allowed for a quick return to the desired temperature after a rapid drop, such as was encountered whenever the tent door flap was opened

for a long period of time. **B)** The contents of the shelving unit could be viewed without requiring eye protection from outside the tent through a UV light-filtering window in the door flap. All cabling for the powered components inside the tent were routed through access ports to an electrical outlet. The oil-filled radiator and extraction fan were powered constantly, while the fan heater and humidifier were powered through connected temperature and humidity controller boxes respectively. The controllers were responsive to environmental probes suspended from the ceiling inside of the tent, and only powered the connected heater and humidifier components when the set internal temperature or humidity values dropped below the set point. The breeding lights were connected to a 24-hour segment timer socket switch, which allowed the lights to be powered only during set hours according to the desired lighting cycle. **C)** Despite the rudimentary environmental feedback control system, the interior of the tent maintained a consistent temperature and humidity without any additional inputs over periods of up to 96 hours, after which the humidifier tank would need to be refilled. At constant environmental control settings of 28.5°C and 70% RH when measured by probes at the tent ceiling, the temperature (red line) fluctuated by around 4°C and the average RH (blue line) by around 5% throughout the measured period. The largest changes in both values occurred in line with changes in the lighting period, with the dramatic temperature increase notable during the light photoperiod (area in white) likely relating to the significant heat output of the breeding lights when powered. Temperature was seen to drop in the dark photoperiod (area in grey) when the lights were not powered. Conversely, the rise in RH level during this photoperiod likely resulted from the lower temperature level. Fluctuations within the same photoperiod were generally much lower than between photoperiods, with temperature changing by around 0.5°C and RH by 3%. Only a modest drop in temperature was encountered when the tent door was left open for a two-hour period during colony maintenance activities (area between the purple dotted lines), with both temperature and RH quickly returning to their original values within an hour of sealing the tent.

4.2.4 – Developing a Preliminary Protocol for BSF Embryo Microinjection

With the improved insectary environment, numerous clutches and cages of the BSF T4 strain could be reared in tandem to produce a reliable output of eggs. Using these, the first rounds of embryonic microinjection were attempted. The initial protocol developed shared the same general steps as embryo microinjection for germline transformation protocols commonly used in *Drosophila,* mosquito, and assorted Lepidoptera and Tephritid species[513,523–525]. While these steps vary in methodology, materials used and specific timings, the general order of processes remains the same; Egg collection, washing and dechorination, egg alignment, microinjection proper, post-injection incubation and transfer of hatched larva to a feed medium.

The initial protocol tested in this study was based off of the established method for injection in *Ceratitis capitata* routinely used in the insectary, which was viewed as a general structure to build from and adapt as needed[526]. In brief, BSF embryos were collected within 40 minutes of being laid, and arranged in a line horizontally using a steel needle onto a line of heptane

glue on an injection slide. Embryos were then covered with a thin layer of Halocarbon oil 700, and injected with a small volume of sterile, filtered molecular-grade water directly into the posterior pole with a quartz needle pulled according to the dimensions used for *Ceratitis* microinjection. Following injection, embryos were incubated on the slide in the insectary climate condition and checked for development and emergence 3 days later.

After several injection rounds with this protocol, injection survival was found to be 0% (n = 312 embryos across 12 egg clutches / injection slides). Untreated and non-injected controls, comprising of eggs either transferred from a clutch to a separate egg pot (n = 25 eggs per clutch) or aligned but not injected (n = 15 per clutch) respectively, conversely showed high survival rates of 88-100% and 80-100% respectively, after removing slides populated with eggs from clutches later deemed to be unfertilised. 3 days post-injection, the injected G0 eggs showed minimal development and characteristic signs of injection-related damage, including bursting or significant leakage at the injection site.

In structurally intact G0 eggs, large transparent voids or cavities were present within the translucent egg yoke material, indicative of injecting excessive volumes of liquid. When using lower injection pressures to inject less volume, this phenotype occurred less frequently, and survival was increased to around 5% (table 3). Despite this, hatching embryos appeared lethargic and showed no signs of feeding when transferred to a source of feed, and so were not able to be retained.

Injections were then performed without the halocarbon 700 oil layer, as the current published BSF microinjection protocols are roughly split on the use of an oil cover or injecting embryos while exposed to the air[253,422,435,440]. Despite no obvious bursting or over-filling, survival following injection without the oil layer remained around 5% (n = 170 eggs over 8 clutches / injection slides), with hatched neonates not surviving long enough to be transferred into feed. Due to the unsuitability of this protocol in the current form, each composite stage was examined for optimisation separately, in an attempt to first increase survival following injection before attempting transformation.

4.2.5 – Establishing an Embryo Collection Window for Microinjection

With a reliable production of fertile embryos established, investigations were made regarding the ideal timing of egg collection and alignment for injection. On the day of injection, the eggy from a productive cage was replaced with an empty one. The cage was then checked every 30 minutes for signs of questing or actively laying females. Females were allowed to finish laying for another 30 minutes before collecting and exchanging a fresh eggy, with the collected embryos being 30-60 minutes old. Large, non-overlapping clutches were harvested as usual using a knife into an egg pot, which was closed tightly to seal in moisture and protect the eggs from exposure.

Freshly laid eggs (under 30 minutes old) had a characteristic pale off-white colour, which transitioned to increasingly darker shades of yellow over the course of egg development. Embryo age seemed to correspond with fragility in the egg chorion when eggs were manipulated with a fine needle or brush, with young and pale embryos much more likely to be damaged or burst than embryos aged 1-3 hours, which had a yellow and more turbid appearance. This is consistent with egg melanisation and tanning processes in other insects, a process that protects the egg from desiccation and damage from the environment[527]. While successful transgenesis is generally encouraged by injecting eggs at the earliest stage possible, it remained to be established if the handling and manipulation of recently laid BSF eggs prior to injection imposed any fitness cost, and if shorter periods between egg laying and alignment correlating with a lower hatch rate.

Concerns were also given to the impact of environmental stress on young embryo viability. Arranging a neat line of embryos and proceeding with injection would require a period of exposure to the comparatively lower measured temperature (19-24°C) and relative humidity (40-55%) of room conditions than that of the closed insectary environment. Preliminary experiments indicated that large egg rafts could tolerate being open to room conditions for up to an hour without impacting hatch rate (data not shown). However, eggs were likely to experience a different rate of temperature sensitivity and desiccation when laid out individually as apposed to being in a large mass, where eggs in centre of the clutch may be protected and experience more moderate conditions. The link between embryo age and the degree of chorion hardening would also likely impact their ability to tolerate cold and desiccation for extended periods.

Eggs from clutches aged 30 to 180 minutes old were aligned in 30-minute intervals. Eggies were closely monitored for laying events and the time recorded so that the completion of laying could be identified to within 10 minutes at the time of device collection. For the 30-minute condition, clutches were immediately collected into an egg pot before aligning 40 eggs onto a glass slide with a metal needle. For timepoints over 30 minutes, eggies were removed from the cage but the embryos left undisturbed to age in the insectary environment for the required time before transfer to egg pots and alignment. The time taken to complete the alignment, during which individual eggs were separated from the egg mass and exposed to the external environment, was designated as the exposure time. For each embryo age timepoint, slides were additionally subjected to a 20-, 40- or 60-minute exposure time, before being placed into a covered petri-dish in the insectary environment.

Younger eggs were initially more difficult to align without bursting or bending them with the needle, whereas the older eggs could be handled much more reliably. After sufficient practice, however, 30-minute old eggs could be reliably aligned without discounting any due to obvious mechanical damage. Survival following alignment was high in all cases (85.0%-100.0%). Logistical regression analysis found a non-significant relationship between embryo age at alignment and survival ($B1=0.01$, $p = 0.0655$). In conclusion, embryos of all ages proved similarly resilient to the processes of alignment and exposure, meaning the correct timing of injection could therefore be determined according to other factors, such as apparent development stage.

4.2.6 – Early Embryo Morphology and Developmental Staging for Microinjection

The embryos should preferentially be injected before cellularisation takes place and as close to the posterior pole as possible. The morphology of freshly laid BSF embryos was examined at several stages in developmental progression *via* a recorded timelapse. Under magnification, the anterior and posterior poles of the egg became apparent from as early as 15 minutes AEL (**Figure 28A**). The anterior side of the egg appeared slightly thinner, with a steeper taper from the midsection to the pole than at the posterior end. The posterior pole of the embryo yolk showed a steady retraction from the chorion surface from as late as 15 minutes to 3 hours AEL, leaving a clear void between the dense cytoplasmic compartment and the vitelline membrane on the internal side of the chorion[528]. This gap was indicative

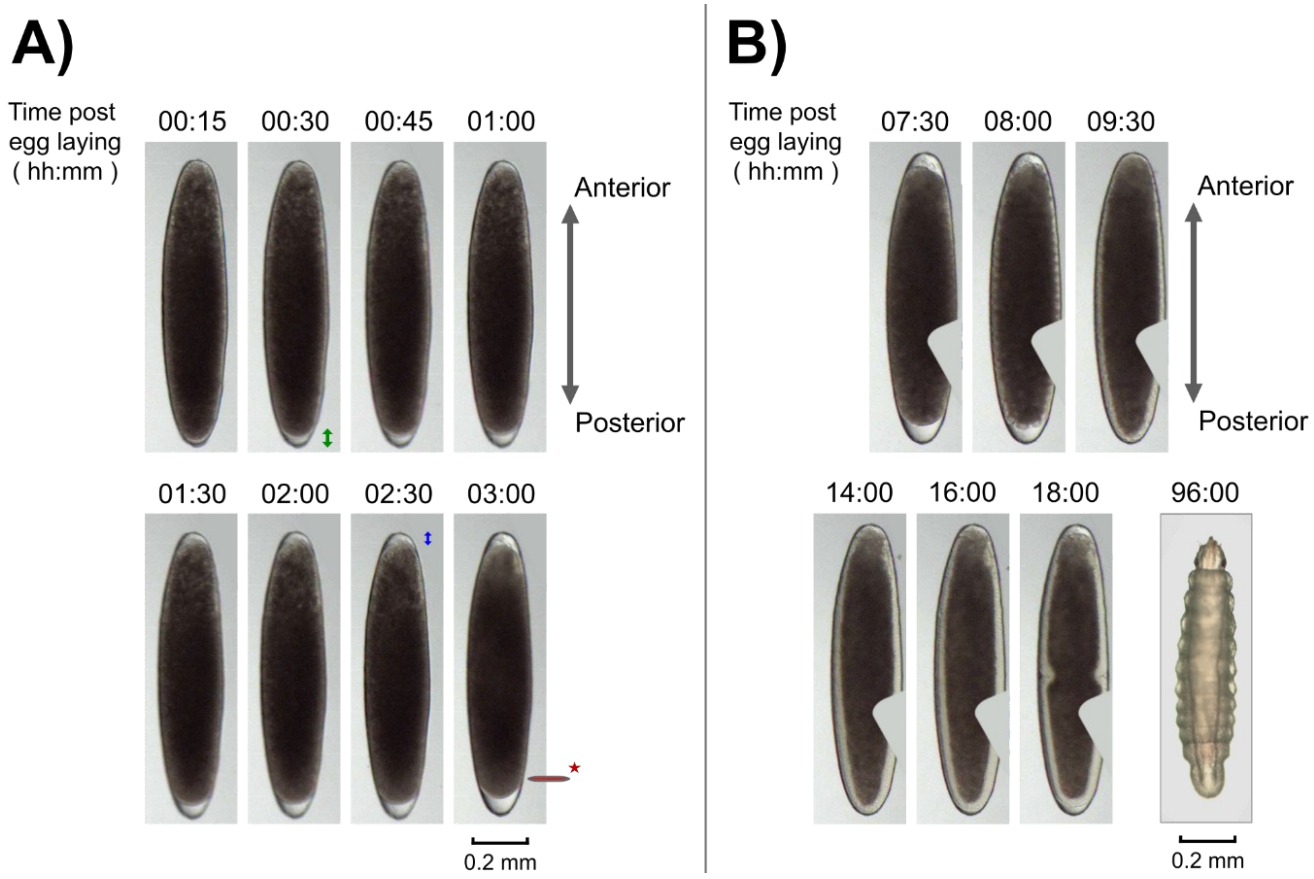


Figure 28: Early morphological signatures of embryo development in BSF: A) Dorsal view of a single developing embryo recovered and imaged at 26°C from 15 minutes after being laid (AEL). Timelapse images were taken at 5-minute intervals from 15 minutes to 3 hours AEL. The anterior-posterior body axis is shown by the black arrow, with the ideal site and direction of embryo microinjection indicated by the red star indicating a needle point in the 3-hour AEL image. The green and blue arrows indicate the relative distances covered by the dense syncytium layer during the "pole retraction" phase at the posterior and anterior poles respectively, with the steady growth of the void spaces between the egg yolk and chorion being a characteristic indicator of both embryo polarity and developmental progression. Polarity can also be identified by the small differences in width across the egg, with the posterior half appearing slightly thicker than the anterior. Additionally, the yolk at the posterior pole appears darker and denser than yolk found at the anterior, and has a more defined and less "feathered" edge. **B)** After 3 hours AEL, the imaging interval was increased to 30 minutes until 18 hours AEL. While little morphological change was visible in the embryo after the end of pole retraction at 3 hours AEL, signs of cellularisation and segmentation at the edges of the yolk started to become apparent from 7:30 hours AEL. This growing layer of cells, indicative of a transition from syncytial blastoderm to cellular blastoderm, expanded to fill the void space at both poles, and became progressively more defined until around 16 hours AEL. An invagination at the embryo midsection began to develop from 18 hours AEL, suggesting entry into the gastrulation stage and the formation of defined germ layers. Microinjections for genetic modification should ideally be performed in the syncytial blastoderm stage well before the start of cellularisation, with the morphological evidence presented here indicating a reasonable injection window as between 15 minutes up to some time before 7:30 hours AEL at 26°C, with additional research necessary to narrow down an optimal range further. For reference, a hatched neonate larva imaged at a similar magnification in a separate experiment is depicted next to the developing embryo at 18 hours AEL. To obscure a software error message that appeared on all images taken in the 7:30 – 18:00 hours AEL window, images in B) were edited with a grey blur on the lower right side of the egg, but have otherwise been unaltered.

that the egg was viable and in active development, and useful for distinguishing fertilised from unfertilised eggs. From around 2:30 hours AEL, the yolk at the anterior pole also showed signs of retraction, although moved a lesser distance than seen in the posterior pole.

No obvious changes occurred in the embryo morphology between 3:00 and 7:30 hours AEL, after which a rippling activity was noted around the edges of the yolk (**Figure 28B**). This likely signals the start of cellularisation, with a single layer of cellular blastoderm surrounding the yolk becoming apparent between 8:00 and 9:30 AEL[426]. With this, the retracted poles were seen rapidly extending back towards the chorion, with the outer cellularised layer becoming more defined. The timelapse was stopped around 18 hours AEL, with an invagination beginning to become apparent at the egg midline and indicating the first signs of gastrulation[529].

It should be considered that for the duration of imaging, embryos were incubated in a room at a slightly lower temperature than used in standard rearing (26°C), which could have resulted in a slower development than usual. This time course of images suggested that the BSF embryo remained as a syncytial blastoderm for potentially up to 7:30 hours AEL, after which signs of cellularisation were clear. While injections as late as 7 hours AEL were not attempted, this data at least indicated that cellularisation within a realistic injection window of 1-3 hours was unlikely, with the timing of injection likely being much less stringent than more-rapidly cellularising embryos such as *D. melanogaster*[426]*.* This also demonstrated that the posterior pole could be easily identified based on embryo morphology and the relative speed at which both poles retract from the vitelline membrane.

4.2.7 – Optimising Embryo Alignment and Microinjection Site

With an injection window now appearing more defined, attention was turned to embryo pretreatment and alignment for injection. While washing and bleaching of embryos for dechorination was considered, the chorion was significantly easy to pierce with an injection needle without these steps. Embryos appeared significantly sensitive to moisture, with any exposure to water or dilute bleach solution completely ablating development in all attempted cases. Manual dechorination also proved impossible using mechanical agitation with a needle tip and double-sided tape, resulting in burst eggs, and so the dechorination step was omitted from the protocol.

As the microinjection needle would push up against the eggs during injection, they would have to be immobilised after alignment. Various methods for immobilisation were employed inspired by other insect microinjection protocols, including brushing the eggs onto a strip of double-sided tape, a line of heptane-based adhesive, lining them up against a cellulose membrane, and placing them directly on a glass coverslip. While no method was detrimental to survival, it became apparent that the egg would sufficiently adhere directly onto a glass coverslip due to its moist outer coating, presumably involved in binding the egg clutch together. To stop the egg from being pushed when touched with the needle, two glass coverslips were stuck together with an adhesive, with a small overlap at the edge. Eggs could be immobilised by their coating on the bottom coverslip, and pushed up against the side of the top coverslip as a support against any lateral movement (**Figure 29**).

At this point, the specific site of injection was re-examined. All previous injections had been carried out on eggs aligned horizontally, with injection directly in the pole. This involved bypassing the large void between the vitelline membrane and to reach the retracted egg yolk with the needle tip, which would result in backflow of the yolk into the space when the needle was pulled out. Although this method works sufficiently for injection in *C. capitata,* for example, the significantly low injection survival experienced with this injection method suggested that injection may be more effective from a different angle.

By injecting from a lateral direction to a vertically aligned embryo, the chorion could be bypassed and needle inserted into the egg yolk without traversing and significant void space between the two. The posterior yolk could still by accessed from the side, with the needle needing less travel inside of the egg to complete the injection. When attempting new injections with this method, survival was seen to increase dramatically from around 5% to 20%, with the only difference between injection sets being the embryo orientation and direction of injection (**Table 3**). At this point, signs of transient GFP expression were seen in surviving larva injected with *Vasa*-MAD7 and the *piggyBac* transposase helper plasmid I-hyPBase. This suggested that integration with the produced constructs in injected BSF was possible.

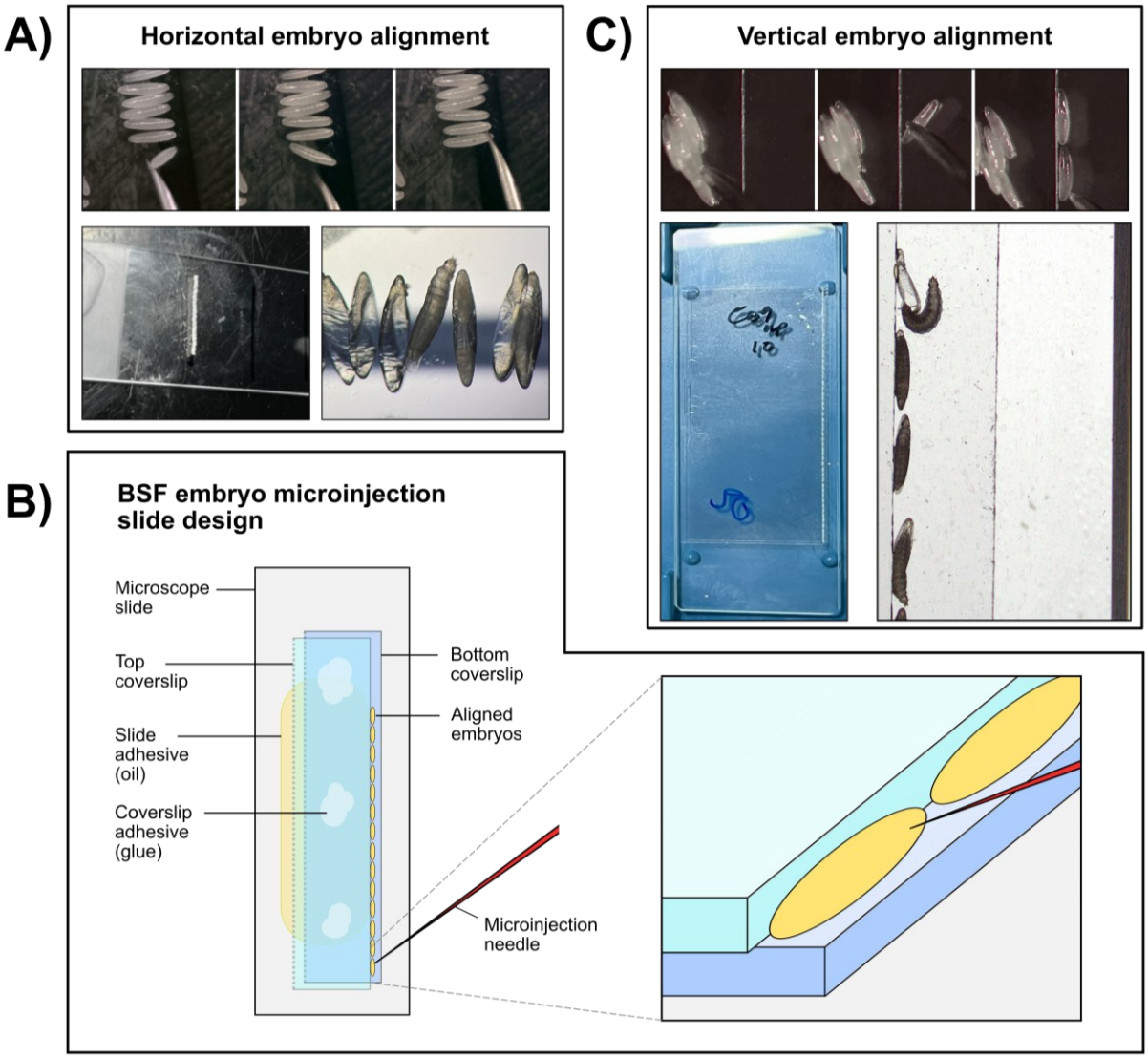


Figure 29: Embryo alignment methods used for microinjection. A) In the initially employed microinjection protocol, embryos were aligned horizontally onto a line of heptane-based glue, using a marked line of ink as a guide. As only one egg pole would be facing towards the injection needle, the posterior pole had to be identified pre-alignment. Survival of aligned, non-injected eggs was high (>95%), with embryo development and neonate hatching identified under magnification. **B)** The final slide design used for microinjection was composed of two glass coverslip sections glued together, which were then adhered to a microscope slide with oil to allow easy removal following injection. The coverslips were glued at a staggered angle, leaving a small (1-2 mm) overlap for vertical embryo alignment. Embryos were brushed until they lay flat on the bottom coverslip, with one edge pushed up against the top coverslip to keep the embryo in place during injection. Both poles of the embryo would be exposed on one slide for microinjection, with the needle arranged at a 20°-45° injection angle. No additional treatment to immobilise the embryos was required. Diagram elements are not to relative scale, with embryo size enlarged for diagram clarity. **C)** Vertical alignment of embryos using the final slide design could be achieved rapidly, with the coverslip edge being useful for creating a straight line and pole identification not being necessary at this stage. Between 42-50 embryos could be aligned per slide when prepared using two halves of a 50 x 22 mm coverslip, with space available for labelling. Vertically aligned non-injected embryos had a survival rate comparable to horizontally aligned embryos. The glued coverslip assembly could be removed from the microscope slide following injection and placed in agar feed vials, allowing suspended neonates immediate access to feed upon hatching.

4.2.8 – Optimising Post-Injection Neonate Survival and Rearing

Despite appearing more mobile than the previous injection survivors, the produced G0 larva were lost when they were transferred into feed. It was apparent that the amount of food provided for the few injection survivors was too much, and would harden or foul before the larva grew large enough to efficiently regulate the moisture content of the feed. However, smaller amounts of feed would quickly dry out, trapping the larva. It was clear that an alternate feed would be required for the safe rearing of small numbers of G0 neonates.

For this, a sucrose, yeast and agar-based diet commonly used for *D. melanogaster* rearing was adopted (**Figure 30A**[530]. At a 1.5% agar content, the feed provided a solid matrix to support the feeding larva, while still containing a high moisture content. In addition, the presence of propanoic acid and p-Hydroxy-benzoic acid methyl ester as preservatives negated the issue of fouling and the growth of mould observed in the standard chicken feed-based diet. BSF neonates tolerated the agar diet well, and could be observed growing successfully even when added in small numbers (**Fig. 30B**). As the food did not need to be conditioned by larval processing, low mortality rates were seen when G0s were transferred into this diet, with additional agar being added once the feed was exhausted (**Fig. 30C-D**). Larva were successfully incubated in agar-containing vials until the point of prepupa, when they were removed to a dry container for pupal transition (**Fig. 30E**).

4.2.9 – Refining Microinjection Needle Parameters

The most significant optimisation made to the microinjection method was the alterations made to the injection needle. Previously, an injection needle pulled with a thick taper was used for microinjection, with the same needle pulling parameters as used successfully in *D. melanogaster, A. gambiae, C. capitata* and *A. ludens*[531]. However, in an effort to increase G0 survival further, alternate needle designs were produced for use in BSF injection (**Figure 31**).

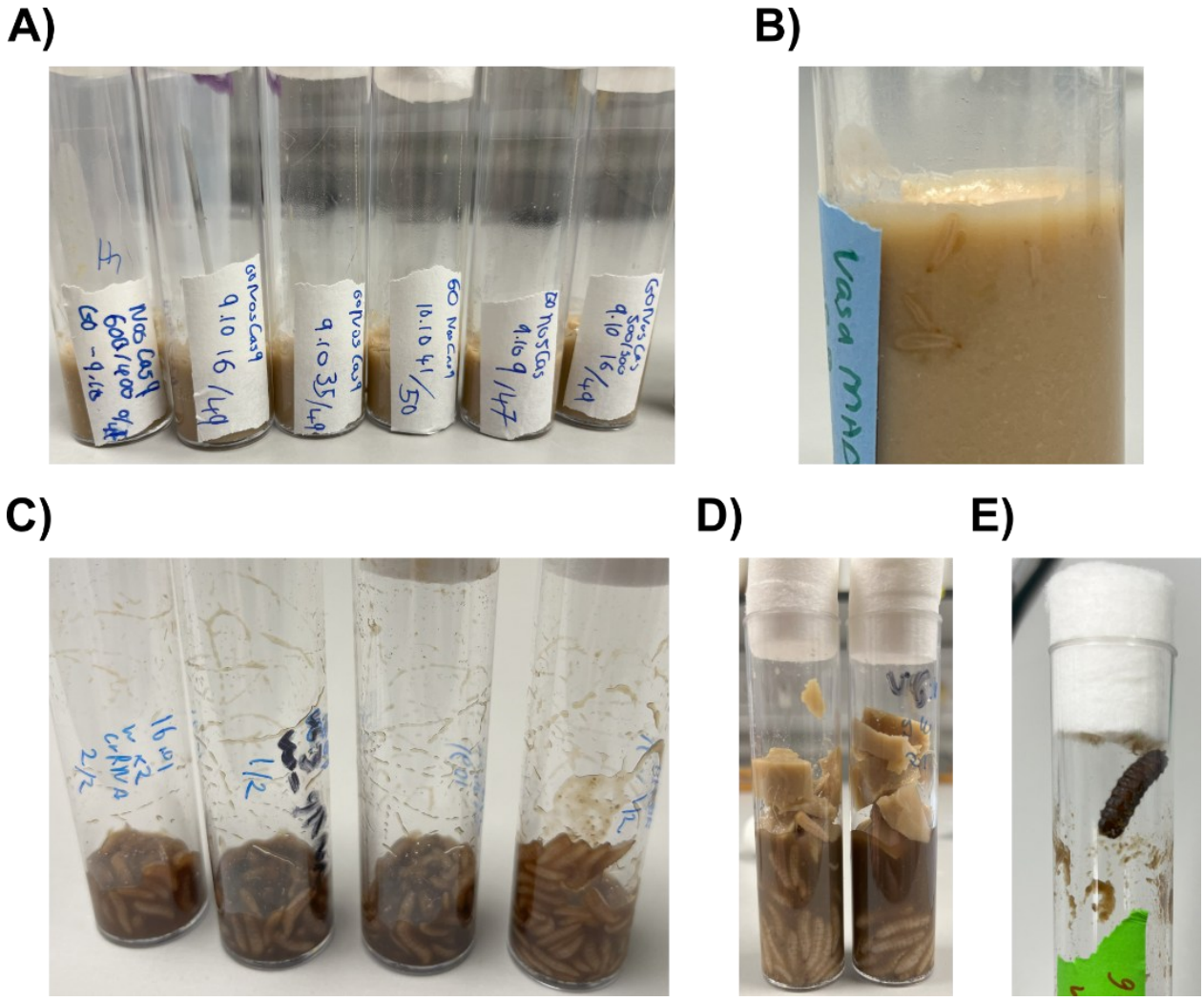


Figure 30: Agar-based feed vials allowed for the efficient self-transfer of hatching neonates and safer rearing of larvae at low densities. A) When aligned using the vertical slide design described previously, injected embryos could be incubated in a feed source immediately following injection. The glass coverslip assembly was removed and placed vertically in an agar-feed vial, with the embryo line suspended above the feed source. Up to three coverslip assemblies, containing up to 150 eggs, were arranged in one feed vial, and removed in the days following predicted egg hatching to establish the injection survival rate. **B)** Early larvae could be seen with the naked eye around the walls of the vial from around 5 days post hatching. Regardless of number, feeding larvae were able to regulate the consistency of the feed without the need for stirring or disrupting the agar layer. The feed was also not seen to mould, dry out or create visible pools of moisture that may be harmful to early neonates. As few as a single neonate can be safely reared in a single vial. **C)** Agar feed becomes exhausted as larvae grow, apparent as a dark brown and liquid consistency. However, larvae were still viable in this condition and show no signs of drowning. For high larval densities, such as pictured here, larvae could safely be transferred at this stage to a small amount of standard diet in a meal 1 container, providing adequate space for further growth. **D)** Fresh agar feed could be added directly to the vials for continued rearing. Larvae could sufficiently incorporate the incoming feed, meaning that feed mixing was not required. **E)** Larvae could be sufficiently reared until the prepupal stage in vials, with migratory prepupa becoming apparent by their colour and movement out of the feed to the upper vial walls. Prepupa were removed manually to a dry container for pupation, or the vial opened and placed in a way that allowed easy migration out of the vial.

Needle settings were adjusted on the needle puller to produce a capillary with a long, fine point with a shoulder far from the tip. This made for a characteristically thin and delicate needle, as opposed to the bee stinger type stout needles used for *D. melanogaster* injection. Whereas needles are often pulled to have a thicker taper and shoulder to provide enough force to puncture through the chorion of other insect species, this type resulted in clearly traumatic injections and burst eggs in BSF (**Fig. 31A i**). Success was found in a series of needles with progressively longer, sharper points (**Fig. 31A iv-v**). Producing needles of this type allowed for much cleaner incision into the BSF embryo requiring minimal pressure, indicated by the lack of resistance or surface deformation at the injection site. The thin needle also leaves minimal damage to the embryo's integrity; leakage of syncytial fluid through the puncture site was not encountered after removing the needle despite the high internal pressure, indicating formation of a seal. While survival was sufficient following injection with standard needles with a thicker taper, the thin needle provided the highest survival, keeping it's point for longer without clogging and allowing easier control over injection volume. Due to the elongated tip and small opening, injection compensation pressure had to be kept significantly higher than with needles with thicker openings to prevent clogging in the small capillary tip. By not injecting under a layer of oil, compensation needed to be even higher to resist the high internal pressure of the embryo causing backflow of syncytial fluid into the needle tip and clogging it.

The inclusion of phenol red dye in the injection mix, while not essential, was also found to be a useful aid for visualising the amount of mix entering the egg, allowing for adjustment of injection pressure to increase or decrease flow. Embryos injected with too much fluid would not usually burst, but would arrest and show a characteristic bubble or empty space where there would usually be translucent egg fluid. A lack of flow seen inside the egg usually indicated a clogged needle. This could be resolved by depressing the clean button, or touching the needle to the edge of the coverslip. After several injections, the needle would clog more frequently and would require breaking back by more vigorous scraping on the coverslip, producing a needle with a slightly larger tip and faster flow as a result. Compensation pressure would be decreased to recreate the desired flow rate before injecting further, with multiple rounds of breaking back and pressure reduction possible before the needle was too clogged or broken back to inject smoothly or without excessive pressure, at which point the needle was exchanged.

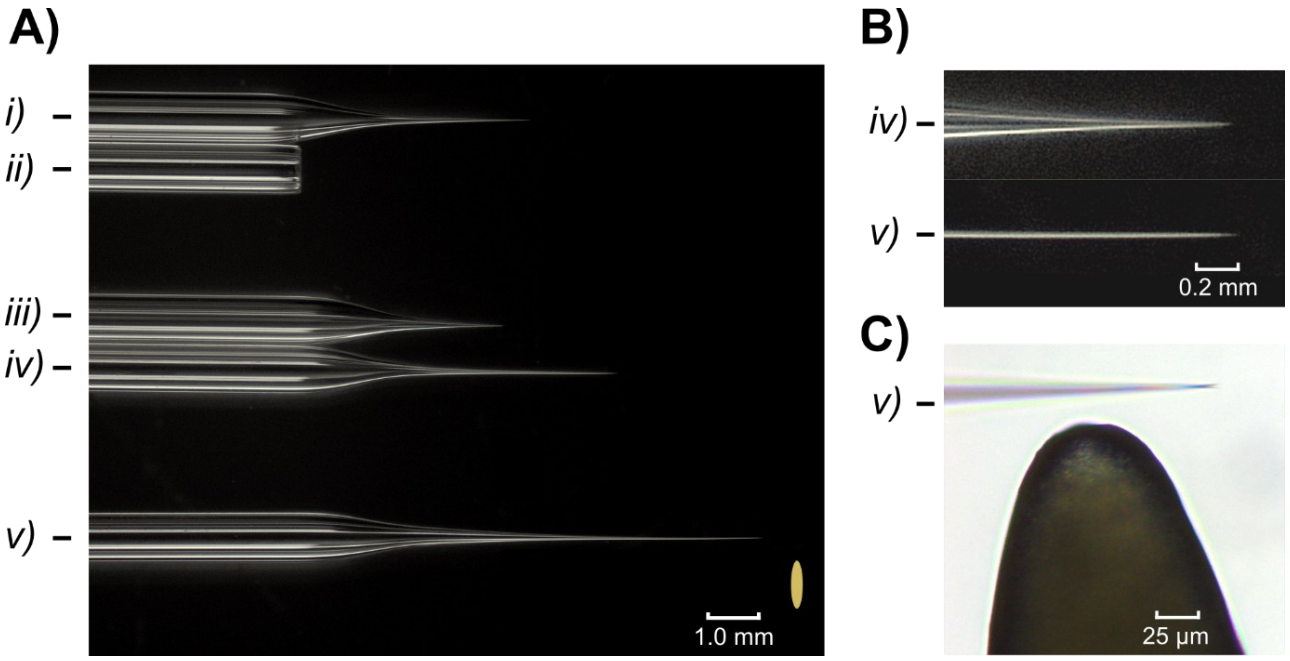


Figure 31: Needle parameters for embryo microinjection. A) Several types of needle for microinjection were pulled and tested for suitability in this study. The initial type used **(*i*)**, pulled to a specification successfully used within the lab for *A. gambia, C. capitata* and *D. melanogaster* embryos, featured a short taper and and fine point. All needle tips were pulled from the capillary shown in **(*ii*)**. With needle **(*i*)** showing significant deformation to the embryo cuticle when pushed against it, and causing internal leakage when removed from the yolk, other needle designs were attempted. Needle **(*iii*)** was a classic "beestinger" design employed in the injection of zebrafish embryos, with a wide shoulder and thick taper. The wider tip was found to cause excessive cuticle damage and leaking when inserted into the egg. Needle **(*iv*)** was pulled to the same specification as needle **(*iii*)**, but with a longer tip to allow more tapering and a thinner point being inserted. This was found to show lower embryo surface deformation and easier insertion into the embryo. This was expanded on in needle **(*v*)** with a comparatively much longer taper allowing for a very fine needle thickness at its point. Minimal deformation was observed at the chorion when inserting this needle, and the small injection scar showed no obvious leakage of embryo contents when the needle was removed. An embryo graphic is included to illustrate needle tip thickness relative to embryo width. **B)** At high magnification, needle **(*iv*)** had a gradual tapering and higher thickness over the distance of needle tip inserted into the embryo (~0.2 mm) when compared with needle **(v)**, which was thinner and more uniform at the tip. **C)** The tip of needle **(*v*)** relative to the width of the embryo pole when mounted to the injection apparatus. By injecting through the side of the embryo chorion into the yolk, the vittelline gap was bypassed and a limited needle travel (~25 µm) was required for visible injection of material into the yolk.

With the improved needle facilitating rapid, less damaging injections, the pace of the injection process saw a significant increase. Some needles were able to keep their sharpness for over 5 slides worth of injections, or over 250 eggs, at which point they would be exchanged for a fresh one. The timing of injection was around 15 minutes per slide of 50

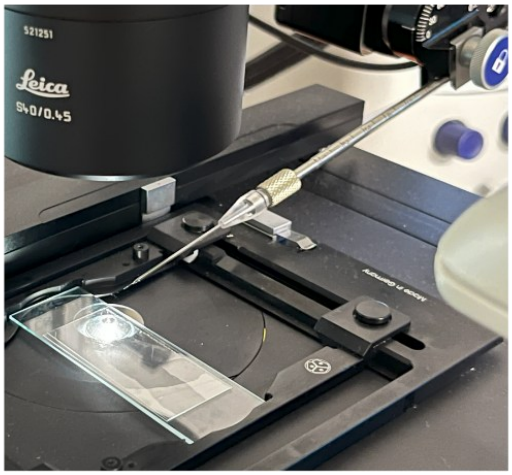
eggs, or longer if multiple needles required opening and exchanging. Alignment of 50 embryos on a preprepared slide took around 20 minutes or less, meaning the minimum time for completing injection was around 40 minutes from the point of egg collection, to aligning, preparing the needle for injection and injection proper.

While a routine of alternating alignment and injection per slide was used to build the protocol, with the long injection window and tolerance of remaining exposed to room conditions, with time and practice it was more convenient to align 4-6 slides back-to-back and inject the entire batch, reducing the time used for swapping between machines and allowing a needle to be used for multiple slides without sitting long enough to clog. Injecting in batches like this could mean 200-300 eggs could be aligned and injected within a 2-hour period.

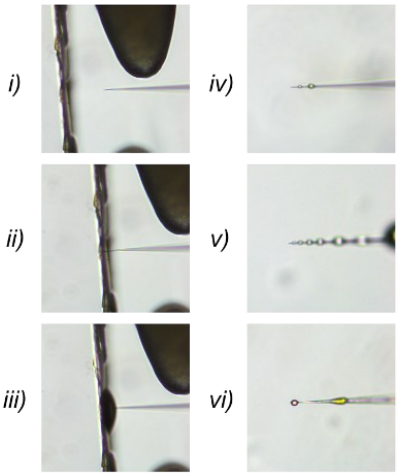
4.2.10 – An Improved Embryonic Microinjection and Transgenesis Protocol in Black Soldier Fly

Following the steps of optimisation described, a highly efficient method of BSF microinjection and transgenesis was established. The optimised method and protocol steps described in sufficient detail for repeatability in the methods section (**Chapter 2.2**). The major contributions to high injection survival and transgenesis are summarized in **Figure 32**. Following the developed protocol, high rates of transient expression was noted in G0 injected larva, which can be used as an indicator of successful establishment of a transgenic line following crossing (**Figure 33**). Finishing this section, the injection data relating to the optimisation process is summarised in **Table 3**, while the data associated with the transgenic lines later described in this study is summarised in **Table 4**.

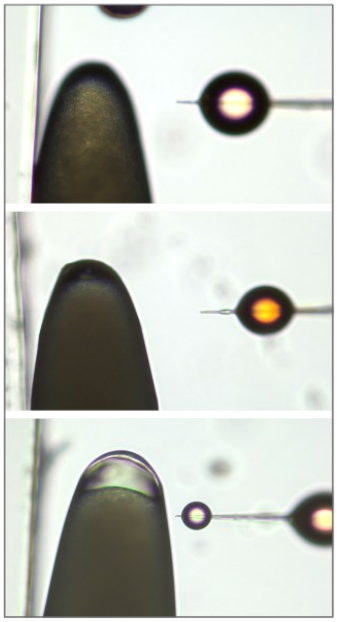
A) Injection plane



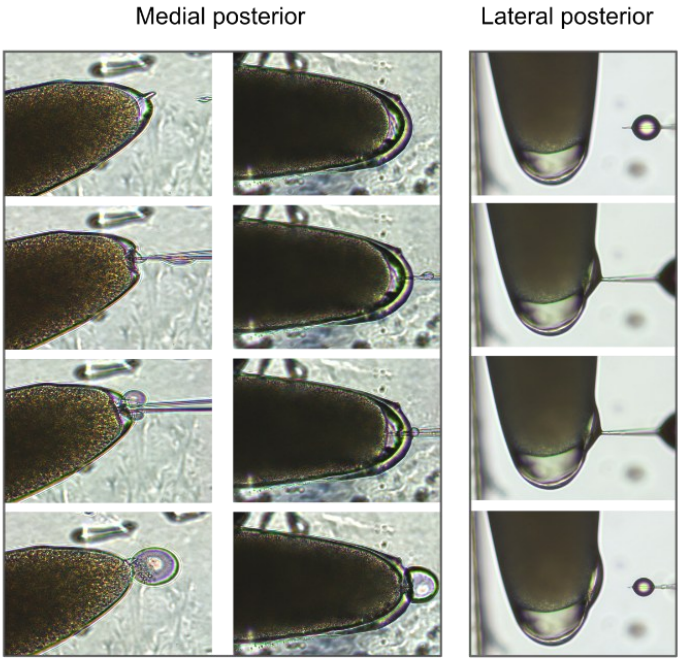
B) Needle priming and flow rate



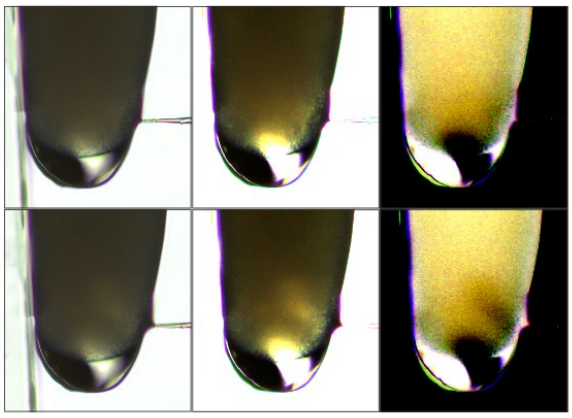
C) Egg polarity



D) Injection site



E) Target injection volume



F) Post injection feed and transfer

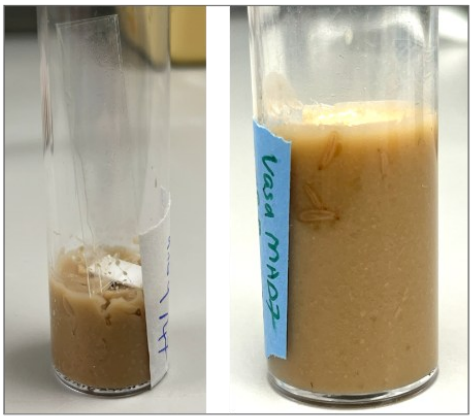


Figure 32: Protocol optimisations and embryo features identified in BSF microinjection setup. **A)** Embryos were aligned vertically and injected laterally into the posterior pole. The needle was fixed perpendicularly to the exposed side of the embryos at an attack angle of between 20° - 45°. Embryos were not explicitly aligned in the same dorso-ventral axis, and no difference in survival was observed when injecting at the angles within this range. **B)** Tip opening was performed when injection needles became clogged or were first mounted **(*i*)**. The needle tip was brought into contact with the side of a coverslip **(*ii*)** until liquid was seen to flow from the tip ***(iii)***. The tip was then retracted and excess pressure applied through the microinjector to increase the flow of liquid and trapped air through the needle **(*iv-v*)**. Once purged, the backpressure supplied to the needle through the microinjector was adjusted until a steady, continuous efflux of small droplets of injection mix was seen from the tip **(*vi*)**. **C)** The morphology of developing embryos allowed for identification of the anterior and posterior poles when preparing to inject. The yolk at the anterior pole had a translucent, “bubbly” appearance, while yolk was uniform and dense at the posterior. Retraction of the posterior pole from the chorion as the embryo develops left a characteristic vitelline void space, with the yolk compartment having a clear boundary. **D)** Posterior injections were initially performed by inserting the needle at the embryo horizontal midpoint, passing through the vitelline space before entering the yolk compartment. Removing the needle from the medial posterior site commonly led to leakage of the yolk to the exterior (left) or into the vitelline space (middle), with minimal survival observed following injection. Vertical alignment allowed for lateral injection into the posterior pole. Lateral injections were performed close to the posterior yolk boundary through the embryo side, bypassing the polar vitelline space (right). A smaller internal needle travel was required to reach the yolk, leaving a smaller injection scar with minimal exterior leakage observed, and the boundary between the yolk and vitelline space remaining intact. **E)** As efflux from the needle point was continuous, the appropriate volume for injection was determined by observing for signs of injection mix entering the yolk and the microinjection pressure adjusted accordingly. At the correct pressure, a small cloud of pale mix was seen forming in the yolk from the inserted needle tip (bottom left). False colour images are included (bottom middle, bottom right) to clearly indicate the ideal injection volume and location in the posterior yolk. The needle should be removed after a volume of mix is observed in the egg to avoid excessive injection volumes. **F)** As described above, placing embryo-aligned coverslips into agar feed vials prior to the expected hatch date removes the need for *ad hoc* manual transfer of hatching neonates, avoiding the risk of starvation or migration away from the slide if not transferred immediately. Between 1 and 150 hatching neonates were reared successfully in agar vials following self transfer from coverslip assemblies following injection, with no immediate adjustment to feed amount being necessary for different injection survival rates.

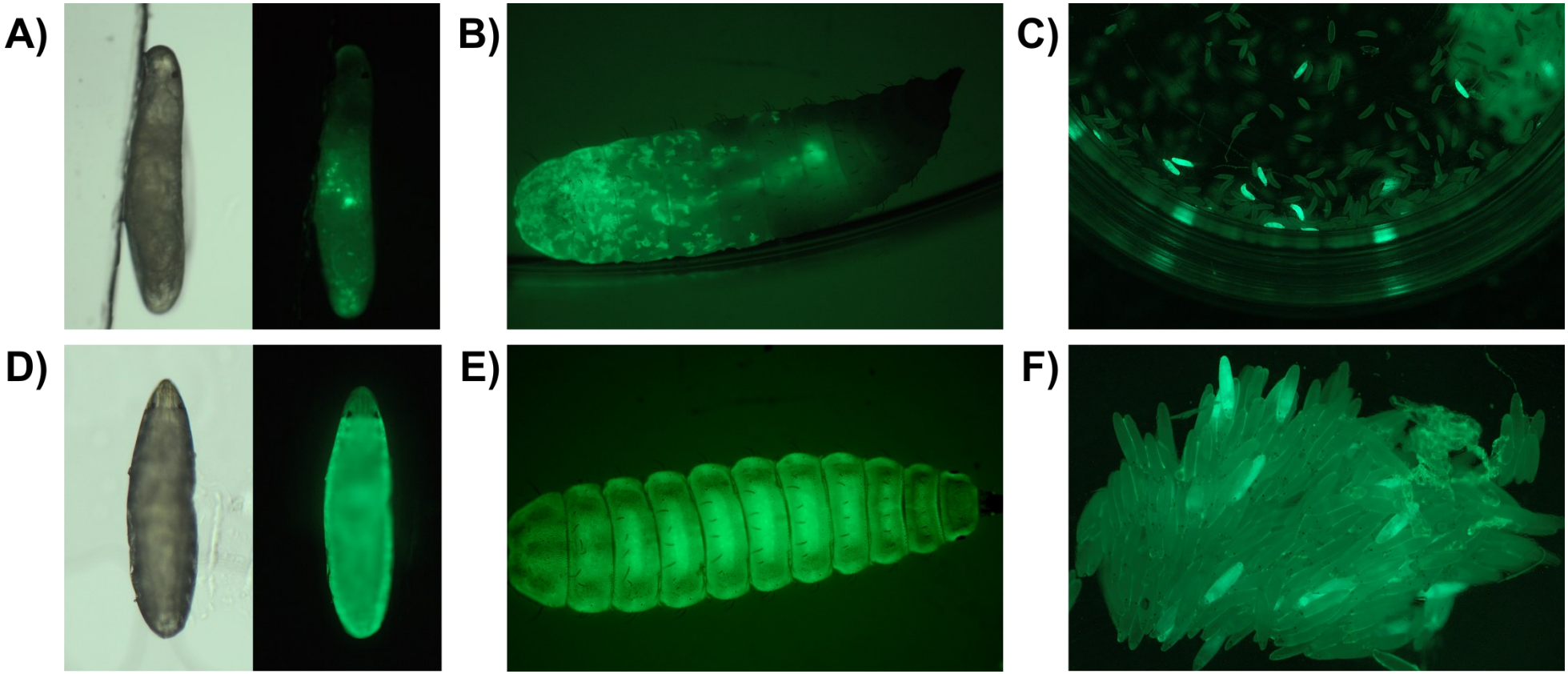


Figure 33: Transient and stable expression of fluorescent marker transgene at the embryo and larval stages. A) Embryos injected with Hr5IE1-GFP-encoding *piggyBac* constructs displayed a mosaic pattern of marker expression from around 72 hours post injection, localised around the posterior injection site. **B)** Large larva (~12 days post hatching) retain clear signs of transience emanating from the posterior pole, indicating successful construct integration and expansion of transgene-encoding tissues. **C)** Transgenic G1 neonates, with the construct integrated across the organism, can be detected upon hatching via stable expression of the *Hr5IE1*-GFP marker. Fluorescent larva can be manually separated from non-transgenic siblings in the same egg clutch for rearing separately. **D-E)** Transgenic G1 that stably encode the marker across the organism can be distinguished from transiently-expressing G0 predecessors from the late egg development stage throughout all larval stages. **F)** Clutches that contain transgenic individuals can similarly be identified *via* fluorescence prior to embryo hatching, allowing for an earlier screening of egg containers before transfer to feed is required.

Microinjections performed following the initial protocol

Construct 1	Construct 2	Eggs injected	G0 larva hatched	G0 larva transients	G0 pupae
Water (injection control)	-	**470**	**24** (5.1%)	**0**	**0**
***Hr5IE1*-GFP** (500 ng / µL)	**I – HyPBase** (300 ng / µL)	**251**	**16** (6.4%)	**0**	**0**
***Nanos*–Cas9** (500 ng / µL)	**I – HyPBase** (300 ng / µL)	**221**	**0**	-	-
***Vasa*–Cas9** (500 ng / µL)	**I – HyPBase** (300 ng / µL)	**164**	**0**	-	-
***Nanos*–MAD7** (500 ng / µL)	**I – HyPBase** (300 ng / µL)	**74**	**0**	-	-

Included vertical embryo alignment, injection site at lateral posterior

Water (injection control)	-	**352**	**72** (20.5%)	-	-
Injection buffer, pH 7.2 (injection control)	-	**251**	**50** (19.9%)	-	-
***Nanos*–MAD7** (500 ng / µL)	**I – HyPBase** (300 ng / µL)	**69**	**0**	-	-
***Vasa*–MAD7** (500 ng / µL)	**I – HyPBase** (300 ng / µL)	**546**	**83** (15.1%)	**23** (27.7%)	**0**
U6:1 crRNA-*wx*2 (500 ng / µL)	**I – HyPBase** (300 ng / µL)	**524**	**116** (22.1%)	**0**	-
U6:2 crRNA-*wx*2 (500 ng / µL)	**I – HyPBase** (400 ng / µL)	**252**	**56** (22.2%)	**0**	-

Table 3: Chronological embryo microinjection experimental survival data before process optimisation. Water, injection buffer or plasmid *piggyBac* donor and helper constructs at a combined DNA concentration of between 800-900 ng / µL were injected according to the methods outlined in sections 4.2.4 and 4.2.7. With the initially employed microinjection protocol, the rate of successful G0 hatching ranged between 0.0% and 6.4%, with no transient marker expression observed and no G0 survival encountered beyond the hatched neonate stage. The change in how embryos were aligned to allow lateral injection into the posterior yolk, therefore avoiding penetration through the void space at the embryo pole, increased injection survival and G0 hatch rate to between 15% and 22%. The first indication of transient marker expression in surviving G0 was noted in around 28% of hatched neonates injected with the *Vasa*-MAD7 donor construct, although no G0 pupa were recovered from any of these injection sets. Based on this data, a post-injection G0 rearing method was produced to improve the survival of hatched G0 neonates to facilitate outcrossing.

	Construct 1	Construct 2	Eggs injected	G0 larva hatched	G0 larva transients	G0 pupae	G0 adult males	G0 adult females	G0 adults total	G0 cross	G1 clutches total	Transgenic containing G1 clutches	Transgenic G1 larva
Thinner injection needle with increased injection pressure, agar diet	***Nanos*-Cas9** (500 ng / µL)	**I – HyPBase** (400 ng / µL)	**480**	**213** (44.3%)	**106** (49.8%)	**85** (39.9%)	**43**	**36**	**79** (92.9%)	**40 male G0** x **32 female G0** 20 male WT x 18 female WT	**24**	**4** (16.7%)	**33**
	***Vasa*-MAD7** (500 ng / µL)	**I – HyPBase** (400 ng / µL)	**348**	**176** (50.6%)	**84** (47.7%)	**45** (25.6%)	**17**	**18**	**35** (77.8%)	**17 male G0** x **18 female G0** 15 male WT x 25 female WT	**11**	**3** (27.3%)	**54**
	***Dm*_PUb GFP / DsRed** (500 ng / µL)	**I – HyPBase** (400 ng / µL)	**240**	**201** (83.8%)	**36** (17.9%)	**28** (13.9%)	**11**	**16**	**27** (96.4%)	**11 male G0** x **16 female G0** 29 male WT x 14 female WT	**16**	**2** (12.5%)	**17**
	Cas9 protein (700 ng / µL)	**sgRNA-*w*** (500 ng / µL)	**330**	**157** (83.8%)	**0** (*mosaics*)	**89** (56.7%)	**36**	**31**	**71** (79.8%)	**36 male G0 x 31 female G0** 14 male WT x 19 female WT	**26**	**6** (23.1%)	**84** (*w* mutants)
	U6:3 gRNA-*w* (500 ng / µL)	**I – HyPBase** (400 ng / µL)	**285**	**179** (47.6%)	**102** (57.0%)	**95** (53.1%)	**45**	**47**	**92** (96.8%)	**45 male G0 x 47 female G0** 15 male WT x 13 female WT	**9**	**4** (44.4%)	**78**
Increased DNA concentration, G0 agar vial self-transfer	**U6:3 crRNA-*wx2*** (600 ng / µL)	**I – HyPBase** (400 ng / µL)	**290**	**242** (83.4%)	**155** (64.0%)	**152** (11.4%)	**67**	**58**	**125** (82.2%)	**67 male G0** x **55 female G0** 13 male WT x 25 female WT	**29**	**5** (17.2%)	**> 250**
	U6:1 gRNA-*w* (600 ng / µL)	**I – HyPBase** (400 ng / µL)	**381**	**348** (91.3%)	**218** (11.4%)	**201** (62.8%)	**94**	**88**	**176** (87.6%)	**90 male G0** x **88 female G0**	**34**	**8** (23.5%)	**54**
	U6:3 crRNA-GRGx6 (600 ng / µL)	**I – HyPBase** (400 ng / µL)	**290**	**218** (83.4%)	**110** (62.6%)	**73** (33.5%)	**32**	**32**	**64** (87.7%)	**32 male G0 x 31 female G0** 28 male WT x 29 female WT	**12**	**3** (25.0%)	**124**
	***Vasa*-Cas9** (600 ng / µL)	**I – HyPBase** (400 ng / µL)	**350**	**311** (75.2%)	**271** (87.1%)	**260** (83.6%)	**101**	**112**	**213** (81.9%)	**101 male G0** x **110 female G0**	**19**	**7** (36.8%)	**> 238**
	U6:2 gRNA-*w* (600 ng / µL)	**I – HyPBase** (400 ng / µL)	**300**	**235** (78.3%)	**157** (66.8%)	**114** (48.5%)	**42**	**47**	**89** (78.1%)	**42 male G0** x **47 female G0** 18 male WT x 23 female WT	**13**	**5** (38.5%)	**43**
	***Act5C*-Cas9** (600 ng / µL)	**I – HyPBase** (400 ng / µL)	**380**	**217** (57.1%)	**124** (57.1%)	**78** (35.9%)	**33**	**34**	**67** (85.9%)	**33 male G0** x **34 female G0** 17 male WT x 16 female WT	**8**	**4** (50.0%)	**> 250**
	***Act5C*-MAD7** (600 ng / µL)	**I – HyPBase** (400 ng / µL)	**306**	**238** (77.8%)	**163** (68.5%)	**112** (47.1%)	**56**	**45**	**101** (90.2%)	**56 male G0** x **45 female G0** 4 male WT x 15 female WT	**17**	**7** (41.2%)	**148**

Table 4: Chronological embryo microinjection survival and transgenesis data for established transgenic and gene-edited lines. The combination of refining the parameters of the injection needle with the previous embryo alignment and injection steps substantially increased the number of G0 neonates surviving injection (44.3% - 91.3% G0 hatch rate across all injection sets). A general increase in injection survival was noted over time, which is likely representative of more practice and comfort with the microinjection method as the injection schedule progressed. G0 pupa were recovered for the first time following the introduction of a post-injection rearing method using agar diet, with previous G0 neonate survivors assumed to have had poor survival due to the conditioning of the standard chickcrumb-based diet at low larval densities. G0 flies were self-crossed in a single cage, with wild-type flies supplemented to increase cage density. Transgenic G1 larva, or *w* mutants in the case of the Cas9 / sgRNA-*w* RNP injection set, were recovered from a minimum of two independently collected G1 clutches, with transgenics recovered from between 12.5% - 50.0% of all G1 clutches collected depending on the injection set. The proportion of transient neonates and G0 pupa survival when compared to the number of hatched G0 both increased with increased plasmid DNA concentration and the addition of a neonate self-transfer step, which was generally reflected in the higher number of transgenic G1 larva recovered. Neither of these steps was considered essential for generating transgenic lines, however, with the combination of injection site selection, needle optimisation and the use of an improved G0 feeding medium likely to be the key improvements facilitating the microinjection process in this study.

4.2.11 – Characterization of Promoters for Constitutive Transgenic Marker Expression

For the planned CRISPR transgenic line crosses, the inheritance of one or both CRISPR transgene constructs would need to be easily tracked in the progeny. The nuclease constructs were therefore designed to include a GFP marker and the RNA constructs a DsRed marker, which each fluorescent protein not affecting the identification of the other. To further distinguish the type of construct encoded, different candidate promoters were used to control the expression of the two markers.

The Hr5IE1 promoter was chosen for GFP expression in nuclease encoding projects. Although since reported, use of this promoter had not been described in BSF at the time of cloning[440]. However, as the IE1 promoter element had previously been characterized in the species without the Hr5 enhancer element, it was assumed that marker expression with the combined elements would be strong and constitutive as seen with other insect species[422,532,533].

As seen previously, Hr5IE1 was highly efficient for producing transient expression in G0 individuals produced during transgenesis. Strong GFP expression could be seen localised around the posterior site of injection in transient individuals from around 3 days post injection. The phenotype remained as the larva grew, with a clear mosaic patten of GFP indicating integration in the posterior tissues. Mosaicism was also noted in adult flies with this promoter (not shown, but distribution and intensity of expression was comparable to *Opie2* – marked transient adults depicted later).

In full transgenic G1 generations and above, GFP expression was noted in all development stages when using *Hr5IE1* (**Figure 34**). As with transients, expression was first indicated around 3 days after laying in egg clutches, with identification of transgenic and non-transgenic siblings in a clutch becoming clear (**Fig. 34A-B**). All larval stages were associated with strong expression in the majority of the cuticle, with the ability to differentiate between heterozygote and homozygote *Hr5-IE1* encoding transgenics possible due to a slight increase in fluorescence intensity in homozygotes (**Fig. 34C-D**). Marker intensity began to decrease in pre-pupa in line with cuticle darkening and pupal formation, although strong expression was still apparent near the centre of the pupa (**Fig. 34E**). In male and female adults, the *Hr5IE1*-marker was distributed in the majority of soft cuticle tissues connecting the hard, black cuticle abdominal plates and thorax cuticle sections (**Fig. 34F-G**). Expression

was also apparent throughout the translucent abdominal window, as well as the off-white cuticle of the mouthpiece, halters and the tibia and tarsus sections of the leg. During dissections, the fat body and ovaries in the abdomen also appeared strongly fluorescent (data not shown). *Hr5-IE1* was considered a highly successful promoter for strong and constitutive marker expression in BSF, at all development stages and in multiple tissue types.

Knowing that *Hr5IE1* would lead to a broad distribution of marker expression in the nuclease lines, the CRISPR RNA-expressing lines were designed to preferentially express DsRed in a distinct tissue type. The 3xP3 promoter was selected for these constructs, understood to direct marker expression to the eyes and neuronal tissues in multiple insect species[534–536]. Despite the two construct types used in this study encoding different fluorescent markers, the promoter was chosen for additional characterisation of potential transgenesis tools in BSF, where few transgenesis components commonly employed in other insects have been disseminated in the wider literature.

While performing the transgenesis injections, it became apparent that while G0 survival was similar between eggs injected with the *Hr5-IE1* and *3xP3* encoding constructs, no transient DsRed expression was recorded and no transgenic G1 recovered from G0 crosses. With suspicions regarding the functionality of the untested *3xP3* in BSF, alternatives were explored for producing clear marker expression.

In a prior study involving transgenic *C. capitata, piggyBac* constructs were designed with the GFP and DsRed markers being expressed under a 2.0 kb proximal polyubiquitin (PUb) promoter isolated from *D. melanogaster*[417,477]*.* Producing strong marker expression at multiple life stages, the promoter represented an endogenous regulatory element from a highly conserved and globally expressed gene in eukaryotes[537]. While a promoter region from the PUb orthologue endogenous to BSF has since been successfully characterized, the conserved pattern of expression between species suggested that the *D. melanogaster* promoter (*Dm* PUb) would likely show similar activity[440].

A cassette encoding the *Dm* PUb-eGFP and PUb-DsRed genes in tandem, each with an SV40 terminator, was amplified from a *piggyBac* construct developed during the aforementioned study in *C. capitata*[477]*.* Once cloned into the *piggyBac* backbone produced earlier, the PUb GFP / PUb DsRed construct was used in transgenesis injections (**Table 4**).

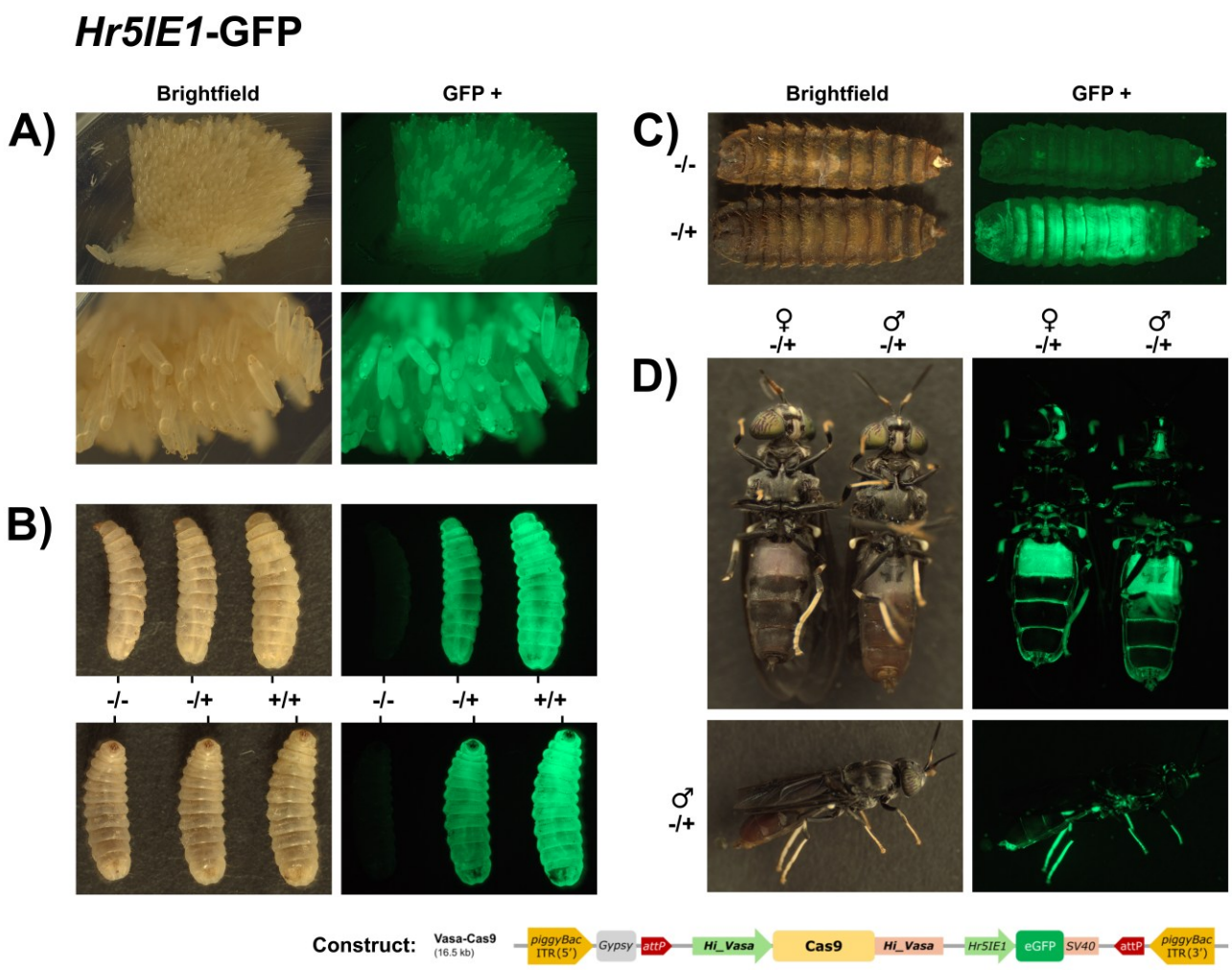


Figure 34: The *Hr5IE1*-GFP promoter-marker combination permits clear expression of transgenic marker in multiple tissues throughout all BSF development stages. A) *Hr5IE1*-GFP encoding individuals could be distinguished from non-transgenic siblings from as early as 72 hours AEL. Mixed transgenic / non-transgenic containing clutches were reared without separating the larva for simplicity, with fluorescent transgenics being isolated at the adult breeding stage for cage set up and line maintenance. **B)** Strong marker expression was observed in the majority of tissues throughout the larval stage in transgenic heterozygous (-/+) and homozygous (+/+) for *Hr5IE1*-encoding constructs. A slight, but notable increase in fluorescence intensity was also apparent in homozygous larva. **C)** Marker intensity was decreased in pupa at the periphery of pupal casings, but remained strong in the central section throughout pupa development. **D)** *Hr5IE1* marker expression was distributed throughout the soft and translucent cuticle tissues in male and female transgenic adults, with characteristic fluorescence observed in the abdominal window, tarsi and mouthpiece sections. All transgenics depicted encoded the *Vasa*-Cas9 *piggyBac* construct, although marker tissue distribution and intensity was identical for all *Hr5IE1*-GFP encoding lines.

Despite only a weak transient expression of GFP being recorded at G0, G1 positive for the construct were recovered following G0 cross (**Figure 35**). Although not apparent at the embryo stage, like with stronger *Hr5IE1*-associated expression, both fluorescent markers appeared broadly expressed in the neonate stage after hatching. Larval stages continued

to display high marker expression, which became less evident in the pupal stage (data not shown). Adults also showed clear expression, although with a different distribution than encountered with *Hr5IE1*; expression was weak around the mouthpiece, halters, legs and soft connecting cuticle tissues, while a striated pattern of fluorescence was seen through the abdominal window tissue, contrary to the level plane of expression seen with *Hr5IE1.* Additionally, a striking marker pattern was visible through at dorsal thorax through the black cuticle tissue. As PUb expression is high in striated muscles, the marker genes presumably illuminate the extensive thoracic flight musculature responsible for controlling the wings during flight[538].

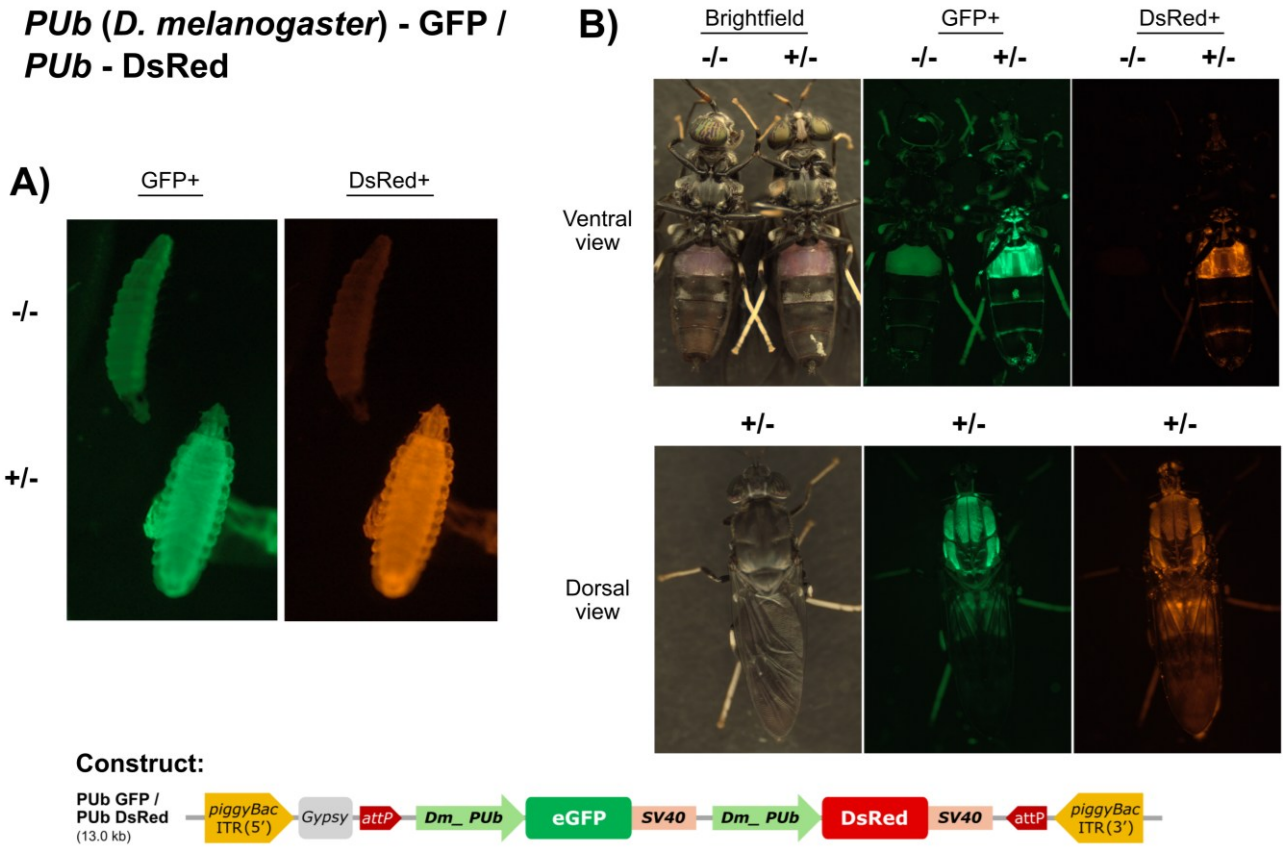


Figure 35: The *D. melanogaster PUb* promoter was effective for flourescent marker expression at both larval and adult stages. A) A construct with tandem eGFP and DsRed markers under two copies of the *D. melanogaster Polyubiquitin* promoter allowed for sufficient identification of transgenic individuals from the neonate up to the adult stage. **B)** The pattern of *PUb*-driven marker expression appeared similar to that seen when using the *Hr5IE1* promoter, with the exception of high additional expression around the presumed flight musculature in the adult thorax when viewed from the dorsal side.

Concurrently with the PUb characterisation, the smaller synthetic promoter *Opie2* was evaluated for marker expression in RNA-expression constructs. If functional, this element was preferable to *Dm* PUb due to its smaller size (541 bp) when introduced with the sufficiently-large transgenesis constructs. Analogous to baculovirus *Hr5IE1, Opie2* is derived from the IE1 promoter from *Orgyia pseudotsugata* multicapsid nucleopolyhedrovirus (OpMNPV), although it lacks an additional upstream enhancer element[539]. Generally understood to be weaker than *Hr5IE1*, a similar pattern and distribution of marker expression was expected with *Opie2*[540]. When cloned into the RNA-expression constructs, *Opie2* produced clear transient expression of the DsRed marker in G0 injected larva around the injection site (**Figure 36A**).

In *Opie2*-transient adults, a speckled pattern of expression presumably corresponding to transgene integration was visible around the abdominal soft cuticle tissues and through the translucent window. G1 individuals depicted a strong global expression of DsRed at the larval, pupal and adult stages in the same tissue types as *Hr5IE1* (**Fig. 36B**). Interestingly, marker was not detected in developing embryos or for several days in neonates post hatching. Expression started to become apparent around 4 days post hatching, predicted to be the 3rd instar, originating from a small patch in the abdomen presumed to be the fat body store (data not shown). Expression was significant from this point onwards, with homozygote larva appearing visibly pinker than heterozygote and *wild-type* counterparts. Despite the larval phenotype, it was noted that *Opie2*-DsRed fluorescence appeared much fainter in adults than encountered with the PUb-DsRed transgene, and much less than GFP in the *Hr5IE1* nuclease transgenics. With an evident G0 transience phenotype, however, all RNA-expression constructs were subsequently converted from encoding *3xP3*-DsRed to *Opie2*-DsRed for simplified production of these transgenic lines.

Although the *H. illucens* PUb promoter has since been characterised for transgene expression in the species, the validation of the *D. melanogaster* PUb variant still increases the repertoire of effective transgenic promoter candidates for use in future studies. Similarly, the *Opie2* promoter, with it's use not previously described in BSF, demonstrated high activity and suitability for marker expression in transgenic constructs. With increasing numbers of validated promoter candidates with distinctive expression profiles, more flexibility and complexity can be afforded in the design of future BSF transgenesis studies.

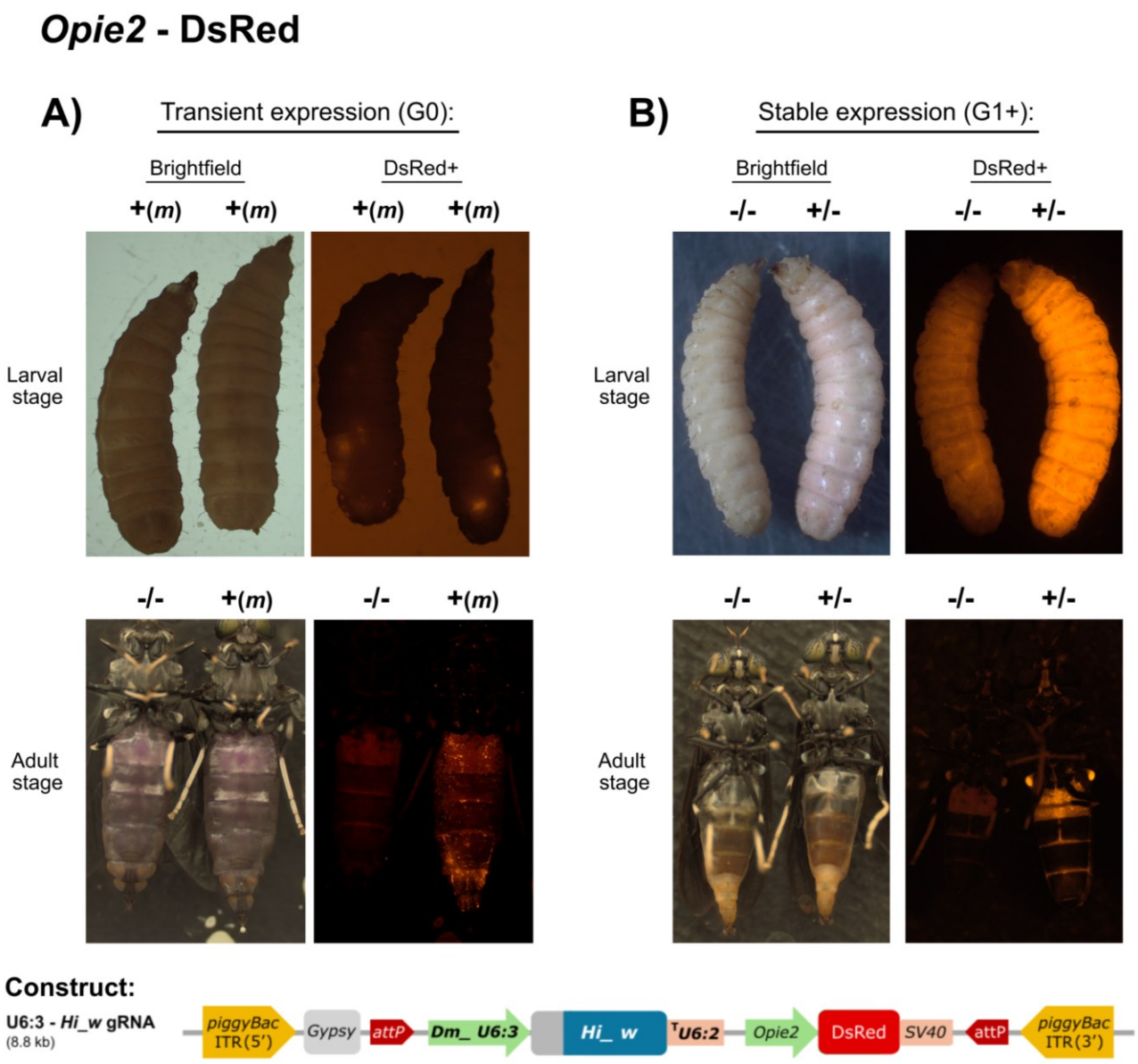


Figure 36: The synthetic *Opie2* promoter was effective for marker production at the larval, pupal and adult stages. A) While DsRed marker expression under the *Opie2* promoter was not evident in the embryonic stage and early larval instars, the promoter allowed for sufficient transient expression in G0 from late larval into the adult stages with a similar distribution to *Hr5IE1* promoter-associated marker expression. **B)** With stable construct integration at the G1 and later stages, *Opie2* permited a substanital level of marker expression in later larva to the adult stage, with expression found to be evident in larva from around 4 days after egg hatching. The combination of high larval expression with the DsRed marker gave transgenic larva a characterisitc pink hue, which was notable to the naked eye in larger larva without the use of a flourescence screening system. The images of transient and stable *Opie2*-linked marker expression depicted here utilised late-stage larva (6-10 days after hatching) and adults from the G0 and G1 generations of the U6:3-*Hi_w* gRNA transgenic line respectively.

4.2.12 – Production of a BSF *White* Knock-Out Line by Cas9-RNP Microinjection

With efficient production of *piggyBac*-mediated transgenics using the improved microinjection method, the gRNA-*w* sequence designed and tested earlier in the study was validated for mutagenic activity *in vivo.* Although briefly described in a patent, the analysis and knock-out phenotype of *H. illucens w* had not been covered in a publication at the time of the experiment performed in this study[435,541]. As the later planned validation of gene editing by transgenic cross depended on a clear mutant phenotype, the $w^{-/-}$ line was produced through embryonic microinjection of a classical Cas9-RNP complex.

As the mutant phenotype was expected to be detectable in the eye spots in neonate larva, a set of injections were performed by first targeting the site in the egg where eyes were expected to be located, near the anterior pole. The 157 G0 injection survivors showed no clear mosaic phenotype at either the larval or adult stages (data not shown). To maximise the chance producing hereditable mutations, a second set of injections targeting the posterior pole was performed. A high survival rate following injection was recorded (83.8%), although *w* mosaicism was not noted in the injection survivors when examined at larval or adult stages. A G0 cross was performed with the 36 male and 31 female adult survivors. As low breeding activity was observed initially, the number of each sex in the cage was increased to 50 by supplementing T4 *wild-type* flies. 26 fertilised clutches were produced from the cross, with each being examined for the presence of G1 neonates with the *w* phenotype. 84 mutant neonates were encountered across 6 clutches (23.1%), which were pooled in a single vial of agar diet for rearing until adulthood.

The $w^{-/-}$ larva had a clear phenotype that was apparent at the point of eye-spot development in 3-day old eggs (**Figure 37**). Mutant larvae lacked the otherwise distinct red eye-spot pigmentation seen in BSF neonates, appearing at this stage as if they lack eyes completely. The phenotype became less apparent as the larva grew and the eye structure became more pronounced, presumably due to an increasing contribution of other cuticle pigments to the appearance of the eye (image not shown). The phenotype became apparent again at the adult stage, with mutants having uniformly grey compound eyes, as opposed to the green and purple striped pigmentation characteristic of *wild-type* BSF.

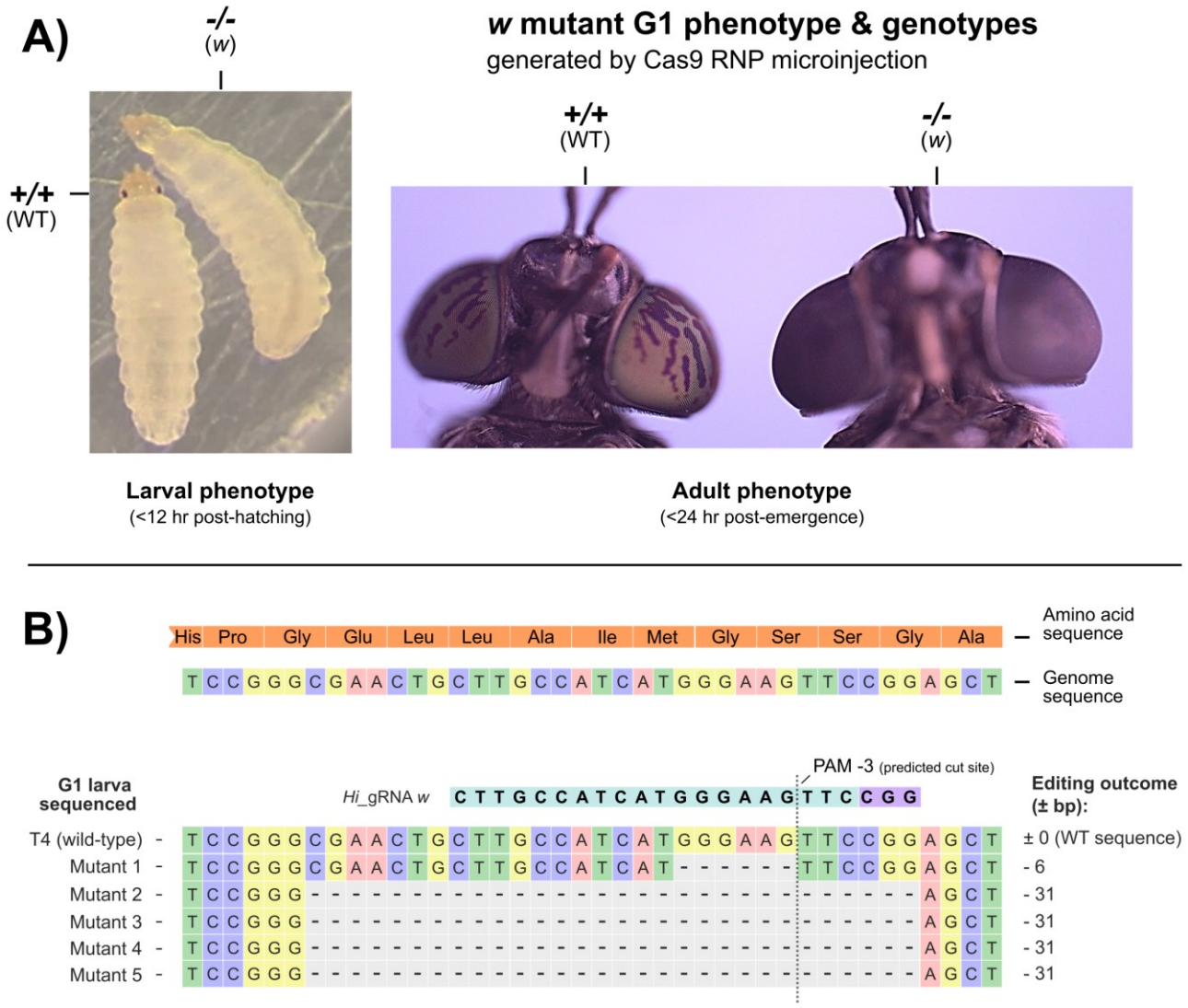


Figure 37: BSF *w* mutants have a clear eye pigmentation phenotype at the larval and adult developmental stages. A) *Wild-type* larva with a functional copy of *white* displayed a clear red pigmentation in the eye spots, whereas null *w* mutants had minimal pigmentation throughout the neonate and larval stages. Additionally, WT adult flies featured green compound eyes with a distinctive purple patterning, while mutants had solid grey-black colouration with no obvious pattern of pigmentation. **B)** Sanger sequencing of the *white* target from several G1 larvae with the *w* phenotype, alongside a red-eyed WT, suggested indel formation around the predicted gRNA cut site in the predicted mutant individuals. Four of the five mutants possessed the same 31 bp frameshifting deletion encompassing the gRNA target sequence, which may indicate a common G0 parent. No obvious difference was seen in the degree of the *w* larval phenotype between the frameshifted mutants and one larva with a two amino acid in-frame deletion. This could evidence that the highly-conserved amino acids at the chosen target site are essential for gene function.

Five of the pooled G1 mutant larva were sampled for gDNA extraction and Sanger sequencing of the *w* target site, with the T4 *w* amplicon submitted as a non-mutant comparison. Only two indel profiles were detected in this initial screen, with one larva encoding a 6 bp deletion immediately downstream of the expected gRNA cut site (**Fig. 37B**). This genotype resulted in two amino acid deletions and one substitution at the 101-103

position in the *w* protein coding sequence, but interestingly produced no frameshift. This further supports the conclusion that the residues in this position are critical to the function of *w,* supported by the relative conservation at this region across different species. The remaining four larvae, presumably from common G0 parents, encoded a large frame-shifting deletion of 31 bp, spanning from -25 to +6 bp relative to the predicted cut site. Only a single sequence trace was detected during sequencing for each of the *w* mutant larvae (data not shown), consistent with genotypes that are homozygous for a single indel repair outcome inherited from both G0 parents.

While the possibility remains that the genotypes encountered represent the most common repair outcomes following cleavage at this locus, a more plausible explanation could be that only one of the mutant alleles encoding a relatively small deviation from the wild-type sequence was amplified by PCR prior to sequencing. An allele with a large insertion or deletion may not have been effectively amplified under a PCR condition and primer set optimised for the wild-type sequence, thereby resulting in only one of the mutant alleles being produced for analysis. While the null phenotype for the dominant *white* gene does indicate the absence of a functional wild-type allele in the G1 mutants, alternative sequencing techniques such as long-range PCR would likely be required to confidently detect mixed mutant genotypes of alleles with significantly different lengths or gene structures in the G0 progeny.

The $w^{-/-}$ mutant line was maintained successfully for 3 more generations, with no apparent changes in insect behaviour, mating activity or fecundity observed. Later Sanger sequencing of the progeny of a G3 egg clutch confirmed the inheritance of the 31 bp deletion genotype across the population (data not shown). This early characterisation of the *w* line provided a clear reference for the phenotype to look for in the progeny of later transgenic crosses performed to test the functionality of the endogenous CRISPR toolkit.

4.3 – Conclusion

In this chapter, a contained, climate-controlled insectary space for rearing all BSF life stages was successfully built, allowing for the establishment of a limited colony of *wild-type*, transgenic and knock-out mutant lines. Several steps were taken to produce a detailed, simple and efficient embryo microinjection method for establishing BSF transgenic lines using the *piggyBac* transposon system. This encompassed optimisations in egg collection and alignment methods, the timing and location of injection, injection mix composition, injection needle parameters and manipulation, and post-injection rearing of limited numbers of hatching larva.

The described microinjection technique proved highly reliable and easily reproducible with limited equipment, and led to injection survival rates higher than, to our knowledge, any other BSF microinjection method encountered in the literature. Similarly, high numbers of transgenic G1 could be reliably produced from a single injection session with around 300 G0 injections performed. With the constructs employed in this study, stable transgenic lines could be established with as few as 15 transgenic G1. The *Opie2* and *Dm*-PUb promoters, previously uncharacterised in this species, were successfully employed for transgenic marker expression at multiple development stages. Finally, the *w* target gene and the previously designed gRNA-*w* spacer were validated *in vivo* to produce a clear *w* mutant phenotype, confirming the target as an appropriate readout for benchmarking the efficiency of the endogenous CRISPR toolkit for creating gene edits upon cross of nuclease and CRISPR RNA transgenic lines.

Chapter 5: **Evaluating a Transgenic Cross Strategy for Endogenous Gene Editing in Black Soldier Fly**

5.1 – Introduction

After optimising the embryonic microinjection method for transgenesis in BSF, several *piggyBac* constructs from the CRISPR component library generated earlier in this study were successfully incorporated into the BSF genome as stable transgenic lines. 5 of the 6 produced nuclease constructs, identified by GFP marker expression, were integrated to yield the *Nanos*-Cas9, *Vasa*-Cas9, *Vasa*-MAD7, *Act5C*-Cas9 and *Act5C*-MAD7 lines. Regrettably, a *Nanos*-MAD7 could not be established in time to give a sufficient number of generations needed for full analysis and cross experiments to be performed; the process of characterisation took 2 – 3 generations, and around 4 - 5 months, for the other nucleases.

In total, 5 CRISPR RNA-expressing lines were generated, identified by expression of the DsRed marker. 3 Cas9-compatible guide RNA-expressing lines were produced for targeting the BSF *w* gene for knock out, differing by the Pol III promotor used (U6:1, U6:2 or U6:3). Each gRNA-*w* line encoded the same gRNA spacer sequence that was validated earlier in this study with a T7 RNA polymerase-synthesised sgRNA, both *in vitro* by Cas9 cleavage assay and *in vivo* through microinjection as a Cas9-RNP complex for the *BSF w* mutant line ($w^{-/-}$). Additionally, 2 multi-target crRNA-expressing lines were generated for crossing to MAD7 lines: U6:3-crRNA *w*x2 encodes an array with 2 spacers for target sites in *w,* while U6:3-crRNA GRGx6 targets the 6 growth regulatory gene candidates identified earlier for producing a larger larval mass or decreased development time as a knock-out phenotype.

While plasmid variants of both the crRNA constructs integrated into BSF were produced encoding the U6:1 and U6:2 promotors, these were not chosen for use in transgenesis. This decision was made due to the amount of additional space that would be required for rearing the extra 4 lines, in addition to the nuclease lines, that was unavailable with the current insectary setup. The characterisation and comparison between the three U6 promoters was therefore limited to the Cas9-gRNA lines only, and U6:3 chosen as the single Pol III promoter to continue with when establishing the MAD7-crRNA lines. The *D. melanogaster* U6:3,

although the only produced U6 promotor not endogenous to BSF, was selected for transgenesis because of its proven track record for effective gRNA expression across multiple different insect species, as well as its promising performance with MAD7 crRNA array expression as indicated earlier in S2 cell culture experiments.

As mentioned in a previous section, the *3xP3* promoter originally chosen for DsRed marker expression in the RNA lines and was later exchanged for the *Opie2* promoter when no *3xP3* line could be established, or at least easily identified. Although this lacks the benefit of a clear differential expression of markers by tissue when crossed with the Hr5IE1-GFP positive nuclease lines, *Opie2* still provides a strong and distinctive pattern of marker expression across multiple life cycle stages. Although lacking the embryo and neonate marker expression profile encountered with Hr5-IE1, *Opie2* positive RNA lines could be identified quickly at the early larval stage, which was estimated to be around the 3rd instar based on larval size and age. DsRed marker expression was evident in larva when still in the smaller meal container size used for the first feed, and so screening and separating larva at this point was still trivial compared to rearing all possible transgenics to the adult stage before identification of a *3xP3* associated marker, which would be expected according to recently published data characterising the promoter in BSF[440].

Having established an efficient method for BSF transgenesis, the transgene-based CRISPR toolkit would need to be characterized, both as the individual components and when combined *via* crossing lines. As detailed later, significant hurdles to establishing a stable transgenic line can still be encountered after recovery of multiple transgenic G1 larva, which while not specific to BSF may still be exacerbated in this species by their breeding requirements required for their artificial rearing. Certain transgenes, although viable when integrated at G0 in the germline but in an otherwise mosaic distribution, may result in an associated fitness cost when fully integrated in all tissues at the G1 stage. While the procedures spanning the embryonic microinjection to successfully generating transgenic G1 may be the most technically challenging, safeguarding and expanding the genotype into scaled-up stocks of defined, homogenous integration sites from G2 and later generations still requires attention.

Once stable G2 lines were produced, presuming a single transgene integration event per egg clutch, the BSF lines were characterized for the specific location of the *piggyBac* insertion in the genome. A rough examination of potential inheritance bias for each construct was also made through backcrossing heterozygote transgenic individuals to a non-transgenic strain and screening the progeny. Once characterised individually,

combinations of the nuclease and RNA-encoding lines were crossed to yield a transheterozygotes F1 progeny with both CRISPR transgenes, identified by dual marker expression. The activity of both endogenously expressed Cas9 and MAD7 were then assessed in this and proceeding backcrossed generations by screening for the *w* mutant phenotype, and examined on the genetic level by DNA sequencing of the expected target site.

Although the mutant phenotype generated by the MAD7 x U6:3 crRNA GRG x6 cross would be more difficult to accurately categorise than the *w* mutant, this cross could still be made and transheterozygotes generated with the intention of examining any GRG mutant-associated phenotypes in a later growth assay. A larval growth assay was developed for use informed by several experiments outlined in multiple BSF developmental studies, including a recent preprint characterising mutants of Lipid storage droplets I (LSDI) for their larval mass and development phenotypes[439]. Although ultimately not performed due to an inability to detect indels in each GRG target gene over 3 backcrossed generations, the assay is described in the materials and methods section.

This chapter describes the preliminary characterisation of the CRISPR transgenic lines and then depicts the three nuclease-RNA line crosses and subsequent genotyping performed in this study. With the crosses proving apparently unsuccessful for generating mutants as intended, the chapter concludes by outlining several follow-up experiments in an attempt to deduce if the transgenic CRISPR components show any functionality independent of the original crossing strategy.

5.2 – Results and Discussion

5.2.1 – G1 to G3: Establishing Stable Transgenic Lines

For the nuclease constructs, resulting transgenic G1 neonate progeny from the G0 cross of injection survivors were identified by GFP marker expression and transferred upon hatching by needle point into a new source of feed, rearing them separately from their non-transgenic siblings. For egg collections that contained more positive G1s than could be counted accurately, indicated in table 4, the entire egg pot was added to a feed source and the larva reared collectively for a few days until large enough for easier separation of transgenics without the risk of overexposure. As *Opie2* was not evident until later larval instars, each G1 clutch produced was reared in its entirety until a point where *Opie-2* maker screening and transgenic isolation was possible.

Once at the adult stage, a G1 cross was set up using all female G1 and males from the T4 wild-type strain. Assuming each G1 female encoded a single transgene integrated at one genomic locus, and no transgenes were provided by the male, all transgenic G2 from a single egg clutch would inherit at the same integration site as the mother. Female G1 were chosen for the G1 cross, as males are considered more likely to mate multiple times than females, increasing the chance that any one G1 genotype would occur in multiple G2 clutches. Each G2 clutch was carefully collected to minimise the chance of contamination with eggs from other females, which could potentially mix transgenics of different integration sites in one clutch. With this in mind, G2 clutches were then considered as an isoline homogenous for a single transgene locus, with both the male and female transgenic then able to be crossed independently or to each other to expand the isoline without contamination and produce G3 heterozygote and homozygote progeny.

Each successful G1 cross produced multiple transgenic G2-containing egg clutches. As both male and female G0 survivors were crossed to each other in a single G0 cross, in an attempt to increase the chance of breeding and egg laying, each G1 larva in a single clutch produced by one female could have inherited a transgene integration from either parent. Given that the G0 parents are likely mosaic for different integration sites in their germline tissues, and that multiple gametes from each parent were likely used to produce the different individuals in the clutch, there is no guarantee that all G1 from one clutch inherited the transgene at the same integration site. Therefore, all transgenic G1 resulting from a G0 cross were pooled and reared together, with each being treated as potentially representing a different

integration for the same transgene from its direct siblings and other transgenics produced from the same G0 cross.

Due to the nature of the G0 cross, it was theoretically possible that G1 individuals could inherit multiple copies of the transgene at different integration sites, either from one or both parents. Although it was not ruled out, this was considered unlikely for the G0 cross and screening method employed in this study; each transgenic G1 clutch, even those containing >250 transgenics, also contained several-fold more siblings that didn't inherit transgenes, too many to realistically count to establish a specific transgenesis rate (number of transgenic larva / total number of larva) for each G0 pairing. With transgenics representing a small fraction of these clutches, it was assumed that transgenesis rate was low enough to make multiple integrations in one individual relatively improbable. As a distinction between heterozygotes and homozygotes could later be made in Hr5-IE1 and Opie2 – encoding lines based on the intensity of marker expression, the uniform expression seen during G1 screening also suggested that multiple integrations were likely not encountered.

For the majority of constructs, 33 to >250 transgenic G1 were recovered (**Table 4**) and developed as expected with no clear fitness cost (~80% surviving to adulthood) in the agar diet under the otherwise standard rearing scheme. The first 5 generated lines, which included transgenics marked by the *Hr5IE1*, *Opie2* and *PUb* promoters as well as the Cas9-RNP injected *w* $^{-/-}$ line, had an average G1 neonate (**G1-*n***) to adult (**G1-*a***) survival rate of 83.9% (SD = 4.7%), with no significant difference found in pairwise comparisons (X^2 = 5.85^{e-31} - 1.08, p = 0.29 – 1.00).

***Nanos*-Cas9** = 87.9% (G1-*n* =33, G1-*a* =29)

***Vasa*-MAD7** = 79.6% (G1-*n* =54, G1-*a* =43)

PUb-GFP/PUb-DsRed = 88.2% (G1-*n* =17, G1-*a* =15)

w $^{-/-}$ = 85.7% (G1-*n* =84, G1-*a* =72)

U6:3 gRNA-*w* = 78.2% (G1-*n* =78, G1-*a* =61)

For later injected constructs that produced over 100 transgenic G1, neonate to adult survival ratios were not specifically recorded, with the adult G1s separated by sex and used to set up reciprocal crosses to the T4 wild-type strain. Three constructs, however, were not successfully resolved from G1 to G2 for expansion and stabilisation of the line. Both U6:1 gRNA-*w* and U6:2 gRNA-*w* G1, although recovered as larva at around the same frequency as with the preceding constructs, suffered from a lower neonate to adult survival rate:

U6:1 gRNA-*w* = 31.5% (G1-*n* =54, G1-*a* =17)

U6:2 gRNA-*w* = 51.2% (G1-*n* =43, G1-*a* =22)

This resulted in a suboptimal G1 crosses being set up for both lines with the few surviving females (U6:1 gRNA-*w* = 6, U6:2 gRNA-*w* = 9). As suppressed breeding was expected due to low cage densities, T4 flies of both sexes were added to bring the number of each sex up to 25 for a total of 50 flies per cage. Additionally, reciprocal cages were set up using the G1 males for both lines (U6:1 gRNA-*w* = 11, U6:2 gRNA-*w* = 13). As male breeding behaviour was only apparent throughout this study with a density of >15 males, T4 males were supplemented to bring the male counts to 20. 30 female T4 were provided for excess mating opportunity while retaining a 40:60 male to female fly ratio. Despite this, few clutches were collected in each of the cages, with a low fertilisation rate that was characteristic of low – cage density:

♀ **G1 U6:1 gRNA-*w*** = 7 clutches, 4 fertilised, 0 transgenic.

♀ **G1 U6:2 gRNA-*w*** = 3 clutches, 1 fertilised, 0 transgenic.

♂ **G1 U6:1 gRNA-*w*** = 13 clutches, 10 fertilised, 0 transgenic.

♂ **G1 U6:2 gRNA-*w*** = 10 clutches, 3 fertilised, 0 transgenic.

As with all Opie2 transgenics, fertilised G2 clutches were each reared independently for 4-5 days until the marker visualisation was expected. However, none of the G2 clutches collected from either the U6:1 gRNA-*w* or U6:2 gRNA-*w* G1 crosses contained any positive transgenics. While the failure to recover G2 from the cross can be convincingly explained

as resulting from low cage density suppressing BSF breeding behaviour, as often encountered with other BSF strains, the survival of G1 neonates to adulthood with these constructs was notably lower than observed with the previous lines. Additionally, the G1 cross performed for PUb GFP / DsRed utilised a similarly low number of G1 females (n = 7), yet produced 4 transgenic-containing G2 clutches when crossed under the same cage supplementation conditions. Despite the low number recovered, the G1 neonate to adult survival rate for PUb GFP / DsRed was 88.2% compared to 31.5% and 51.2% as seen with U6:1 gRNA-*w* and U6:2 gRNA-*w* respectively, with both being significantly different to each of the 5 lines examined previously ($X^2 = 3.98e^{+1} - 5.60$, $p = 2.69e^{-10} – 6.05e^{-3}$). This could evidence a link between the fitness of G1 during development and the success of transgenic G2 production. However, due to the numbers involved and confounding variables in the rearing process, it would be difficult to claim that the specific transgene content of the two gRNA G1 pools was solely responsible for the lower fitness recorded without replicating more G1 from repeated rounds of injection.

However, the likely fitness cost associated with the *Act5C*-Cas9 transgene on survival of G1 transgenics to adulthood was much more evident. 4 out of 8 clutches produced in the G0 cross contained transgenic G1, with over 250 GFP positive larva collected and transferred to an agar feed vial. Growth proceeded as expected for around 3 days, after which the *Act5C*-Cas9 G1 larva appeared clearly smaller than a comparable number of T4 (wild-type) larva transferred on the same day to a separate vial. All transgenics appeared to then cease growing around 5 days post hatching, with a range of lengths between 2 – 4 mm placing this at around the 3rd larval instar (transgenics were not imaged).

The larvae were more motile that expected, and displayed the similar behaviour of crawling up the vial wall and away from the feed as seen with prepupa when they cease feeding and search for a dry environment to pupate. Larvae were transferred to separate vials to rule out any issue with the feedstock, but exhibited the same behaviour. Newly hatched, non-transgenic T4 larvae were added to the same vials as the *Act5C*-Cas9 G1s and showed no feeding deficiency, continuing to grow and overtaking the older G1s in size in a matter of days. G1s showing food avoidant behaviour were seen settling on the walls of the vial far away from feed and began to darken, suggesting precocious pupation.

None of the over 250 *Act5C*-Cas9 G1 developed further, with efforts to refeed them ceasing around 21 days after initial hatching, by which point all larvae were immobile and appeared as small, flat and pupa-like. As such, no G1s could be crossed to establish *Act5C*-Cas9 as a transgenic stock. Although striking, the fitness cost associated with high Cas9 expression

has been described in many other species, with the nuclease proving toxic and high burden when coupled with strong constitutive promoters[385,542–544]. The BSF-endogenous *Act5C* has previously been shown to fulfil this criteria in an earlier study where *Act5C*-driven GFP expression had a clear global tissue distribution that increased in intensity throughout the larval stage[422]. Although the *D. melanogaster Act5C* promoter variant employed for this study has not been specifically described in BSF, the high conservation between the orthologous genes and the assumed stability of expression shared by both species indicates that Cas9 was expressed under a similar profile as found in the study with GFP.

The *Ac5C*-Cas9 G1s collected were distributed evenly across the 4 transgenic-containing clutches obtained, suggesting that the toxic phenotype was more likely related to the specific transgene rather than the site of integration, although this cannot be ruled out as the integration sites of the perished G1 were not later determined. An analogous line, *Act5C*-MAD7, showed no obvious fitness cost at the larval stage in either the G1 or later generations following line establishment, the only difference between the two constructs being the specific nucleases driven by *Act5C*, leading to the conclusion that high Cas9 expression caused significant enough stress or metabolic burden for the larvae to undergo precocious pupation.

5.2.2 – Analysis of Transgene Inheritence Bias

After being unable to establish transgenic lines encoding the U6:1 gRNA-*w*, U6:2 gRNA-*w* or *Act5C*-Cas9 constructs, 4 nuclease-encoding and 3 gRNA-encoding BSF lines were successfully maintained for later CRISPR cross experiments. At the G2 stage, isolines for each construct were reared to adulthood, where general estimations of transgenic size, number and sex ratio compared to non-transgenic siblings were made. Initially, a minimum of 3 separate G2 clutches were reared for each line in independent feed containers, with each representing a possible different integration site to be later determined. However, the rapid scaling of space, resources and maintenance time for rearing one egg pot to the pupa stage, followed by the small window for adult screening and cross setup made maintaining multiple independent isolines for each construct impractical for the limited insectary space. When no obvious differences were found in terms of fitness between isolines for the same construct, one isoline was then selected to represent each transgenic construct going forward. From the G3 stage, the isolines were considered stabilised and expanded enough for characterisation and use in CRISPR cross experiments.

During standard maintenance, as well as after performing the first of the CRISPR crosses, one line had a notably biased inheritance of the transgenic component when screened at the adult stage. The *Nanos*-Cas9 line appeared to produce substantially lower ratios of transgenic to non-transgenic flies than the expected 50:50 ratio following heterozygote cross to the T4 *wild-type*. As neonates and larva were difficult to manipulate for screening and impossible to separate by sex, screening for inheritance bias was performed at the adult stage. Because of this, it remains undetermined whether the bias in inheritance occurred at the stage of fertilisation, egg development or during the course of larval development, assuming a transgene-related fitness cost was associated with the low observed proportion of viable transgenic flies.

To quantify potential biases in how each transgenic construct was inherited, a backcross to the T4 WT line was performed for each of the 4 nuclease and 3 RNA-expressing lines. In each case, 85 heterozygous transgenic G2 females were crossed to 80 - 100 non-transgenic T4 males, with the progeny collected and screened at the G3 stage to quantify inheritance patterns. Due to limitations on space, 3 fertilised clutches were collected and pooled in a single container for rearing to adulthood from each cross. This meant that the individuals screened were an average of three clutches and that inter-clutch variability in inheritance could not be assessed. A power calculation was performed to determine the number of individuals to screen per line for detecting a 10% deviation from the expected transgene inheritance of 50% (n = 387, power = 0.8, significance level = 0.05). The screening data and statistics used to calculate inheritance rates for each line are in presented in the appendix (**Supplementary Figure 3**).

At G3, *Nanos*-Cas9 had a significantly lower transgene inheritance than expected Mendelian rate at 35.1% when considering both sexes (G-test goodness of fit – G = 42.677, p = 6.5e-11, n = 473) (**Figure 38**). *Nanos*-Cas9 inheritance appeared not to be influenced by sex, with males showing 35.0% inheritance (G = 22.272, p = $2.3e^{-06}$ n = 243) and females 35.2% (G = 20.408, p = $6.2e^{-06}$, n = 230). The other transgenic lines displayed paired inheritance rates of between 45.0% and 53.7%, with only *Act5C*-MAD7 and U6:3 crRNA-GRG x6 appearing significantly lower than expected (*Act*5C-MAD7 = 45.2%, G = 7.304, p = 6.9e-3, n = 792) (U6:3 crRNA-GRG x6 = 44.98%, G = 5.429, p = 2.0e-2, n = 538). As inheritance for these lines deviated from the expected Mendelian rate by less than 10%, they were not considered at risk for continued line maintenance.

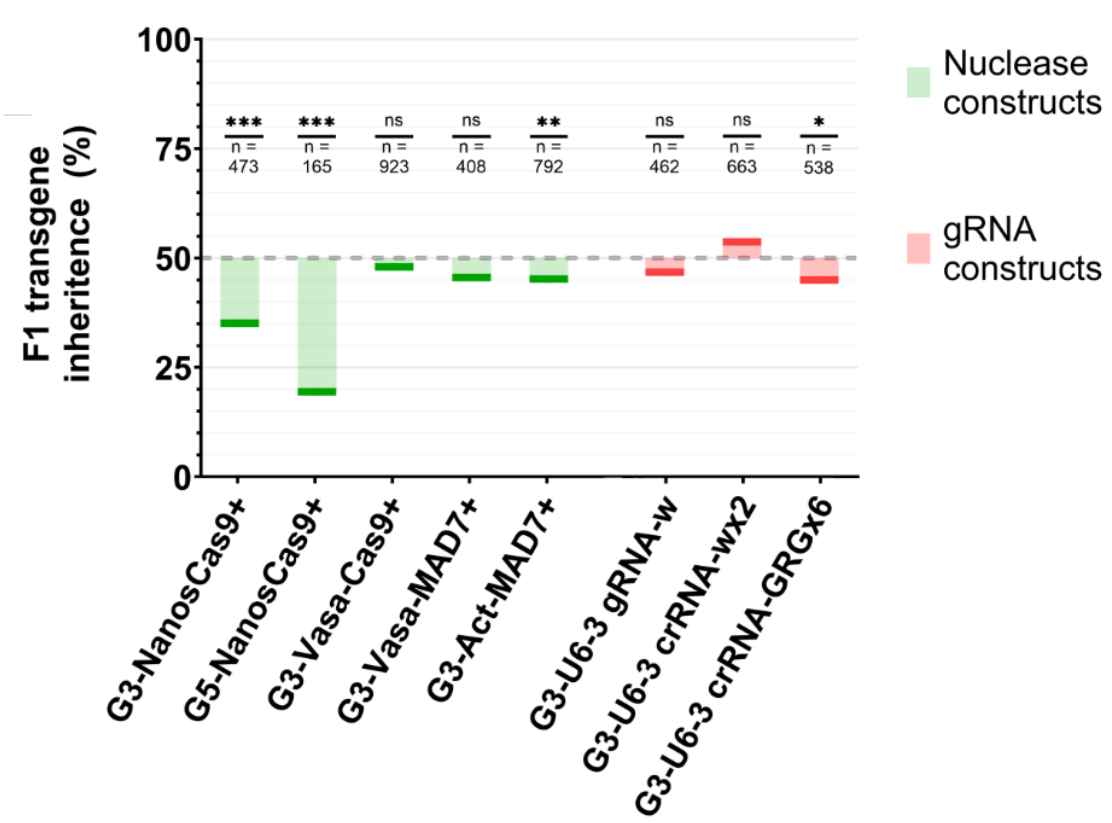


Figure 38: Transgene inheritance bias is only apparent in the *Nanos*-Cas9 transgenic line. Inheritance bias was defined as significant deviation from the 50% expected inheritance rate. Statistics were based on the G-Test of Goodness of Fit test, with screening data and statistics test results in appendix. Although significant deviation was found for *Act5C*-MAD7 and U6:3 crRNA-GRGx6, only *Nanos*-Cas9 deviated below 45% inheritance, with significant implications to fitness and the ability to maintain the line. ns, *, ** and *** are equal to p >0.05, <0.05, <0.01 and <0.001 respectively.

The low recovery of *Nanos*-Cas9 transgenics, however, suggested that the construct posed a non-trivial fitness cost. Using the same cross conditions, inheritance was calculated again at G5, by which point all other lines had continued to show no inheritance bias while the *Nanos*-Cas9 line produced notably smaller clutches with a low proportion of transgenic eggs. Transgene inheritance was calculated at 19.4% at G5, with only 32 adult transgenics recovered from 2 pooled clutches totalling 165 individuals (G = 65.463, p = 5.9e-16). The surviving transgenics had an obvious deficit in fitness, with adults either failing to fully eclose from pupal cases, lacking wings or having deformed genitalia in male flies (**Supplementary Figure 4A**). With only 22 male and 10 female G5 *Nanos*-Cas9 adults recovered, the cross for G6 larva produced no viable embryos and the line could not be maintained further.

With only the *Nanos*-Cas9 line displaying a deteriorating inheritance pattern and viability across multiple generations in this study, several possibilities for the accumulating fitness cost were examined. One possibility considered related to direct toxicity associated with

transgene content, specifically the *Nanos*-regulated expression of Cas9. The potential for Cas9-related toxicity in this line could be supported by earlier observations made with the *Act5C*-Cas9 line produced in this study, from which it was hypothesised that high and constitutive expression of Cas9 may have supressed larval growth sufficiently enough to collapse the line at G1. However, the *Nanos* promoter element utilised here was selected due to the conserved pattern of germline-restricted expression from the endogenous gene encountered in other species, with *Nanos*-Cas9 transgenic constructs being frequently utilised in other insect models with few reports on associated fitness costs[385,387,477]. Even if the *Nanos* regulatory elements isolated in this study permitted some leaky expression in non-germline tissues, it was deemed unlikely that Cas9 levels would accumulate enough to cause the type of toxicity as seen in the *Act5C*-Cas9 line. In support of this, the *Vasa*-Cas9 line was not associated with any observed fitness cost across multiple generations, despite these transgenics sharing a high degree of overlap in the level, timing and tissue-specific context of Cas9 expression with the *Nanos* construct.

Direct toxicity relating to inappropriate expression of Cas9 also does not explain the progressive deterioration of the *Nanos*-Cas9 line over multiple generations. Another explanation could be due to indirect toxicity relating to persistent Cas9 expression, in which a low level of promiscuous Cas9 activity causes an accumulation of off-target deleterious mutations throughout the genome. In this scenario, lower fitness may only become apparent after several generations of compounding mutagenesis. However, off-target activity of Cas9 is understood to still be largely dependant on the presence of a gRNA, with the nuclease protein being essentially catalytically inactive outside of the RNP complex[473]. With no discernible gRNA sequence being present in the BSF genome or the *Nanos*-Cas9 construct, it was deemed unlikely that the presence of Cas9 alone produces significant mutagenic activity in the line.

5.2.3 – *piggyBac* Integration Site Mapping and Assessing the Impact of Transgene Position on Fitness

Outside of Cas9-related effects, the most likely explanation for the low observed fitness in the *Nanos*-Cas9 line was thought to be related to the specific site of transgene integration. With no clear fitness deficit in any of the three isolines being evident at the G2 stage, only one isoline was chosen for maintenance of *Nanos*-Cas9, unfortunately precluding any later comparison of fitness between isolines featuring the same construct at different integration

sites. The *piggyBac* constructs were expected to have integrated randomly at TTAA sites distributed across the BSF genome, although the *piggyBac* transposon system does show an apparent bias for integration in transcriptional units[419]. Depending on the region, the expression level of the integrated transgenes could be influenced by surrounding genomic enhancer or suppressor elements. This could lead to variability in the efficiency of transgene production from otherwise identical constructs depending on their specific genomic context and even orientation on the chromosome, with these influencing factors being collectively known as position effects[545].

The *gypsy* insulator element, originally from *D. melanogaster*, was added into the constructs to help reducing the influence of the local genomic environment on the behaviour of the transgenes, effectively flattening their expression level to be less affected by adjacent endogenous genetic features and allowing the associated transgene regulatory elements to be the main influencer over gene expression. As the transgenics in this study showed clear marker expression, it suggested that each construct had integrated in loci that were not significantly restrictive to transcription. However, the possibility remained that the *Nanos*-Cas9 construct may have integrated in a particularly highly-transcribed region of the BSF genome, and that strong enhancer elements adjacent to the transgenes may have an influence on Cas9 expression independent of the restrictive *Nanos* elements.

However, an accumulating loss of fitness over multiple generations more strongly points to the loss or gain of function in native genes resulting from a *piggyBac* insertion at an inappropriate locus. With a preference for integration in highly transcribed regions, there is an inherent risk that the transgenic cargo sequence may integrate within an endogenous gene in a manner that significantly impacts the level or context of gene expression, or interrupt the reading frame all together. This tendency of *piggyBac* constructs to integrate close to active genes has even been exploited as a mutagenic screening technique in *D. melanogaster*, from which several essential genes have been identified by determining where an insertion event may have disrupted native gene functions in flies with compromised fitness[546,547].

To examine the possibility of positional effects on the fitness of the *Nanos*-Cas9 line, the location of each *piggyBac* insertion site in the BSF genome was investigated. gDNA samples from the one isoline for each construct was extracted from pools of 10 female transgenic flies. The full promoter, nuclease and terminator components for each nuclease line were then amplified by PCR and sequenced, with no obvious mutations being detected in the transgenic flies. For the guide RNA lines, the U6 promoter and RNA expression cassette

were validated in the same manner. Maker genes were not sequenced, as they clearly remained functional regardless of potential sequence deviations.

After sequencing the internal construct components, several attempts were made to identify their specific genomic loci by both inverse PCR (iPCR) and a modified iPCR technique known as splinkerette PCR[548]. Both techniques have been successfully employed for resolving *piggyBac*-mediated integrations in BSF[422,441]. The success of these techniques in the current study was mixed, however, despite optimisations to the restriction, ligation and PCR reaction conditions, varying gDNA input concentration, and the use of different primer sets. Of the 8 transgenic lines maintained, only *Nanos*-Cas9, *Vasa*-Cas9, *Vasa*-MAD7 and U6:3 gRNA-*w* were successfully mapped to a genetic locus (**Figure 39**).

Despite only representing around 1.5% of the BSF genome, 2 of the 4 transgenes mapped onto chromosome 7, which is understood to be the sex-linked or functional X chromosome of the species[429]. Aside from a detailed sequence assembly, little has been described regarding the function or accessibility of the X chromosome in BSF, with different insect models showing significant divergence in both the structure of their sex chromosomes and their roles in sex-determination. Although possibly only arising due to chance, the two X integrations mapping to within 3 mb of each other in a total genome size of 1 gigabase could indicate that the region has a more accessible chromatin structure or was more transcriptionally active than others at the time of *piggyBac* transposition[418]. Both *Vasa*-Cas9 and U6:3 gRNA-*w* constructs integrated inside genes near the beginning of the transcribed region, on chromosome 1 and 6 respectively, which likewise aligns with the apparent bias of *piggyBac* integration towards transcriptional units.

Of the three genome-mapped constructs associated with no obvious fitness effects in transgenic BSF, only *Vasa*-MAD7 had integrated outside of an active gene. On chromosome 7, this construct was located 12.7 kb upstream of LOC119660950, a gene corresponding to a homolog of *unc-13*. A regulator of vesicle maturation and fusion at the synapse, this gene was assumed to be sufficiently distant enough from the site of *piggyBac* insertion to have a low risk of significant disruption as a result[549]. In contrast, the *Vasa*-Cas9 construct was determined to have integrated within the first intron of the 5'UTR of LOC119652388, a homolog of the adenosine deaminase-like protein gene *ada*. Although the 5' and 3' splice sites defining the intron boundaries were not disrupted, the presence of the 13 kb transgenic construct at this site may well have an effect on the proper splicing, and therefore activity of the gene.

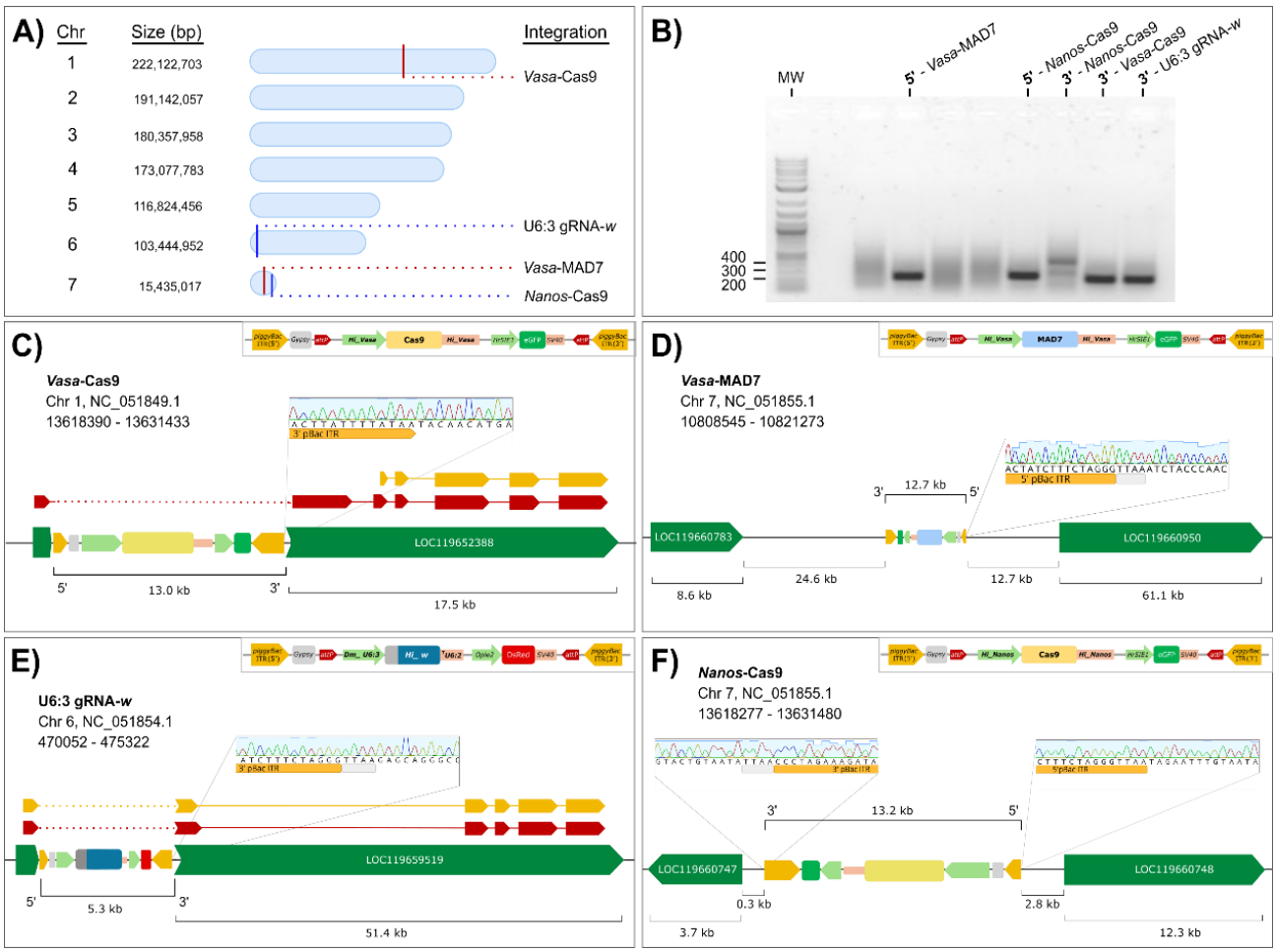


Figure 39: The sites of genomic integration of four transgenic *piggyBac* constructs determined by inverse PCR. A) Relative chromosomal position of the successfully verified transgenic insertions. **B)** PCR bands from inverse PCR spanning the junction between genomic sequence and the 5' or 3' *piggyBac*-ITR. Bands were extracted for Sanger sequencing and aligned to the BSF genome to determine the exact site and orientation of the construct. Unlabelled lanes correspond to samples from other lines where amplification was unsuccessful, with no sequencing therefore performed. **C-F)** Each sequenced insertion site suggested *piggyBac* insertion either within or closely adjacent to annotated protein-coding genes, as identified by sequence chromatograms spanning the *piggyBac* ITR and the BSF genome. While *Vasa*-MAD7 and *Nanos-Cas9* integrated between two discrete genes, *Vasa*-Cas9 was detected in an intron in the 5'UTR sequence, and U6:3 gRNA-*w* integrated within the first protein coding exon of LOC119659519, likely resulting in a translational frameshift for that gene. The sequence alignments presented here encompassed >100 bp of downstream genomic sequence for each insertion site, with only the transgenic to genomic DNA junction shown here for illustrative purposes.

Additionally, the U6:3 gRNA-*w* construct was found to have integrated within the protein coding sequence of LOC119659519, resulting in a translational frameshift in the gene for zinc finger protein *schnurri*, a transcription factor with central roles in cell differentiation and development[550]. *ada* and *schnurri* are both considered essential haplosufficient genes in *D. melanogaster,* meaning heterozygotes retaining at least one functional allele would still be viable[550,551]. Assuming that the inserted *Vasa*-Cas9 and U6:3 gRNA-*w* transgenes created effectively null alleles for each gene respectively, both appear to be similarly haplosufficient in BSF. Crosses to produce homozygotes were not made for either of these lines, so the fitness of individuals with two disrupted copies of *ada* or *schnurri* was not assessed further.

Despite integrating at a locus outside of annotated gene sequence, the *Nanos*-Cas9 transgene insertion was suspected to be highly likely to significantly impact the function of two adjacent genes. Found on chromosome 7 between the divergently-orientated LOC119660747 and LOC119660748, the *piggyBac* construct was located only 0.3 kb and 2.8 kb upstream of the transcriptional start sites of each gene respectively. For both genes, the close proximity of the insertion was deemed likely to cause a disruption in the endogenous upstream promoter or enhancer elements, attenuating their expression or perhaps ablating it completely.

The likely modification to the expressional level of either of these genes may help explain the gradual decrease in the fitness of the *Nanos*-Cas9 line over generations. The LOC119660748 gene corresponds to *Ssdp*, or Sequence-specific single-stranded DNA-binding protein. An evolutionarily conserved transcription factor with multiple roles in neuronal development, tissue patterning and the Wnt signalling pathway, the gene is considered to be highly dosage-sensitive in *D. melanogaster*[552]. Neuronal overexpression in flies was associated with a detrimental increase in brain volume, oxidative stress and abnormal behaviour, whereas null mutants encountered developmental issues and lethality at the larval stage[552,553]. Located 2.8 kb upstream of the BSF *Ssdp* gene, the integrated *Nanos*-Cas9 construct was considered to have likely avoided disruption of core proximal promoter elements, but may still impact expression level or context through disruption of long-range interactions with distal enhancer elements.

Conversely, the *Nanos*-Cas9 integration being placed only 0.3 kb upstream of LOC119660747 was deemed highly likely to significantly curtail native expression of the endogenous gene. LOC119660747 encodes the glucose-6-phosphate 1-dehydrogenase (G6PD) *Zw*, an enzyme in the pentose phosphate pathway and the catalyst in the primary mechanism for the regeneration of NADPH from NADP[554]. A cofactor in several metabolic

processes, NAPDH is critical for the production of reduced glutathione by glutathione reductase, the key antioxidant for the neutralisation of reactive oxygen species (ROS) within cells[555]. A disruption in upstream G6PD function therefore results in an inability to maintain high levels of reduced glutathione in cells, causing them to be highly susceptible to damage under conditions of oxidative stress.

G6PD genes are understood to have critical roles in facilitating redox homeostasis, metabolic adaptation and immune response in insects[556]. Selection for longer lifespan in *D. melanogaster* found that longer-lived flies posses more active allozymes of *Zw,* and studies on the overexpression of *Zw* using Gal4-UAS transgene pairing found that increased expression resulted in significantly higher levels of reduced glutathione, resistance to hypoxia and paraquat-induced redox conditions, and an extended lifespan of between 16-38%[557,558]. The majority of G6PD loss-of-function research, however, has been conducted on orthologous mammalian genes in the context of metabolic disease. In humans, G6PD deficiency is an X-linked recessive disorder and the most common form of enzyme-associated anaemia worldwide[559]. In the highly oxidative environment of red blood cells, the inability to regenerate reduced glutathione leads to the formation of protein aggregates and damage to cell membranes, resulting in haemolytic anaemia. While G6PD is not strictly an essential gene in a state of low-stress, mutants are far more susceptible to disease when exposed to conditions of oxidative stress, which can arise from many environmental or dietary-related triggers, as well as infection[560].

Interestingly, several features of the observed low fitness profile in the *Nanos*-Cas9 line appeared to correlate with a G6PD deficiency phenotype. As G6PD deficiency has an incomplete penetrance depending on the presence of ROS, the impact of a loss-of-function mutation in *Zw* may have been low in optimal feed and growth conditions, and impaired fitness only becoming apparent when exposed to triggers of redox stress. This could help explain why the earliest generations of *Nanos*-Cas9 flies appeared viable, while fitness appeared to deteriorate over multiple generations; this may be attributable to repeated periods of oxidative stress causing an accumulation of cellular damage and a reduced ability to adapt to later ROS exposure.

Additionally, the fitness of the *Nanos*-Cas9 line appears to resemble the X-linked recessive phenotype relating to G6PD deficiency. As with the human variant of G6PD, *Zw* is found on the X chromosome in BSF, meaning that males are hemizygous for a single copy of the gene while females posses two alleles. In humans, females that are heterozygous for a G6PD mutant allele are typically asymptomatic carriers of the disorder, with one functional copy

present on either X chromosome proving sufficient for viability[560]. Males, in contrast, inherit only one allele on their maternally-derived X chromosome, with G6PD deficiency-related diseases therefore being much more prevalent and severe in males compared with females[559]. In the *Nanos*-Cas9 line, male flies screened at later generations appeared to have a more significant diseased phenotype than female transgenics when screened as adult flies, with deformities in the abdomen and genitalia being significantly more apparent in this sex (**Supplementary Fig 4**).

While no apparent sex bias regarding *Nanos*-Cas9 inheritance was detected in the earlier inheritance test discussed in section 5.2.2, a later transgenic cross of *Nanos*-Cas9 females to U6:3 gRNA-*w* males described in section 5.2.4 did result in a significantly lower number of male progeny (4.1%) encoding *Nanos*-Cas9 when screened as adults than females (25.%), with the proportion of males carrying both transgenes (0.6%) being drastically lower than female transheterozygotes (10.6%). Assuming no specific inheritance bias at fertilisation, the lower proportion of hemizygous *Nanos*-Cas9 males could indicate a much higher susceptibility to stressors like ROS than their heterozygous female counterparts, with these compromised males being less likely to survive through the embryonic, larval and pupal stages to the point of screening as adult flies. The lower proportion of recovered male transheterozygotes could also be attributed to the metabolic burden of two transgenic constructs causing excess ROS generation, with G6PD deficient flies being unequipped to deal with this effectively. As a highly oxidative environment, the early embryo is seen as critical stage at significant risk of ROS-associated lethality in individuals with G6PD deficiency, with hemizygous mutant male embryos found to be non-viable even in low ROS conditions in mice and *Caenorhabditis elegans* models[561,562]. With morphological sexing only possible at the adult stage in BSF, the low observed proportion of *Nanos*-Cas9 carrying males following the transgenic cross described above may likely be attributed to ROS-induced lethality at earlier stages that was not captured during the screening experiments.

With the evidence presented, it was suspected that the low fitness of the *Nanos*-Cas9 line may well have been due in part to disruption of native BSF genes involved in essential cell maintenance and resilience to stress. With each characterised integration site having integrated near to, and likely disrupting, native genes in the BSF genome, the influence of specific transgenic loci following *piggyBac* integration is clearly a consideration for the efficiency, fitness and welfare of future BSF transgenic lines. Although more than sufficient for the type of transgenesis experiments performed in this study, these findings emphasise

the need for the development of site-specific integration techniques for more coherent analysis of transgenic material introduced into the BSF genome.

5.2.4 – Cas9 x gRNA-*white* CRISPR Cross Resulted in No Detectable Mutagenesis

With a partial characterisation of each line performed, three crosses of transgenic lines were performed to combine the CRISPR components and enact gene editing. At G3, before the apparent fitness cost of *Nanos*-Cas9 was determined, a parental (P0) cross was initiated with 80 heterozygote females from the nuclease line and 80 heterozygote G2 U6:3 gRNA-*w* males. From the resulting F1 progeny, three large egg clutches were collected and pooled for rearing to the adult stage for screening. Although Cas9 expression was expected to be restricted to germline tissues, the eyes of each F1 individual were also examined for evidence of a mosaic *w* phenotype, which could indicate either somatic expression of the nuclease or significant maternal Cas9 deposition in the early embryo. Following the rationale of non-biased Mendelian inheritance, the four possible inheritance outcomes of no transgene, nuclease-transgene only, RNA-transgene only and both transgenes should each occur in an equal ratio (1:1:1:1).

No mosaics for the *w* phenotype were identified in the F1 progeny of *Nanos*-Cas9 x U6:3 gRNA-*w* cross, which also displayed a similar bias against the nuclease construct as encountered in the transgene inheritance analysis (**Figure 40**). Interestingly, the inheritance of *Nanos*-Cas9 was significantly biased towards female F1 than males (Fishers exact test: odds ratio - 0.12, p = 2.2e-16), although both sexes inherited at a lower rate than encountered in the *Nanos*-Cas9 backcross to T4 *wild-type*. 25.7% of females inherited *Nanos*-Cas9, around half the expected proportion (G = 133.69, p = <2.2e-16), while males only saw 4.1% inheritance (G = 333.56, p = <2.2e-16). U6:3 gRNA-*w* inheritance, however, conformed to the expected rates with no significant difference found between sexes (odds ratio = 1.007, p = 1.0). The sub-Mendelian inheritance of both constructs in tandem could be explained by the interaction of the specific inheritance rates associated with each transgene:

♀ **25.7%** (*Nanos*-Cas9) x **48.7%** (U6:3 gRNA-*w*) = **12.54%**

♂ **4.1%** (*Nanos*-Cas9) x **48.9%** (U6:3 gRNA-*w*) = **1.99%**

With these rates, dual inheritance was not significantly different than expected for either sex, although the low count of male transheterozygotes (n=2) should be noted (Females: G = 3.1084, p = 0.07789) (Males: binomial test, probability = 0.00626, p= 0.1036). This indicated that there was no apparent increased fitness cost seen in F1 inheriting both transgenes, when CRISPR activity is presumed to be present.

The P0 cross was originally intended to be followed up by crossing male and female transheterozygotes to enrich the population containing both CRISPR components. Due to Cas9 germline restriction, edits in the *w* gene were expected in the gametes of F1 transheterozygotes, which would result in a fully penetrant mutant phenotype when individuals of each sex were crossed. However, this cross could not be performed due to a lack of transheterozygote males. The 55 transheterozygote females were instead crossed to an equal number of F1 U6:3 gRNA-*w* + males. Any *w* mutants encountered in the F2 progeny would be understood to have resulted from maternal Cas9 deposition into the embryo and editing the paternal *w* allele. Interestingly, limited breeding activity was noted in the F1 cage despite being at a sufficient mating density, and only one small clutch was recovered that contained hatching eggs.

From this, only 32 F2 flies were recovered, with 6 transheterozygote females, no transheterozygote males and no indication of *w* mutation of mosaic phenotype. Although the cause for the poor F1 cross output and limited fly recovery remained undetermined, several other lines crossed at the same time were found to suffer from a similar breeding depression, including the $w^{-/-}$ line which regrettably was also lost at the point of G3 cross. This was later suspected to be caused from a wiring issue that caused the UV breeding lights above the cages to occasionally disconnect or flicker, the occurrence of flickering having been previously associated with substantial disruptions to successful mating in BSF[519,563]. Unfortunately, the *Nanos*-Cas9 line had expired before the P0 *Nos*-Cas9 x U6:3 gRNA-*w* cross could be repeated.

Despite the lack of an apparent mutant phenotype, 15 transheterozygote F1 females were sampled following the F1 cross and gDNA extracted from individual ovaries for analysis of indels at the *w* locus. A 287 bp amplicon was produced flanking the target site, which was submitted for Sanger sequencing. As no overlapping peaks were evident in the sequence trace, the files were submitted to the online tool DECODR for Sanger deconvolution and indel identification. Each of the ovarian amplicon samples shared 100% identity with the functional *wild-type w* allele, with no evidence of indel formation detected in the F1 female progeny. The same analysis was repeated with gDNA extracts from the 2 F1

transheterozygote males (extracted from testes) and the 6 F2 transheterozygote females, similarly with no indels detected after Sanger deconvolution (data not shown).

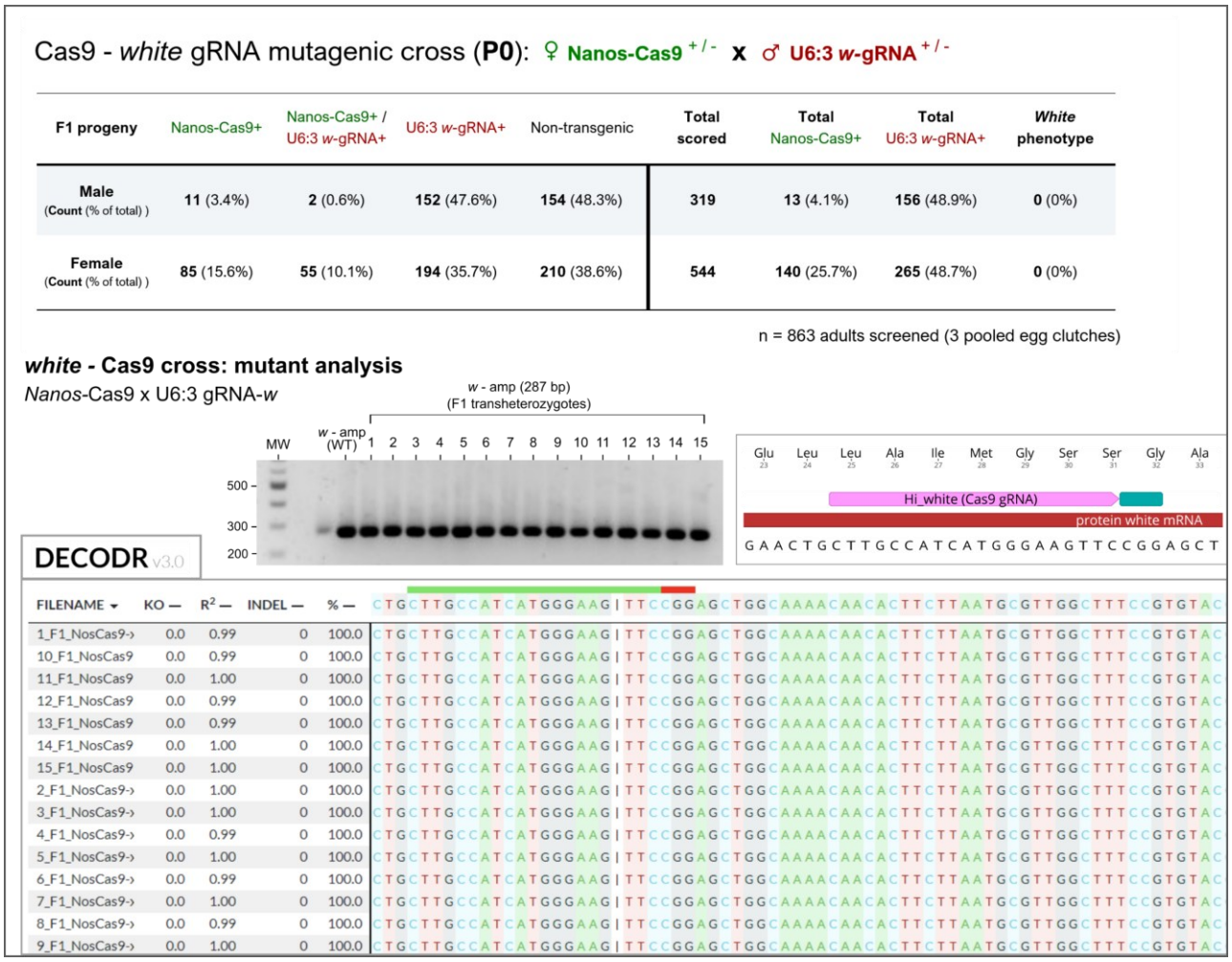


Figure 40: No *white* target site mutation was detected in the F1 transheterozygote progeny of the *Nanos*-Cas9 x U6:3 gRNA *w* cross. The screened F1 progeny of the *Nanos*-Cas9 x U6:3 gRNA *w* cross featured a *Nanos*-Cas9 inheritance rate significantly lower than the expected Mendelian rate of 50%, with inheritance in males (4.1%) being significantly lower than in females (25.7%), while the U6:3 gRNA *w* construct was inherited at the expected frequency in both sexes. A 287 bp amplicon surrounding the *w* target site was amplified from gDNA extracted from 15 female F1 transheterozygote progeny and submitted for Sanger sequencing. Sequence trace alignment and DECODR sequence deconvolution of the target site indicated no clear deviation from the *wild-type white* sequence, suggesting minimal error-prone indel repair activity in the surviving F1 population.

5.2.5 – MAD7 Multiplexing: CRISPR Mutagenic Crosses and Amplicon Sequence Analysis

While the *Nanos*-Cas9 cross remained ineffective for generating mutants in *w*, the associated fitness cost of that transgene was understood to have affected the success of the P0 and subsequent F1 cross. As the *Act5C*-MAD7 line was the last to be established, the MAD7 crosses were set up using the *Vasa*-MAD7 line. While this promoter was expected to show a similar enrichment of germline expression as *Nanos*, there is evidence for higher *Vasa*-associated maternal deposition of Cas9 in *D. melanogaster, A. gambiae* and *C. capitata,* potentially increasing the occurrence of mosaicism in F1 transheterozygotes[385,477,564]. For the first MAD7 mutagenic cross, 90 G3 *Vasa*-MAD7 females were crossed to 90 G2 U6:3 crRNA-*w x2* males.

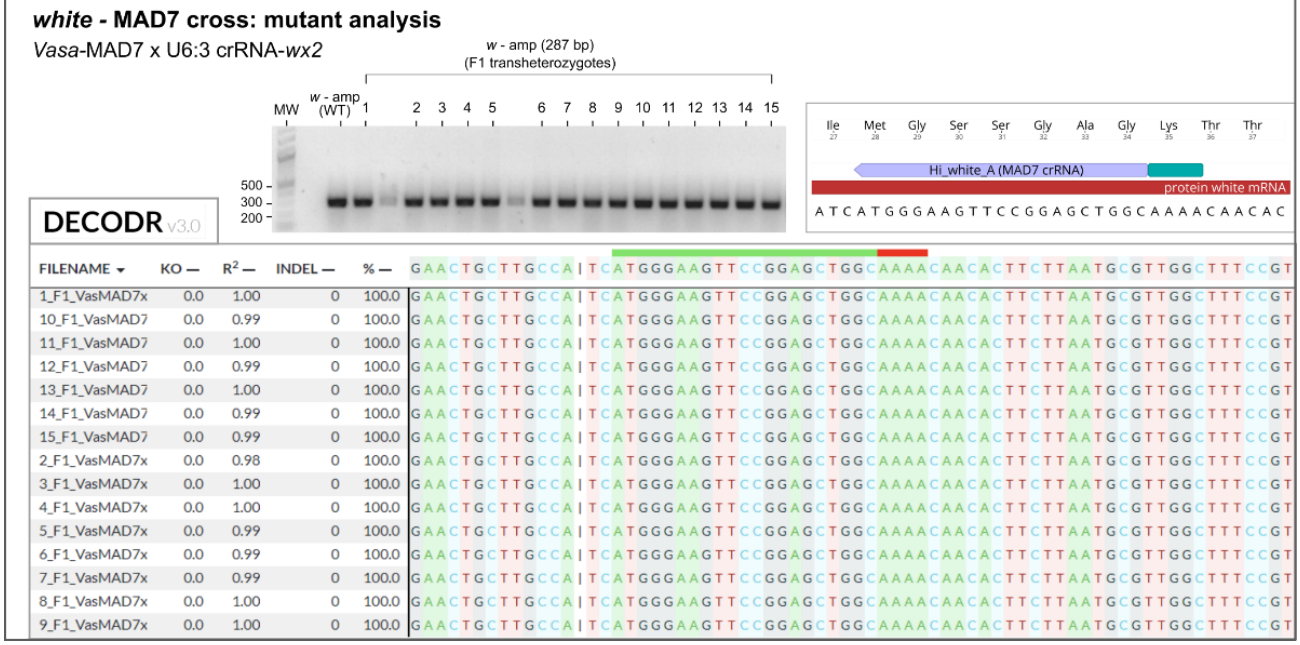
While no mosaicism was recorded in the F1 progeny, the inheritance of both nuclease and RNA constructs followed the expected frequencies in both sexes (**Figure 41**). 90 F1 transheterozygote females were crossed to 90 F1 transheterozygote males to produce F2 progeny at a similar frequency distribution of transgenes (data not shown). Despite a higher fitness, in both sexes, no *w* mutants or mosaics were detected amongst any F2 progeny. Considering that mutagenesis efficiency may be too low to detect in F2, 90 female and male transheterozygote F2 were similarly crossed to produce an F3 population, but again no *w* mutants or mosaics were observed.

w amplicon target production and indel analysis by Sanger trace deconvolution was performed as with the Cas9 cross, with germline tissues sampled from 15 females for the F1, F2 and F3 generations (F1 data shown). No male germlines were sampled in this case, assuming that the enrichment of transgenes coming from both sexes would prove sufficient for mutagenesis in females by the F3 stage. As with the Cas9 cross, no occurrence of indels were recorded. As only the MAD7 crRNA-A target was captured in the *w* amplicon, an amplicon corresponding to the crRNA-B target site was produced from an additional sample of 10 F3 females, similarly indicating no indel formation at the secondary target site (data not shown). The transgenes of one F3 transheterozygote were then fully sequenced to rule out the possibility of deleterious mutant accumulation in the MAD7 coding sequence or crRNA expression cassette. With no evidence of mutagenic activity apparent across three generations of transheterozygotes, the cross was discontinued with the assumption that the constructs were insufficient for generating detectable editing at the designated target sites.

MAD7 – *white* crRNA mutagenic cross (**P0**): ♀ **Vasa-MAD7** $^{+/-}$ **x** ♂ **U6:3 *w*-crRNA (x2)** $^{+/-}$

F1 progeny	Vasa-MAD7 +	Vasa-MAD7 + / U6:3 *w*-crRNA (x2) +	U6:3 *w*-crRNA (x2) +	Non-transgenic	Total scored	Total Vasa-MAD7 +	Total U6:3 *w*-crRNA (x2) +	*White* phenotype
Male (Count (% of total))	**167** (33.1%)	**96** (19.0%)	**130** (25.7%)	**112** (22.2%)	**505**	**263** (52.1%)	**226** (44.8%)	**0** (0%)
Female (Count (% of total))	**196** (34.3%)	**134** (23.5%)	**128** (22.4%)	**113** (19.8%)	**571**	**330** (57.8%)	**262** (45.9%)	**0** (0%)

n = 1076 adults screened (4 pooled egg clutches)



FILENAME	KO	R^2	INDEL	%	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
1_F1_VasMAD7x	0.0	1.00	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
10_F1_VasMAD7	0.0	0.99	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
11_F1_VasMAD7	0.0	1.00	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
12_F1_VasMAD7	0.0	0.99	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
13_F1_VasMAD7	0.0	1.00	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
14_F1_VasMAD7	0.0	0.99	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
15_F1_VasMAD7	0.0	0.99	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
2_F1_VasMAD7x	0.0	0.98	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
3_F1_VasMAD7x	0.0	1.00	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
4_F1_VasMAD7x	0.0	1.00	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
5_F1_VasMAD7x	0.0	0.99	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
6_F1_VasMAD7x	0.0	0.99	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
7_F1_VasMAD7x	0.0	0.99	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
8_F1_VasMAD7x	0.0	1.00	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT
9_F1_VasMAD7x	0.0	1.00	0	100.0	GAACTGCTTGCCA|TCATGGGAAGTTCCGGAGCTGGCAAAACAACACTTCTTAATGCGTTGGCTTTCCGT

Figure 41: No *white* target site mutation was detected in the F1 transheterozygote progeny of the *Vasa*-MAD7 x U6:3 crRNA *w* x2 cross. The screened F1 progeny of the *Vasa*-MAD7 x U6:3 crRNA *w* x2 cross displayed an evenly distributed pattern of transgenic marker inheritance, indicating no significant inheritance bias for either construct. The 287 bp amplicon surrounding the *w* target was amplified from gDNA extracted from 15 female F1 transheterozygote progeny and submitted for Sanger sequencing. Similar to the Cas9 cross, sequence trace alignment and DECODR sequence deconvolution of the target site indicated no clear deviation from the *wild-type white* sequence, suggesting minimal error-prone indel repair activity in the surviving F1 population.

In tandem, the MAD7 cross for simultaneous editing at 6 GRG target sites was performed with 90 G3 *Vasa*-MAD7 and 90 G2 U6:3 crRNA-GRG x6 males (**Figure 42**). As no evident marker was targeted for knockout with this RNA construct, mosaicism could not be easily identified in the F1 progeny, although a potential inheritance cost would be detectable in transheterozygotes if mutation at any GRG target resulted in a significant lack of fitness. However, inheritance of both *Vasa*-MAD7 and U6:3 crRNA-GRG x3 was detected at the expected frequencies in the F1, and later F2 and F3 transheterozygote crosses (**Supplementary Figure 4B**). As with the MAD7-*w* x2 cross, 15 female transheterozygotes

were sampled for indel analysis at F1, 2 and 3. Rather than sequence each of the 6 target sites in every individual, *Maf1* was examined in all 15 gDNA extractions for the 3 generations, with the remaining *ImpL2*, *CRBN*, *sirt6*, *TBC1d7* and *Piezo* target sites only sequenced in 5 extractions per generation (analysis not shown). Nevertheless, no indels were detected by trace deconvolution at any target site at each generation.

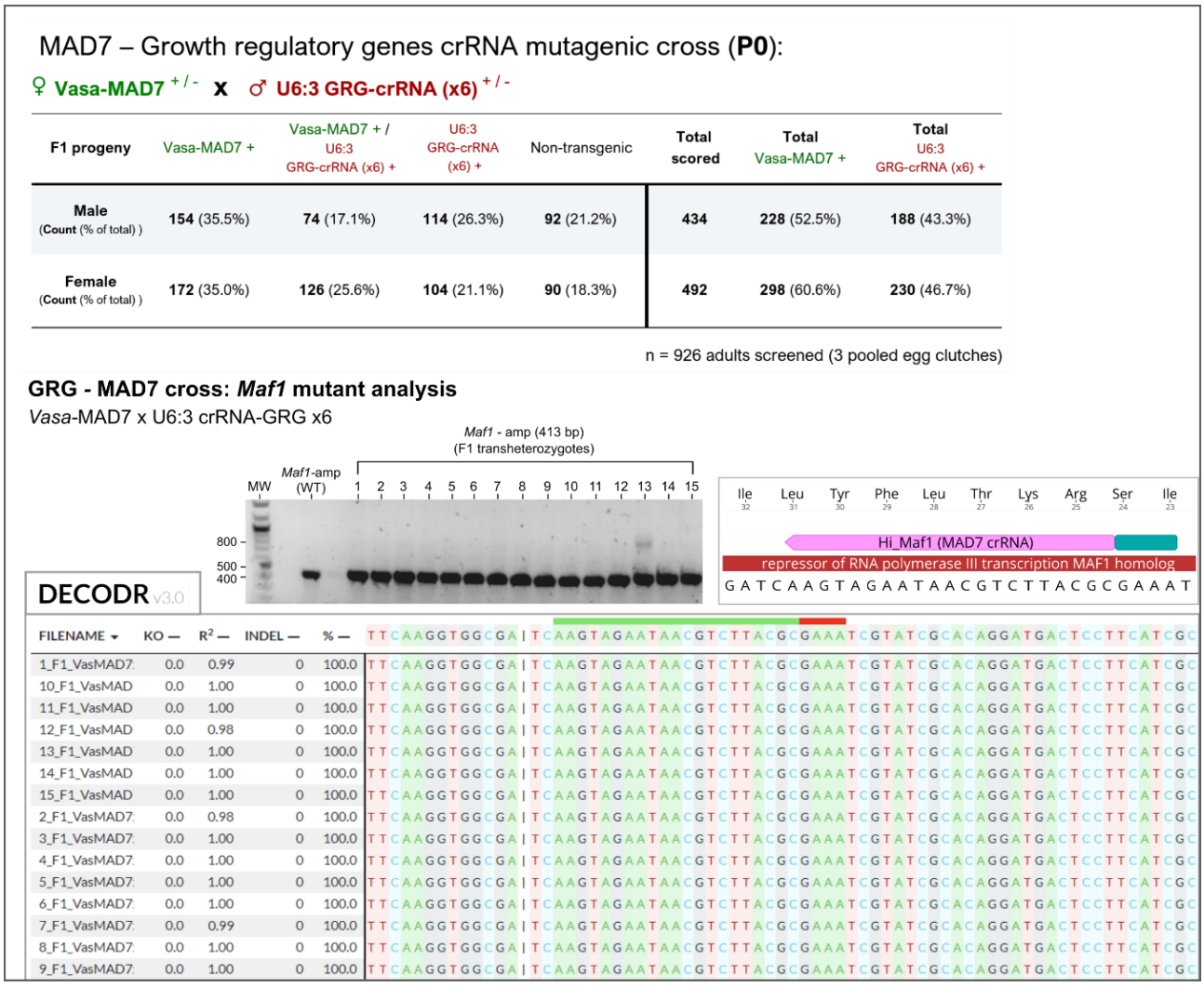


Figure 42: No *Maf1* target site mutations were detected in the F1 transheterozygote progeny of the *Vasa*-MAD7 x U6:3 crRNA GRG x6 cross. The screened F1 progeny of the *Vasa*-MAD7 x U6:3 crRNA GRG x6 cross displayed an evenly distributed pattern of transgenic marker inheritance, indicating no significant inheritance bias for either construct. A 415 bp amplicon surrounding the *Maf1* target was amplified from gDNA extracted from 15 female F1 transheterozygote progeny and submitted for Sanger sequencing. Sequence trace alignment and DECODR sequence deconvolution of the target site indicated no clear deviation from the *wild-type Maf1* sequence, suggesting minimal error-prone indel repair activity in the surviving F1 population.

5.2.6 – Confirming Nuclease Transcription and Translation in BSF Tissues

One of the most likely rationales behind the lack of endogenous CRISPR activity during the cross experiments was that the nuclease components were not effectively expressed. Despite utilising regulatory elements from BSF-endogenous genes, these constructs had not been previously validated in the species and may not lead to the intended level or patten of mRNA transcription. Additionally, these components may be expressed in some tissues, but insufficiently in the germline where they are required for efficient and heritable editing activity to occur. For a qualitative estimation of nuclease transcription in tissues where CRISPR activity would be required, reverse transcriptase PCR (RT-PCR) was performed (**Figure 43A**).

Total RNA was isolated from females of the T4 *wild-type, Nanos*-Cas9, *Vasa*-Cas9, *Vasa*-MAD7 and *Act5C*-MAD7 lines. Pools of ovaries, female carcass (with ovaries removed), and fertilised egg clutches sampled between 1-2 hours post laying were isolated to represent the expected distribution of nuclease expression. Following DNAse-I treatment, cDNA was synthesised for each sample and PCR performed with qPCR primers designed to quantify the levels of Cas9, MAD7 and BSF-endogenous *Actin5C* as a housekeeping gene. The band pattern indicates that *Actin5C* is stably expressed in all lines and each sampled tissue. Bands signifying Cas9 amplification were found in all 3 tissue types for the *Nanos* and *Vasa*-Cas9 lines, and absent in the rest. Although much fainter, MAD7 bands were similarly observed in all tissues of the *Vasa* and *Act5C*-MAD7 lines, despite being more apparent in the carcass than the germline or embryonic tissues. While a more thorough quantitative analysis could be performed for these transgenes using qPCR, this data is sufficient to rule out a complete absence of transcription of either nuclease in the expected tissues for the lines produced in this study. The original agarose gel images for the RT PCR experiments are displayed in the appendix (**Supplementary figure 5**).

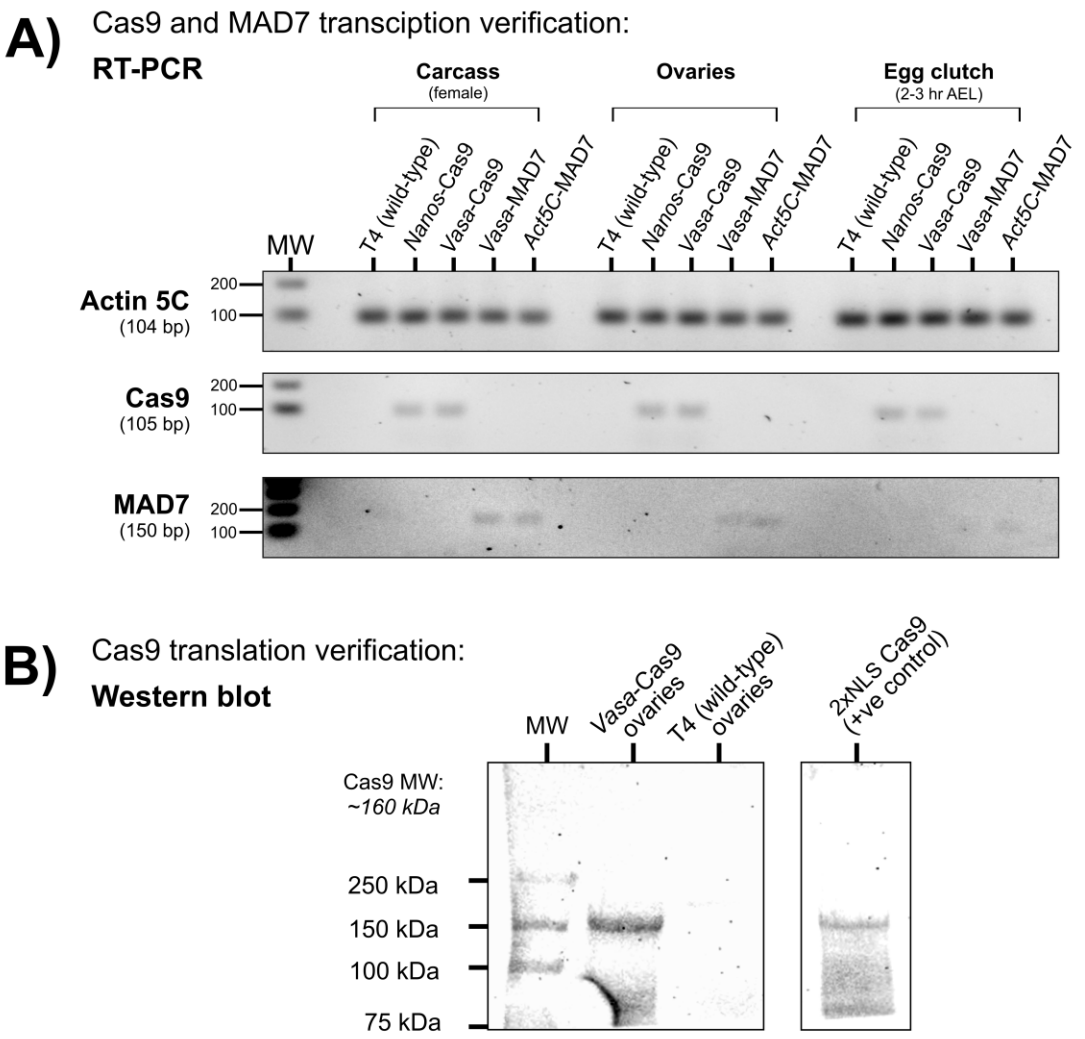


Figure 43: RT PCR and Western blot of WT and Cas9-transgenic line ovarian tissues, carcass and embryos. A) RT PCR for the Cas9 and MAD7 transcripts was performed on cDNAs produced from samples of ovary tissue, fly carcass tissue following ovary removal, and egg clutches taken from female flies of the T4 *wild-type* and *Nanos*-Cas9, *Vasa*-MAD7, *Vasa*-Cas9 and *Act5C*-MAD7 transgenic genotypes. An *Actin 5C* transcript was amplified as a positive control for ubiquitous and endogenous gene expression in all samples. Bands corresponding to the Cas9 PCR amplicon were identified in all tissue samples from the *Nanos*-Cas9 and *Vasa*-Cas9 lines, while MAD7 was identified in all tissues from *Vasa*-MAD7 and *Act5C*-MAD7 lines, and expression of either nuclease was not detected in T4 tissue samples. **B)** An immunoblot for Cas9 antigen was performed on ovary tissue protein extracts from the T4 and *Vasa*-Cas9 genotypes, performed by Dr. Moeez Khan (Imperial College London, August 2024). 10 µg of total protein was assayed for each tissue sample, with 1.0 µg of NLS-Cas9-NLS Nuclease (GenCrispr) loaded in an additional lane as a positive control. Protein bands at the expected molecular weight of 165 kDA for FLAG-NLS-Cas9-NLS were detected in lanes containing the *Vasa*-Cas9 ovarian tissue and Cas9 positive control samples, whereas no band was apparent in the T4 ovary tissue sample.

Although nuclease transcripts could be identified in these tissues, the post-transcriptional regulation associated with *Nanos* or *Vasa* transcripts does not guarantee effective protein translation. While no probe for MAD7 could be acquired, an attempt to detect the presence of Cas9 protein in BSF tissues was made *via* western blot (**Figure 43B**, experiments performed by Dr. Moeez Khan, Imperial College London, August 2024). Tissue samples from females of the T4 and heterozygote *Vasa*-Cas9 lines were prepared, with protein from mature ovaries and embryonic tissues extracted in passive lysis buffer. As the *Nanos*-Cas9 transgenic line was no longer being maintained by the time of the experiment, a comparable sample of heterozygote ovarian tissue in passive lysis buffer could not be obtained. While absent in T4 ovaries, the *Vasa*-Cas9 ovary extract was associated with a clear Cas9 band around the expected size of 165 kDA, with a band at the same size apparent in the positive control sample of purified Cas9 protein.

The immunoblot was performed only once, intended as a quick qualitative estimate of detectable Cas9 signal to rule out the scenario of non-existent expression in the transgenic lines. Although the evidence for Cas9 presence in *Vasa*-Cas9 ovaries was convincing, and so merited inclusion here, several other issues were associated with this blot that limited its reliability and interpretation. Despite repeated attempts, no signal could be generated in any of the samples for the endogenous control gene beta-actin with the antibody used in this study. With no bands being produced for either Cas9 or beta-actin in the T4 ovary sample, the potential for poor sample quality in this tissue hinders its dependability as a true Cas9-negative control.

In the same blot, extracts from fertilised egg clutches were assessed alongside the ovarian tissue samples, with eggs sampled at the 2 hours AEL timepoint for T4 flies, and at 2 and 4 hours AEL for *Vasa*-Cas9. In addition, a *Nanos*-Cas9 2-hour AEL egg sample, originally stored in TRIzol for RNA extraction, was processed for protein extraction and the sample included in the western blot analysis. No clear Cas9 band was seen in *Nanos*-Cas9 eggs, although the low protein sample concentration and quality obtained following the TRIzol protein extraction method was estimated to hinder detection compared with the passive lysis buffer-extracted samples.

Despite the *Vasa*-Cas9 tissues both producing clear bands of the expected Cas9 size, increasing in intensity with the age of eggs sampled, a strong band at the same molecular weight was noted in the T4 egg sample, with smearing apparent in all egg sample lanes. From previous experiments, the primary Cas9 antibody used in this study was understood to bind non-specifically in both transgenic and *wild-type* embryo tissue samples isolated

from *A. gambiae* (personal communication, Dr. Moeez Khan, Imperial College London, August 2024)[565]. As non-specific binding was also evident in the BSF embryo tissue samples examined here, no conclusions could be made regarding the presence of Cas9 protein in transgenic egg clutches from this experiment, with alternative primary Cas9 antibodies being required for any potential future analyses. The original western blot image with the embryo tissue samples is shown in full in the appendix (**Supplementary figure 6**).

5.2.7 - Enabling Endogenous CRISPR Activity in Cas9-Expressing Embryos via gRNA Plasmid Microinjection

Two different nuclease and promoter combinations were assessed during the CRISPR cross, with a potential Cas9-associated toxicity phenotype also noted in the *Act5C*-Cas9 and *Nanos*-Cas9 lines. With evidence from the RT-PCR and western blot experiments supporting nuclease production in transgenic ovarian tissues on a transcriptomic and translational level, the efficacy of the CRISPR RNA-expressing transgenic constructs was interrogated.

Although 9 different target sites were examined during the CRISPR cross experiments, each of the three tested RNAs were designed to be expressed under the control of the U6:3 promoter. Although production of the U6:1 and U6:2 gRNA-*w* transgenic lines was earlier attempted, the G1 pools encoding these constructs clearly had a sub-optimal fitness when compared with G1s transgenic for U6:3 gRNA-*w*. Although several factors may have contributed to the inability to generate transgenic G2s for the constructs encoding BSF-endogenous U6 promoters, their corresponding low fitness at G1 could also indicate significant constitutive gRNA expression. While these identified endogenous promoters correspond closely with those employed for transgenic shRNA expression in a separate study, no evidence exists describing the use of *D. melanogaster* U6:3 in this species[422]. Sequence alignment of each promoter sequence, as shown in an earlier section, also indicates potential differences between U6 regulation in the two species, most evident in the TATA box. Taken together, these data suggest that U6:3, and potentially *D. melanogaster* Pol III promoters more generally, may be insufficient for high transgenic expression of CRISPR RNA.

Due to the small size and extensive secondary structures associated with CRISPR RNAs, cDNA synthesis and RT-PCR with tissues from the U6:3 regulated RNA lines was not attempted. Despite failing to detect indels in transheterozygote progeny during the earlier cross experiments, the potential utility of the nuclease-expressing lines for producing CRISPR-associated gene editing was considered in an alternate context. Transgenic studies in several other species employ nuclease-expressing lines for gene editing by injecting a chosen gRNA sequence directly into transgenic embryos, which act as the nuclease source. Despite microinjection still being required during this scheme, the stable and characterised endogenous source for the nuclease component can significantly increase editing efficiency and improve the reproducibility of these experiments. Omitting the nuclease component from the injection makes the method technically easier and less resource intensive to achieve, and likely contributes to a lower nuclease-associate toxicity when compared with high-concentration Cas9-RNP injections. This process could be further improved by replacing synthesized sgRNA with plasmid DNA encoding a gRNA expression cassette, with the associated injection mix having significantly higher stability. Additionally, the target spacer sequences could be quickly exchanged in injection plasmids through simple cloning methods, such as those described in this study.

With the original *w* gRNA spacer sequence proving sufficient for activity, the U6:1, U6:2 and U6:3 gRNA-*w* plasmid constructs were injected directly into *Vasa*-Cas9 embryos. Eggs originated from heterozygous transgenic mothers and T4 *wild-type* fathers and were injected between 1-2 hours AEL (G0). Between 240 and 265 eggs were injected per construct using the same methods as performed when generating transgenics, with editing activity expected to be the highest in the tissues originating near the posterior injection site. RNA-expression plasmid concentration was standardised at 600 ng / µL.

While this crossing setup would result in half of the injected eggs being non-transgenic for *Vasa*-Cas9, this was preferred for several reasons. Crossing male and female heterozygotes would lead to a mix of heterozygote and homozygote G0 injectants that could receive different dosage levels of Cas9 and produce variable editing efficiencies. A homozygote *Vasa*-Cas9 and *wild-type* cross, which would produce an entirely heterozygote transgenic progeny, was also avoided. Microinjection into transgenic lines had not been previously attempted, and a potentially lower injection survival in transgenic eggs was considered likely, due to the predicted additional burden of Cas9 expression and likely active gene editing in early transgenic G0 embryos. As such, microinjection of both heterozygote transgenic and non-transgenic siblings prior to transgenic identification by marker

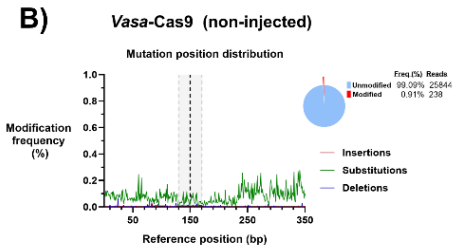
expression would allow for detection of a substantially reduced survival rate in transgenic injectants.

From the G0 neonates that survived injection, no significant difference was found between transgenic and non-transgenic survival rates following injection with any of the U6 constructs (**Figure 44A**). Injection survival rate was generally lower than encountered when injecting constructs for transgenesis, although this was inclusive of both transgenic and non-transgenic G0. Due to low numbers of G0 transgenic neonates and previous experience with poor crossing success at low cage densities, the decision was made to sacrifice more individuals at the G0 stage at the expense of the later mutagenic cross, and focusing on making a full characterisation of indels at the molecular level.

5 transgenic G0 neonates were sampled for each injection condition for pooled gDNA extraction, along with 5 transgenic non-injected neonates for a reference genotype. The remaining transgenic G0 were reared until adulthood, with ovaries from 5 females being sampled for each injection condition for gDNA isolation. Ovaries were chosen as tissues most likely to contain mosaic gene edits, having derived from progenitor cells with high expected Cas9 transgene expression and close in proximity to the site of RNA-plasmid injection. The remaining female adult G0 for each construct were crossed to the male G0. However, as increasing cage density by addition of wild-type flies would reduce the chance of detecting *w* mutants from two G0 parents in the G1 generation, the cages were not supplemented, with minimal G0 breeding activity observed and no fertilised G1 clutches collected for each condition.

A)

Construct	Eggs injected	Neonates hatched	Transgenic Neonates hatched	Transgenic neonates sampled	Transgenic male adults	Transgenic female adults	Total transgenic adults	Transgenic females sampled
Vasa-Cas9 (non-injected)	100 (*aligned only*)	94 (94.%)	49 (52.1%)	5	–	–	–	–
U6:1 gRNA-*w* (600 ng / µL)	240	73 (30.4%)	31 (42.5%)	5	11	13	24 (92.3%)	5
U6:2 gRNA-*w* (600 ng / µL)	240	98 (40.8%)	48 (49%)	5	20	20	40 (93.0%)	5
U6:3 gRNA-*w* (600 ng / µL)	265	125 (47.2%)	54 (43.2%)	5	25	21	46 (93.8%)	5



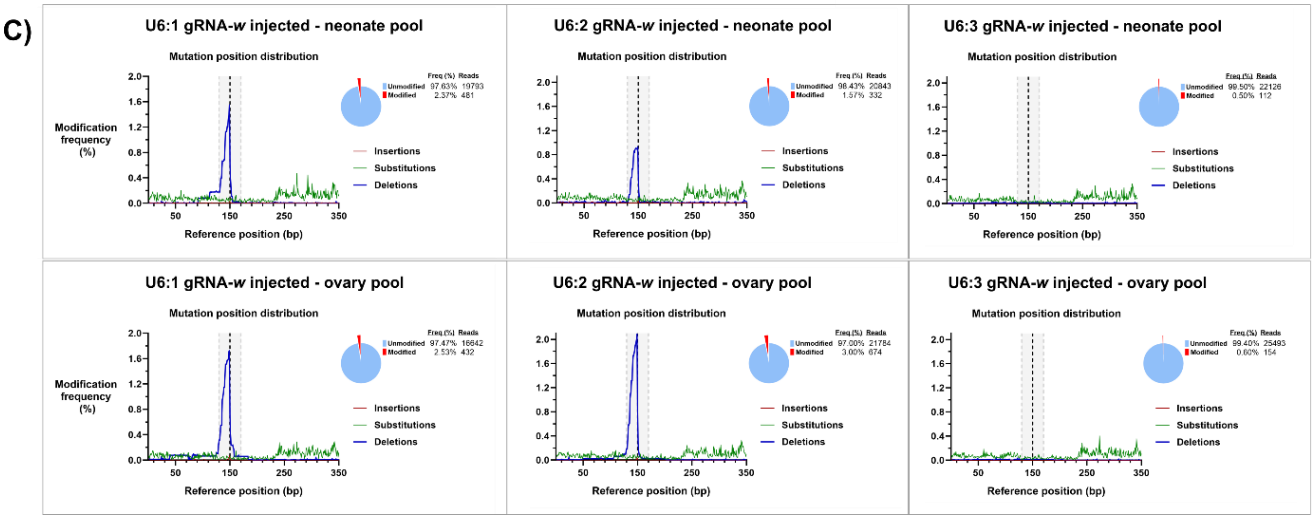


Figure 44: Injection data and amplicon sequencing analysis of *vasa*-Cas9 embryo microinjection with gRNA-w plasmid. A) Injection statistics following embryonic microinjection of eggs produced by crossing male T4 flies with females heterozygote for the *Vasa*-Cas9 transgenic construct. Neonate hatching rates following injection with one of the three gRNA plasmids was between 30.4% and 47.2%, with approximately half of all injection survivors inheriting the *Vasa*-Cas9 transgene. **B)** 5 *Vasa*-Cas9 encoding neonate G0 injection survivors were pooled for gDNA extraction for each injection condition, alongside a non-injected control sample. For each sample, a 390 bp region flanking the *white* target site was amplified and submitted for amplicon sequencing. Due to high sequencing noise seen in the 15 base-pair regions flanking the amplicon in all conditions, the amplicon mutation position distributions were trimmed at both ends and centred around the *white* target site. Sequencing from the non-injected *Vasa*-Cas9 neonate sample indicated a modification frequency of 0.91%, primarily consisting of substitution mutations evenly distributed across the amplicon. As minimal insertion or deletion mutations were detected, this mutant profile was attributed to noise from PCR and sequencing errors. **C)** In addition to G0 neonate samples, ovary tissues of 5 G0 female flies following injection were dissected and pooled for each injection condition. Sequencing of amplicons produced from each tissue sample produced a similar noise profile of evenly-distributed substitutions as seen in the non-injected control. In addition, both tissue types examined were associated with relatively high modification rates in G0 survivors injected with the U6:1 gRNA-*w* (2.37-2.53% modified) or U6:2 gRNA-*w* (1.57%-3.00% modified) plasmids, which were mostly attributed to deletion mutations located within 20 bp of the expected cut site. In comparison, U6:3 gRNA-*w* injected survivors had a low modification rate (0.50%-0.60%) across both tissue types, and lacked the deletion profile seen in survivors injected with the two other promoter constructs, resulting in a mutation distribution similar to the non-injected control sample.

A 390 bp amplicon surrounding the *w* target was produced for the G0 neonate and adult ovary pooled gDNA extractions for each injected construct, and submitted for 2 x 250 bp Illumina amplicon sequencing with an amplicon from a non-injected *Vasa*-Cas9 neonate used as a baseline for mutation detection (**Figure 44B**). 0.91% of the aligned reads from the non-injected pool were classified as modified, although this was represented almost entirely of alleles that differed from the *w* reference sequence by a single base substitution. While substitution was encountered at slightly higher frequencies in a region downstream of the target site in all conditions, substitutions are otherwise distributed evenly over the amplicon with no mutant allele being significantly enriched. This was attributed to be PCR artefacts introduced in production of the amplicon.

The modification frequencies and mutation position distributions were then examined in *w* amplicons produced from injected neonate and ovary tissues (**Figure 44C**). A clear enrichment of deletion alleles was noted in both G0 tissue types subject to injection with U6:1 and U6:2 gRNA-*w.* Within a 20bp window upstream of the PAM site, the frequency of deletion increased dramatically with proximity to the expected cut site, characteristic of a Cas9-generated indel profile. 2.37% and 2.53% of recovered alleles were classified as modified for U6:1 gRNA-*w* G0 neonate and ovary tissues respectively, with corresponding rates in 1.57% and 3.0% in U6:2 gRNA-*w* G0s. Conversely, only 0.5% and 0.6% of reads were modified in U6:3 gRNA-*w* injected neonate and ovary tissues respectively, which were both lower than the 0.93% rate encountered in the non-injected baseline reads. No clear deletion profile was noted within the 20bp target window in either U6:3 G0 tissue, with substitutions evenly distributed across the amplicon representing the majority of identified modifications.

For a fine-grained analysis of specific indel generation at the target site, the alleles with most recorded reads that differed from each other within 50 bp around the expected cut site were aligned (**Figure 45**). For each gDNA sample, the top 13 unique alleles in the amplicon were mapped in the order of their respective frequencies; as the vast majority of alleles encoded the *wild-type* sequence, the 12 most common alternative alleles could then be visualised and compared. As seen in the global mutation distributions, reads from the non-injected, U6:3 gRNA-*w* neonate and U6:3 gRNA-*w* ovary amplicons showed comparable allele profiles around the *w* target site. The *wild-type* sequence represented 97.21% of reads of the non-injected neonate amplicon, and 97.96% and 98.31% in U6:3-injected neonates and ovaries. The next most frequent alleles represented only 0.08% to 0.12% of sequences in the three groups, or between 18 and 26 reads. In each case, the 12 most frequent alternative

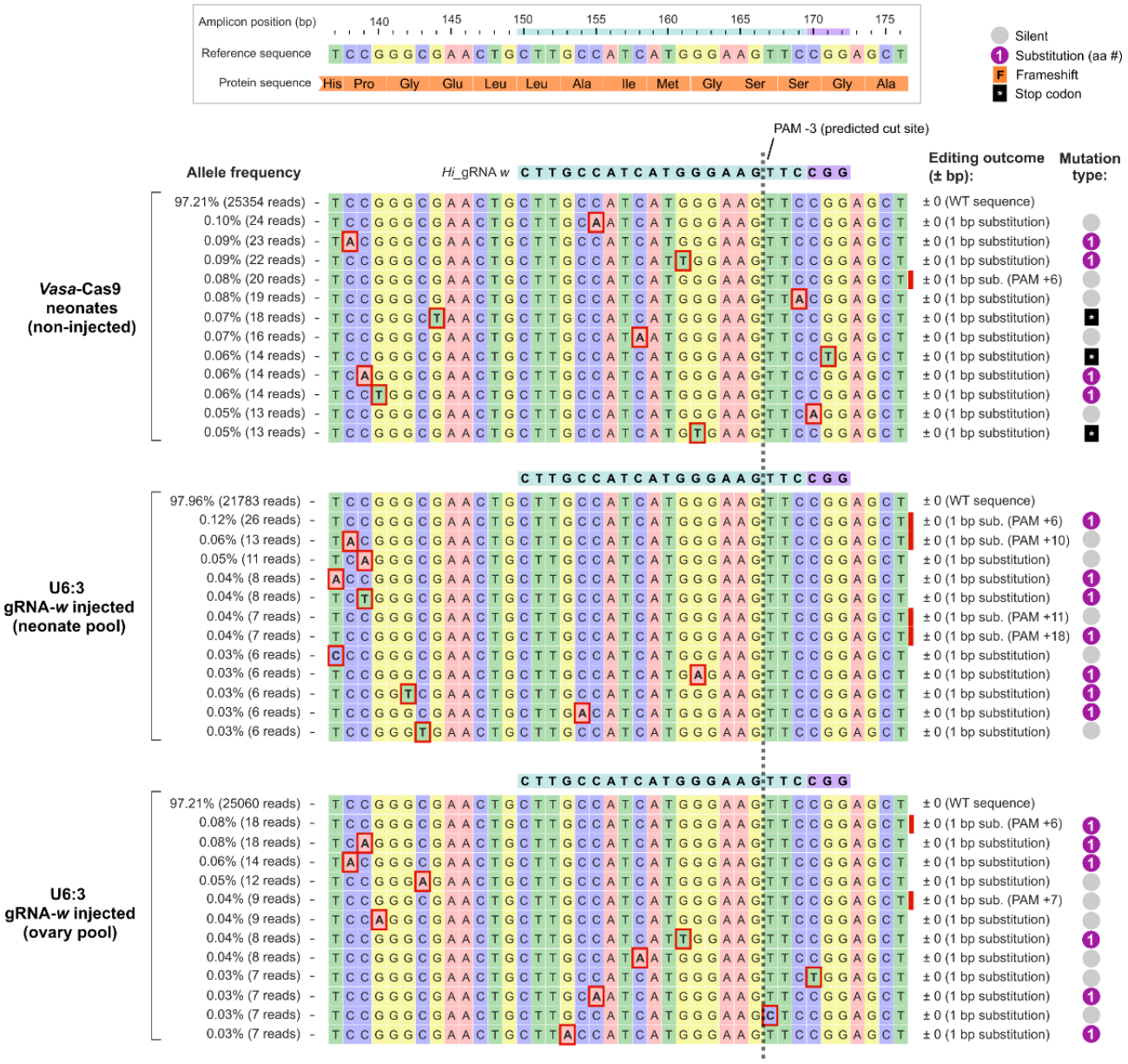


Figure 45: Allele frequency and alignment following amplicon sequencing of non-injected and U6:3 gRNA-*w* injected *Vasa*-Cas9 tissues. Post sequencing, the unmodified *wild-type* and the 12 most common alternative alleles, defined as having modified sequence within 20 bp of the cut site, were aligned in order of number of detected reads. Deviations from the unmodified reference sequence are annotated in the alternative alleles. No insertion or deletion mutants were among the most common reads for the non-injected and both U6:3 gRNA-*w* tissue conditions, with each of the common modified sequences differing from the unmodified reference by a single base substitution. The pattern of substitution is similarly evenly-distributed across the examined quantification window, and not enriched close to the expected cut site. This distribution, in addition to the low overall counts of modified reads, was attributed to sequencing noise, with the profiles generated being deemed to not be due to true target site-specific mutagenic activity.

alleles in the target window represented single base substitutions, which were evenly distributed and not clustered around the expected gRNA cut site. With all full-length reads aligned to the reference amplicon sequence, a comparable distribution of substitutions was encountered at similar frequencies for any chosen alignment window examined in the three tissue groups, indicating no specific enrichment for increased modification rates around the target site (data not shown).

Conversely, the highest frequency alternate alleles in both tissue groups for U6:1 and U6:2 gRNA-*w* injected G0 all showed extensive indel formation centred around the predicted Cas9 cleavage site (**Figure 46**). Although low, slightly higher proportions of the reads were represented by the highest frequency indel alleles, with 47 (0.24%) and 57 (0.26%) reads mapping to top alternate allele in U6:1 and U6:2 tissues respectively. Deletion sizes ranged between 1 and 61 bp, leading to frameshift in the *w* coding sequence in 87.5% of modified alleles. For in frame mutants, 3 single amino acid substitutions were recorded in the U6:2 neonate reads, with U6:1 reads mutants including deletions of 1, 5 and 14 amino acids. A specific 10 bp deletion (-9 to +1 bp relative to the expected cleavage site) was encountered at the highest frequency in all four pools, suggesting that this was the most likely repair outcome. A 1 bp (-1 bp relative to the cut site) and 8 bp (-8 bp relative to the cut site) were also common to each pool.

Amplicon sequencing analysis:
U6:1 and U6:2 gRNA-*w* injected

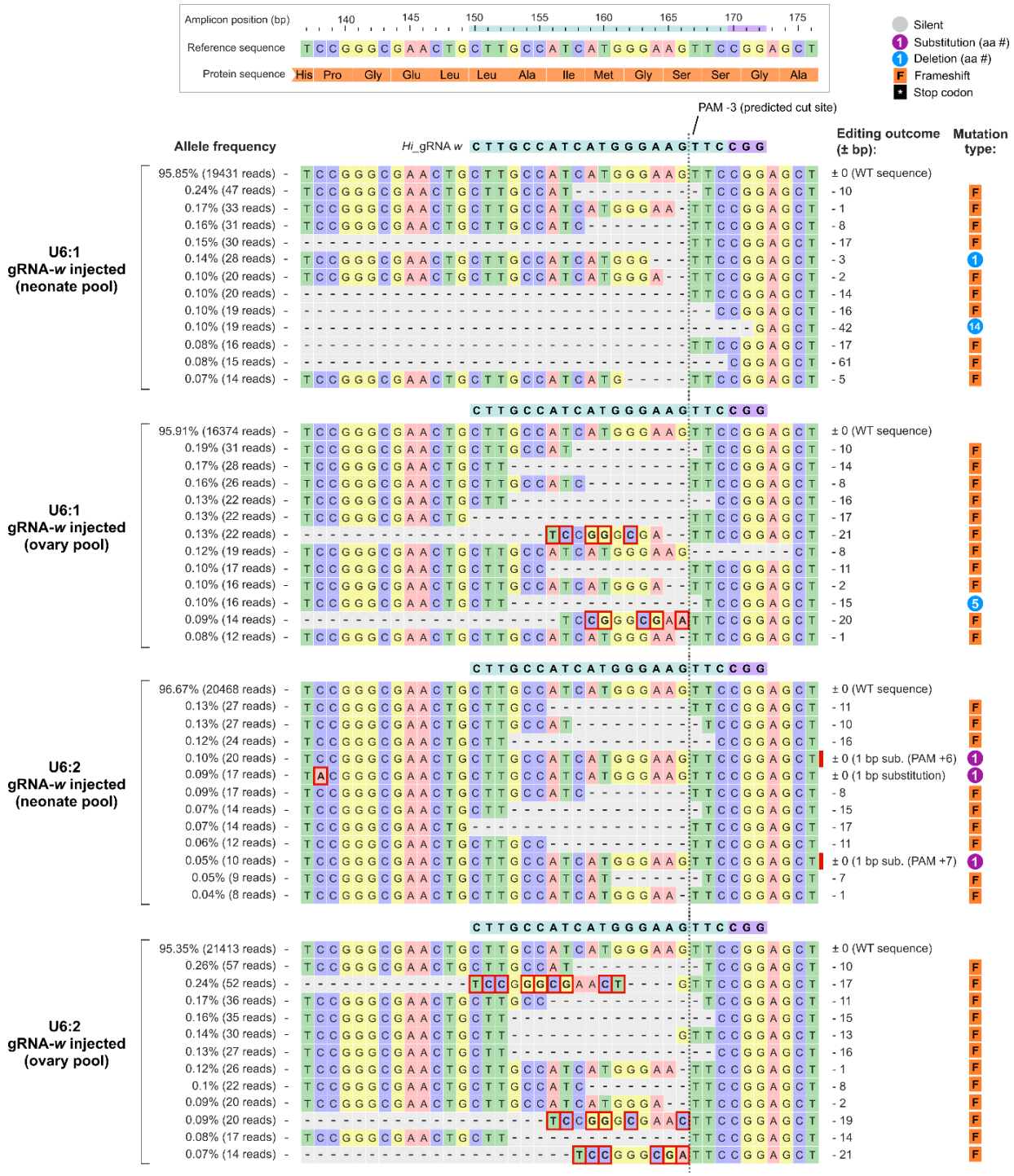


Figure 46: Allele frequency and alignment following amplicon sequencing of U6:1 gRNA-*w* and U6:2 gRNA-*w* injected *Vasa*-Cas9 tissues. In contrast to the non-injected and U6:3 gRNA-*w* injected samples, the U6:1 gRNA-*w* and U6:2 gRNA-*w* injected tissue samples were associated with an overall higher proportion of modified to unmodified reads, although the proportion of mutant sequences was still low overall, presumably due to the mosaic nature of mutation in the G0 stage following injection. The majority of the most common modified reads for each injection condition resulted from deletion mutations ranging between 1 – 61 bases, leading to a frameshift in most cases. In each condition, the pattern of deleted bases was significantly enriched around the gRNA binding site immediately upstream of the expected cut site, further evidencing that the resulting mutant profiles detected were likely to have originated from site-specific targeting by Cas9 nuclease followed by error-prone NHEJ repair.

5.3 – Conclusion

Although the G0 crosses for generation of *w* mutants was not successful, both U6:1 and U6:2 gRNA-*w* constructs demonstrated a clear potential for gRNA expression in BSF, leading to CRISPR activity when injected in Cas9-expressing embryos. Conversely, injection of U6:3 gRNA-*w* did not produce any mutagenic activity, at least that could be detected when compared with the non-injected Cas9 line. This calls in to doubt the efficacy of the U6:3 promoter for RNA expression in BSF, and could reasonably account for why no evidence of gene editing activity was found in the *w* or GRG targets in the progeny of the transgenic CRISPR crosses.

Although >95% of sequenced reads from the U6:1 and U6:2 injection conditions corresponded to non-modified *wild-type* sequence, it remains undetermined how enriched CRISPR activity would need to be in germline tissues to reliably generate a *w* mutant phenotype in self-crossed G0. The G0 indels are inherently mosaic, with a non-even distribution of edited cells across the organism being likely, although the potential for inducing hereditable mutants in the G1 generation may be enhanced due to the proximity of injection to the germline tissues and enriched germline nuclease expression. A future investigation could compare these data with the results of G0 injected with a classical Cas9-RNP complex, establishing editing efficiencies relating to both exogenous and endogenous sources of the CRISPR components.

Chapter 6: **Discussion and Future Directions**

The results of this study are significant as it marks the first time that endogenous CRISPR components have been expressed in BSF, including transgenic nuclease and plasmid-based gRNAs. Although shRNA has previously been expressed from transgenic plasmids, gRNA expression is novel. This is also the first instance where CRISPR nuclease and gRNA do not need to be delivered as protein or pure RNA, respectively, with activity detected from plasmid DNA injection into nuclease-encoding transgenic embryos. Although slightly more attractive than injecting Cas9 RNPs from a technical standpoint, the efficiency of this method compared to classical Cas9-RNP injection still needs to be examined, with DNA injection needing to have a similar probability of success to be useful as a technique. Furthermore, plasmid-based RNA enables the easy testing of multiple targets with flexible and rapid cloning, avoiding the more labour-intensive synthesis of RNA for each target when produced in house. However, the price of synthetic gRNAs has dropped significantly in the time since this study was first conceived, with commercially-obtained gRNAs likely to allow more reproducible mutagenesis experiments than *in vitro* transcribed stocks.

The results presented in the final section of this study suggest that *Vasa*-Cas9 transgenics support nuclease expression suitable for G0 embryo tissue mutagenesis without harmful effects on the animal overall. No significant difference was found in injection survival rates between *vasa*-Cas9 transgenic and non-transgenic eggs, indicating good suitability for injection. Although not attempted in this study, one obvious follow up experiment to the plasmid injection into the *Vasa*-Cas9 embryos would be to inject *w* sgRNA to see if omitting the transcription step in embryos leads to higher Cas9 activity. Additionally, future work should compare the efficacy of Cas9-gRNA RNP injections with those outlined here to determine relative efficiency in G0 editing and G1 mutant production through crossing. Finally, using a homozygous Cas9 line could result in higher editing rates across all injected eggs, allowing for rearing all injectants for G1 mutants in numbers similar to those seen in pBac injections (200-250 eggs per session), with high survival rates supporting effective G1 crosses.

Although it was not possible to establish a stable transgenic line for U6:1 and U6:2 directed RNA expression plasmids during this study, the final injection experiments suggest that these promoters exhibit activity for gRNA expression, leading to CRISPR activity, while U6:3 did not show the same effect. This could explain the lack of activity observed at the cross

stage. Therefore, it's likely that the CRISPR cross system could function as intended when using plasmids with the more active U6:1 and U6:2 Pol III promoters, a hypothesis that could be tested by creating new U6:1 and U6:2 lines.

Unfortunately, due to the considerable time involved with BSF transgenesis, it was not possible to realistically complete the process of establishing the U6:1 and U6:2 gRNA-*w* lines, crossing them with a nuclease line, and examining their progeny within this study, though this would be a clear direction for future work. This study has provided evidence that the U6:3 promoter from *Drosophila* may not be suitable for gRNA expression in BSF. Interestingly, while many *Drosophila* Pol II elements tested, such as *hsp70*, *actin5c,* and *PUb,* perform well in BSF, U6:3 does not, despite showing high activity in other insect models. Differences in promoter structure, particularly an abnormal sequence in the BSF TATA box, may explain the lack of U6:3 activity. Given these findings, it would be advisable to shift focus from U6:3 to endogenous U6 sources within the BSF genome, where many other candidate promoters exist.

Regarding that lack of MAD7 activity seen in transgenic BSF, the most likely explanation is that only U6:3 promoter constructs were paired with MAD7 lines, as U6:3 may have limited activity in BSF. This is supported by observed activity in U6:3-transfected cell cultures compared to constructs with U6:1 and U6:2 promoters. In this study, constructs were designed with tRNA flanking regions, based on studies that demonstrated enhanced stability and activity for Cas12a-associated crRNA arrays due to host tRNA processing. However, a simpler expression cassette, containing only the variable and direct repeat regions, was not tested here.

Additionally, no single-target crRNAs were designed for BSF experiments, as MAD7's strength lies in multiplexing capability. In future research, it may be beneficial to design new crRNAs, including versions without tRNAs, tRNAs at only the 5' or 3' ends, and a simple crRNA with one direct repeat and one variable region. The latter could enable direct comparisons of single-target editing activity between Cas9 and MAD7 in the same genomic region, providing data on indel profiles. MAD7 is expected to produce larger indels than Cas9, with comparable activity shown in various studies. Moreover, the smaller size of MAD7's coding sequence and crRNA compared to Cas9 could simplify transgenesis and construct handling.

Following the conclusion of the experimental work described in this study, a separate group of researchers reported the results of their own work regarding the development of a similar CRISPR toolkit in transgenic BSF. In a 2024 publication, Kou *et al.* examined the expression of the *vasa*, *nanos* and *exuperantia* genes in BSF ovaries, testes and embryos 2 hours post-laying by qPCR[566]. Concluding that *vasa* had the lowest expression in germline tissues of the three genes tested, the researchers proceeded to isolate between 2-3 kb of sequence upstream of the *nanos* and *exu* genes to act as promoters driving the germline-enriched expression of Cas9 in *piggyBac* transgenesis constructs.

After introducing both constructs into BSF, the expression of Cas9 in ovaries and testes was confirmed by RT PCR, after which both nuclease lines were crossed to a third line encoding two BSF-endogenous U6-driven gRNAs against the *white* gene. *w* mosaic mutant larva were detected among the F1 transheterozygote progeny of the cross, indicating successful CRISPR activity following the combination of two germline-restricted nuclease lines with a gRNA expressing line. No fitness deficit was reported for the nuclease constructs over three generations, although transgene inheritance rates were not reported. Interestingly, the study did not mention the crossing of the *w* mosaic individuals of the F1 generation in an effort generate and analyse stable knockout genotypes. Although the proportion of mosaic *w* mutants to non-mutants screened in the F1 generation was not reported, the authors revealed a knockout phenotype of 100% of the transheterozygote F1 progeny when crossing either nuclease construct to a gRNA-expressing line targeting the wing development gene *vestigial.*

From the data reported in this recent publication, it seems clear that the methods followed and directions taken during the course of the current study were well-founded and appropriate for the stated research goals, and would have likely resulted in comparable findings had the project been brought to its conclusion. Assuming that the activity of germline-restricted nuclease constructs generated in this study may have been similar to those discussed in the publication, having been produced through similar gene identification and promoter isolation strategies, it is possible that high rates of mutation and mosaicism in the *white* gene could have been realised through the CRISPR cross strategy outlined at the outset of this study. In line with the findings from the concluding experiments performed in this project, the use of the *D. melanogaster*-derived U6:3 promoter for gRNA expression, which was not employed in the mentioned publication in favour of BSF-endogenous U6 promoters, was likely a large barrier to efficient CRISPR activity in the transgenic crosses performed.

The high rates of mutagenesis at the F1 stage reported in the Kou *et al.* (2024) study, as well as interesting additional data suggesting significant CRISPR activity in nuclease-deficient progeny as a result of maternal Cas9 deposition, firmly support the hypothesis that the modular CRISPR cross strategy is a feasible and highly effective method of knocking out genes in the species[566]. It also reinforces the idea that future work in this area should focus on the use of endogenous U6 promoters, as well as the described germline-regulated promoters, for construct designs with high chance of functioning as intended. However, as the full nature and degree of tissue-restricted nuclease expression was not comprehensively assessed in either of these studies, further research may still be necessary before the employment of CRISPR strategies in BSF that are dependent on a strict pattern of germline-restricted expression, such as might be required with genetic control measures and biosafe containment strategies.

With high feedstock tolerance, rapid growth rates, low general FCR and resilience to foreign contamination, BSFL could represent an ideal production platform, or “chassis” organism, for the synthesis of numerous high-value bio-based commodities. Insect cell culture is already widely employed for the industrial production of recombinant proteins, carrying the advantages of easier culture, increased tolerance and high expression level when compared with certain microbial or mammalian production systems[567]. BSFL has already been indicated as a potential bioreactor, with potential to express high amounts of bulk chemicals or more complex products like AMPs and bioremediatory enzymes[264]. The advantages of this would be a biomanufacturing platform with low feedstock and process costs, simple scalability and low risk of contamination. However, this would need to be coupled with a cheap and efficient post-treatment process for sufficient purification of products from larval biomass, with the cost-effectiveness of BSFL as a biomanufacturing platform remaining to be seen.

Despite the considerable number of avenues for genetic improvement in BSF, further development towards enhancing these traits is currently limited to the few genetic tools available in BSF. With limited experimental detail found throughout the literature, even the basic CRISPR-Cas9 system, *pBac* transformation system and embryonic microinjection genetic transduction technique could stand to benefit from a more rigorous characterisation in this insect, with more data and methodologies being a benefit for other researchers.

Although no clear difference was observed between isolines in terms of marker expression intensity, numerous other factors and position effects could influence the function of the transgenic components at different loci. While outside of the scope of this study, a future

investigation could focus on the comparison of expression and behaviour of identical transgenes at numerous integration sites, examining the degree to which positional effects may enhance or suppress the functionality of non-insulated constructs intended for integration at a specific genomic locus. Understanding the distribution of likely transposon-directed integration sites may also be of use when considering the design of strains that require multiple independent integrations, such as when reconstituting multi-stage metabolic pathways in BSF for commercial synthesis applications[568].

For expanding the current molecular toolkit available to BSF, a significant enabling factor will be the characterisation of site-specific integration using CRISPR-induced HDR. While effective at generating transgenic lines, the random insertion profile of *piggyBac* mobilised transgenes may be limiting in terms of disrupting fitness at certain sites, as suspected with the *Nanos*-Cas9 line that was lost due to low fitness in the current study. With the ability to integrate transgenes at any chosen locus in the genome, different DNA cassettes could be exchanged at these sites using transposon systems like Cre recombinase and *attP* phiC31 integrase systems[569]. With the fundamental CRISPR technology well-established, more novel nucleases such as Cas13 could be employed for transcriptional regulation through RNA-editing, in addition to CRISPR interference / activation fusion proteins and RNAi[570]. More precise, templated repair outcomes at a designated target site could similarly be mediated by engineered nucleases through techniques like base editing and prime editing[571,572].

In addition to GE and transgenesis technologies, several fundamental genetic elements will need to be characterised going forward for controlling transgene expression in BSF. For instance, the activity of multiple constitutive and developmentally-regulated promoters should be assessed for their suitability in maintaining a desired expression level, for both functional transgenes and transgenic markers. To increase the diversity of regulatory elements available to this insect, endogenous, exogenous and synthetic promoters should be analysed for their relative pattern and level of expression. Going forward, even inducible promoters, such as the *Tet* system and heat-shock responsive elements, could be employed in BSF as a controllable switch for activity. If found effective, these could be used for inducing CRISPR or RNA interference activities, overexpression of a transgenic product, or be useful as part of a containment strategy for enhanced biosafety[264].

There are trade-offs for most methods of gene editing activity. The most obvious downside to generating mutants using the CRISPR-cross strategy is the need to generate both nuclease and gRNA-encoding transgenic lines independently, which itself may take more

time and resources for line generation and verification than producing the edit directly through injection of the active CRISPR components. However, nuclease lines can be kept indefinitely once generated, providing a characterised stock that should behave reproducibly when used in later CRISPR crosses. The inclusion of different regulatory elements, as well as the positional effect of the genomic integration site, also allows for more fine tuning of the component's expression, allowing for more specific levels and patterns of activity in different contexts that have a reasonable level of reproducibility.

It is important to recognize that, despite its potential, insect farming must be managed carefully to realise its sustainable and eco-friendly promise. First, it's worth noting that insect rearing, especially for BSF, primarily targets the fish and animal feed markets. Although BSF meal is less environmentally damaging than alternatives like soy or fishmeal, it doesn't fully address the environmental issues linked to livestock—such as land, water, and emissions. While BSF farming offers some environmental benefits, it could be seen as indirectly supporting meat-based diets, which are among the major climate challenges and are expected to grow with rising living standards. Additionally, strict food regulations require that BSF have defined, safe diets before they can be fed to livestock, often meaning BSF are raised on soy or fodder originally intended for livestock. This setup, while potentially more economical or practical, reduces the environmental benefits of BSF farming. As a young sector, insect farming presents an opportunity for policymakers to establish guidelines that contribute to a genuinely sustainable food system rather than reinforcing current inefficient production methods.

The economic viability of insect farming may sometimes clash with its sustainability goals, especially when scaled up to supplement livestock. Emphasis on high-tech solutions like insect farming and transgenic technologies could also serve as an expensive diversion from simpler, more impactful solutions for food security—such as shifting toward plant-based diets and relying on more efficient crops. Insect farming should not be used as a justification for food waste inefficiencies. A significant portion of resources in food production and transport cannot be recovered through insect farming. The waste and emissions tied to overproduction, processing, and transport are better addressed by reducing over-purchasing and adopting less wasteful consumption habits at the individual, cultural, and population levels. It has been suggested that global food production could already be sufficient to meet caloric demands, with food security challenges stemming largely from distribution issues. In this light, greater economic and political efforts are needed to address root causes of food insecurity, including food transport infrastructure, conflict, disease,

poverty, and inequality. Ultimately, promising alternatives like insect farming will only be part of a much larger solution needed to address the dual challenges of feeding the global population sustainably while protecting the environment.

Conclusion

Insect farming offers a promising solution for food waste valorisation, which is an essential step toward enhancing food security in a sustainable and cost-effective manner. Among insect species, the black soldier fly stands out as a key candidate for breeders, thanks to its high tolerance for diverse feedstocks and nutrient-rich larvae. This has sparked interest in developing high-performance insects for animal feed and production purposes. Gene editing and transgenic technologies are anticipated to expand BSF's utility, potentially boosting yields and opening new applications in agriculture, industry, and commerce. While progress is being made in producing genetically edited BSF strains, a comprehensive understanding of the CRISPR system in BSF remains incomplete. Further validation of transgenic and gene editing technologies could help to enable a range of stakeholders to leverage the benefits of BSF genetic engineering for their own applications.

This study contributes to the development of a BSF genetic toolbox to facilitate the screening of endogenous genes of interest. A more efficient BSF transgenesis method was optimised using embryo microinjection and *piggyBac*-mediated random genome integration. Twelve transgenic and gene-edited BSF strains were produced, covering key elements for a segregated, heritable CRISPR system. The study compared three Pol III promoters for guide RNA expression, as well as two germline-specific and four constitutive promoters for nuclease and marker expression. Although no CRISPR activity was observed in the progeny of CRISPR nuclease and RNA-expressing lines after crossing, evidence suggested successful transgenic nuclease expression in the germline when employing the regulatory elements from the endogenous *nanos* and *vasa* genes. However, more work is still required to assess the degree of germline enrichment with the described constructs, and to determine if a significant amount of leaky expression is encountered in non-germline tissues. Additionally, CRISPR RNA expression from two of the three Pol III promoters was indicated following plasmid microinjection into nuclease-expressing embryos, with CRISPR-associated indel activity detected in the germline.

Collectively, the tools developed in this study hold promise for refining a functional endogenous CRISPR system in BSF based on crossing transgenic lines. This approach could enable targeted gene editing without requiring external technology, provided that the lines are established as outlined here. While this strategy may delay mutant analysis, it offers significant advantages by allowing non-specialized laboratories and producers to assess

mutants in their own settings without the fitness costs or efficiency biases associated with injection. This decentralised system for producing and testing transgenic strains supports larger-scale experiments. Although no MAD7 activity was detected in BSF in this study, this approach to mutant generation aligns well with the exploration of multiple genes of interest, supporting more complex gene stacking or metabolic engineering in BSF. The techniques described here are intended to support further advancements in BSF genetic technology and represent a step toward site-specific genome integration via HDR, incorporation of complex transgenes with multiple regulatory elements, multiplexed genome editing, and fine-tuned expression using next-generation CRISPR tools, such as CRISPR activation, inhibition, and prime editing.

Bibliography

1. FAO, IFAD, UNICEF, WFP & WHO. *The State of Food Security and Nutrition in the World 2024*. (FAO ; IFAD ; UNICEF ; WFP ; WHO ;, 2024).
2. Beyene, S. D. The impact of food insecurity on health outcomes: empirical evidence from sub-Saharan African countries. *BMC Public Health* **23**, 338 (2023).
3. FAO, IFAD, UNICEF, WFP, & WHO. *The State of Food Security and Nutrition in the World 2021: Transforming food systems for food security, improved nutrition and affordable healthy diets for all*. (FAO, IFAD, UNICEF, WFP and WHO, Rome, Italy, 2021). doi:10.4060/cb4474en.
4. UN General Assembly. *Transforming Our World : The 2030 Agenda for Sustainable Development*. https://www.refworld.org/legal/resolution/unga/2015/en/111816 (2015).
5. UN DESA. *The Sustainable Development Goals Report 2020*. https://desapublications.un.org/publications/sustainable-development-goals-report-2020 (2020).
6. Minot, Nicholas, Rice, Brendan, Vos, Rob, & Food Security Information Network. *2024 Global Report on Food Crises: Joint Analysis for Better Decisions*. https://www.fsinplatform.org/grfc2024 (2024).
7. Woodhill, J., Kishore, A., Njuki, J., Jones, K. & Hasnain, S. Food systems and rural wellbeing: challenges and opportunities. *Food Secur.* **14**, 1099–1121 (2022).
8. Müller, O. & Krawinkel, M. Malnutrition and health in developing countries. *CMAJ Can. Med. Assoc. J.* **173**, 279 (2005).
9. Coles, G. D., Wratten, S. D. & Porter, J. R. Food and nutritional security requires adequate protein as well as energy, delivered from whole-year crop production. *PeerJ* **4**, e2100 (2016).
10. Devi, B., Devi, J. & Bhattacharyya, N. Subsistence Agriculture—An Approach Towards Food Security in Changing Climate. in *Food Production, Diversity, and Safety Under Climate Change* (eds. Chakraborty, R., Mathur, P. & Roy, S.) 53–62 (Springer Nature Switzerland, Cham, 2024). doi:10.1007/978-3-031-51647-4_5.
11. Institute for Health Metrics and Evaluation. *Global Burden of Disease 2021: Findings from the GBD 2021 Study*. https://www.healthdata.org/research-analysis/library/global-burden-disease-2021-findings-gbd-2021-study (2024).
12. Lowe, N. M. The global challenge of hidden hunger: perspectives from the field. *Proc. Nutr. Soc.* **80**, 283–289 (2021).
13. Sporchia, F. *et al.* Zero hunger: future challenges and the way forward towards the achievement of sustainable development goal 2. *Sustain. Earth Rev.* **7**, 10 (2024).
14. FAO. *Tracking Progress on Food and Agriculture-Related SDG Indicators 2023*. https://reliefweb.int/report/world/tracking-progress-food-and-agriculture-related-sdg-indicators-2023 (2023).
15. IMF. *Joint Statement on the Global Food and Nutrition Security Crisis*. https://www.imf.org/en/News/Articles/2023/02/08/pr2335-joint-statement-by-the-fad-imf-wbg-wfp-and-wto-on-food-and-nutrition-security-crisis (2023).
16. Rother, B. *et al.* Tackling the Global Food Crisis: Impact, Policy Response, and the Role of the IMF. *IMF Notes* **2022**, (2022).
17. Feeding the future global population. *Nat. Commun.* **15**, 222 (2024).
18. van Dijk, M., Morley, T., Rau, M. L. & Saghai, Y. A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. *Nat. Food* **2**, 494–501 (2021).
19. Ranganathan, J., Waite, R., Searchinger, T. & Hanson, C. How to Sustainably Feed 10 Billion People by 2050, in 21 Charts. (2018).
20. Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE* **8**, e66428 (2013).
21. Rasul, G. & Sharma, B. The nexus approach to water–energy–food security: an option for adaptation to climate change. *Clim. Policy* **16**, 682–702 (2016).

22. Springmann, M. *et al.* Options for keeping the food system within environmental limits. *Nature* **562**, 519–525 (2018).
23. Viana, C. M., Freire, D., Abrantes, P., Rocha, J. & Pereira, P. Agricultural land systems importance for supporting food security and sustainable development goals: A systematic review. *Sci. Total Environ.* **806**, 150718 (2022).
24. Kraamwinkel, C. T., Beaulieu, A., Dias, T. & Howison, R. A. Planetary limits to soil degradation. *Commun. Earth Environ.* **2**, 1–4 (2021).
25. Desertification, S. of the C. to C. *The Global Land Outlook :: Land Restoration for Recovery and Resilience*. (United Nations Convention to Combat Desertification, 2022).
26. Almond, R.E.A., Grooten M, & Petersen, T. *WWF Living Planet Report 2020 - Bending the Curve of Biodiversity Loss.* (2020).
27. Olesen, R. S., Hall, C. M. & Rasmussen, L. V. Forests support people's food and nutrition security through multiple pathways in low- and middle-income countries. *One Earth* **5**, 1342–1353 (2022).
28. Artaxo, P. Amazon deforestation implications in local/regional climate change. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2317456120 (2023).
29. Potts, J., Voora, V., Lynch, M. & Mammadova, A. Voluntary Sustainability Standards and Biodiversity: Understanding the potential of agricultural standards for biodiversity protection.
30. Poore, J. & Nemecek, T. Reducing food's environmental impacts through producers and consumers. *Science* **360**, 987–992 (2018).
31. Ingrao, C., Strippoli, R., Lagioia, G. & Huisingh, D. Water scarcity in agriculture: An overview of causes, impacts and approaches for reducing the risks. *Heliyon* **9**, e18507 (2023).
32. Crippa, M. *et al.* Food systems are responsible for a third of global anthropogenic GHG emissions. *Nat. Food* **2**, 198–209 (2021).
33. Chataut, G., Bhatta, B., Joshi, D., Subedi, K. & Kafle, K. Greenhouse gases emission from agricultural soil: A review. *J. Agric. Food Res.* **11**, 100533 (2023).
34. Lynch, J., Cain, M., Frame, D. & Pierrehumbert, R. Agriculture's Contribution to Climate Change and Role in Mitigation Is Distinct From Predominantly Fossil CO2-Emitting Sectors. *Front. Sustain. Food Syst.* **4**, (2021).
35. Smith, P., Reay, D. & Smith, J. Agricultural methane emissions and the potential formitigation. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **379**, 20200451 (2021).
36. Hassan, M. U. *et al.* Management Strategies to Mitigate N2O Emissions in Agriculture. *Life* **12**, 439 (2022).
37. Thangarajan, R., Bolan, N. S., Tian, G., Naidu, R. & Kunhikrishnan, A. Role of organic amendment application on greenhouse gas emission from soil. *Sci. Total Environ.* **465**, 72–96 (2013).
38. Whitton, C., Bogueva, D., Marinova, D. & Phillips, C. J. C. Are We Approaching Peak Meat Consumption? Analysis of Meat Consumption from 2000 to 2019 in 35 Countries and Its Relationship to Gross Domestic Product. *Anim. Open Access J. MDPI* **11**, 3466 (2021).
39. Phelps, L. N. & Kaplan, J. O. Land use for animal production in global change studies: Defining and characterizing a framework. *Glob. Change Biol.* **23**, 4457 (2017).
40. Mottet, A. *et al.* Livestock: On our plates or eating at our table? A new analysis of the feed/food debate. *Glob. Food Secur.* **14**, 1–8 (2017).
41. Ritchie, H. & Roser, M. Land Use. *Our World Data* (2024).
42. Mekonnen, M. M. & Hoekstra, A. Y. A Global Assessment of the Water Footprint of Farm Animal Products. *Ecosystems* **15**, 401–415 (2012).
43. Xu, X. *et al.* Global greenhouse gas emissions from animal-based foods are twice those of plant-based foods. *Nat. Food* **2**, 724–732 (2021).
44. Phillips, S. M. *et al.* Commonly consumed protein foods contribute to nutrient intake, diet quality, and nutrient adequacy23456. *Am. J. Clin. Nutr.* **101**, 1346S-1352S (2015).
45. Ludu, J. S. & Plastow, G. S. Livestock and the promise of genomics. *Genome* **56**, 556–566 (2013).
46. González, N., Marquès, M., Nadal, M. & Domingo, J. L. Meat consumption: Which are the current global risks? A review of recent (2010–2020) evidences. *Food Res. Int. Ott. Ont* **137**, 109341 (2020).

47. Rao, N., Shankar, S. & Dutta, A. Spatial poverty traps and food security: Unpacking the linkages. *Econ. Polit. Wkly.* **59**, 51–57 (2024).
48. Parker, A. *Decentralization : The Way Forward for Rural Development?* https://documents.worldbank.org/en/publication/documents-reports/documentdetail/282091468766807862/Decentralization-the-way-forward-for-rural-development (1995).
49. Wood, A. *et al.* Reframing the local–global food systems debate through a resilience lens. *Nat. Food* **4**, 22–29 (2023).
50. Searchinger, T. *et al. World Resources Report: Creating a Sustainable Food Future*. (2018).
51. Sarkar, D. *et al.* Low input sustainable agriculture: A viable climate-smart option for boosting food production in a warming world. *Ecol. Indic.* **115**, 106412 (2020).
52. Campbell, B. M. *et al.* Reducing risks to food security from climate change. *Glob. Food Secur.* **11**, 34–43 (2016).
53. Wang, Y., Yuan, Z. & Tang, Y. Enhancing food security and environmental sustainability: A critical review of food loss and waste management. *Resour. Environ. Sustain.* **4**, 100023 (2021).
54. Arruda, E. H., Melatto, R. A. P. B., Levy, W. & Conti, D. de M. Circular economy: A brief literature review (2015–2020). *Sustain. Oper. Comput.* **2**, 79–86 (2021).
55. Seberini, A. Economic, social and environmental world impacts of food waste on society and Zero waste as a global approach to their elimination. *SHS Web Conf.* **74**, 03010 (2020).
56. *Global Food Losses and Food Waste: Extent, Causes and Prevention; Study Conducted for the International Congress Save Food! At Interpack 2011, [16 - 17 May], Düsseldorf, Germany*. (Food and Agriculture Organization of the United Nations, Rome, 2011).
57. Awasthi, S. K. *et al.* Changes in global trends in food waste composting: Research challenges and opportunities. *Bioresour. Technol.* **299**, 122555 (2020).
58. Environment, U. N. UNEP Food Waste Index Report 2021. *UNEP - UN Environment Programme* http://www.unep.org/resources/report/unep-food-waste-index-report-2021 (2021).
59. Zhu, J. *et al.* Cradle-to-grave emissions from food loss and waste represent half of total greenhouse gas emissions from food systems. *Nat. Food* **4**, 247–256 (2023).
60. Amicarelli, V. & Bux, C. Food waste measurement toward a fair, healthy and environmental-friendly food system: a critical review. *Br. Food J.* **123**, 2907–2935 (2021).
61. Jeswani, H. K., Figueroa-Torres, G. & Azapagic, A. The extent of food waste generation in the UK and its environmental impacts. *Sustain. Prod. Consum.* **26**, 532–547 (2021).
62. Torres-León, C. *et al.* Food Waste and Byproducts: An Opportunity to Minimize Malnutrition and Hunger in Developing Countries. *Front. Sustain. Food Syst.* **2**, (2018).
63. Haque, M. A., Liu, Z., Demilade, A. & Kumar, N. M. Assessing the Environmental Footprint of Distiller-Dried Grains with Soluble Diet as a Substitute for Standard Corn–Soybean for Swine Production in the United States of America. *Sustainability* **14**, 1161 (2022).
64. Chen, L. *et al.* Strategies to achieve a carbon neutral society: a review. *Environ. Chem. Lett.* (2022) doi:10.1007/s10311-022-01435-8.
65. Sekifuji, R. & Tateda, M. Study of the feasibility of a rice husk recycling scheme in Japan to produce silica fertilizer for rice plants. *Sustain. Environ. Res.* **29**, 11 (2019).
66. Müller, P. & Schmid, M. Intelligent Packaging in the Food Sector: A Brief Overview. 12 (2019).
67. Ishangulyyev, R., Kim, S. & Lee, S. H. Understanding Food Loss and Waste—Why Are We Losing and Wasting Food? *Foods* **8**, 297 (2019).
68. Ong, E. S. *et al.* Palm oil industrial wastes as a promising feedstock for biohydrogen production: A comprehensive review. *Environ. Pollut.* **291**, 118160 (2021).
69. Reguengo, L. M., Salgaço, M. K., Sivieri, K. & Maróstica Júnior, M. R. Agro-industrial by-products: Valuable sources of bioactive compounds. *Food Res. Int. Ott. Ont* **152**, 110871 (2022).
70. Kosseva, M. R. Processing of food wastes. *Adv. Food Nutr. Res.* **58**, 57–136 (2009).
71. Banerjee, S. & Arora, A. Sustainable bioprocess technologies for urban waste valorization. *Case Stud. Chem. Environ. Eng.* **4**, 100166 (2021).

72. Girotto, F., Alibardi, L. & Cossu, R. Food waste generation and industrial uses: A review. *Waste Manag.* **45**, 32–41 (2015).

73. Nath, P. C. *et al.* Valorization of Food Waste as Animal Feed: A Step towards Sustainable Food Waste Management and Circular Bioeconomy. *Animals* **13**, 1366 (2023).

74. Ojha, S., Bußler, S. & Schlüter, O. K. Food waste valorisation and circular economy concepts in insect production and processing. *Waste Manag.* **118**, 600–609 (2020).

75. Roy, P., Mohanty, A. K., Dick, P. & Misra, M. A Review on the Challenges and Choices for Food Waste Valorization: Environmental and Economic Impacts. *ACS Environ. Au* **3**, 58–75 (2023).

76. Redak, R. Introduction to and Importance of Insects. in *Forest Entomology and Pathology: Volume 1: Entomology* (eds. D. Allison, J., Paine, T. D., Slippers, B. & Wingfield, M. J.) 1–17 (Springer International Publishing, Cham, 2023). doi:10.1007/978-3-031-11553-0_1.

77. Stork, N. E. How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? *Annu. Rev. Entomol.* **63**, 31–45 (2018).

78. Rosenberg, Y. *et al.* The global biomass and number of terrestrial arthropods. *Sci. Adv.* **9**, eabq4049 (2023).

79. Beutel, R. G., Friedrich, F., Yang, X.-K. & Ge, S.-Q. *Insect Morphology and Phylogeny: A Textbook for Students of Entomology*. (De Gruyter, 2013). doi:10.1515/9783110264043.

80. Wiegmann, B. M., Kim, J. & Trautwein, M. D. Holometabolous insects (Holometabola). in *The Timetree of Life* (eds. Hedges, S. B. & Kumar, S.) 0 (Oxford University Press, 2009). doi:10.1093/oso/9780199535033.003.0031.

81. Belles, X. Chapter 5 - The holometabolan development. in *Insect Metamorphosis* (ed. Belles, X.) 71–104 (Academic Press, 2020). doi:10.1016/B978-0-12-813020-9.00005-3.

82. Siddiqui, S. A. *et al.* Bioconversion of organic waste by insects – A comprehensive review. *Process Saf. Environ. Prot.* **187**, 1–25 (2024).

83. Galante, E. & Marcos-Garcia, M. A. Decomposer Insects. in *Encyclopedia of Entomology* (ed. Capinera, J. L.) 1158–1169 (Springer Netherlands, Dordrecht, 2008). doi:10.1007/978-1-4020-6359-6_845.

84. Js, R. *et al.* A study to access the significant role of insects in decomposition and nutrient recycling. *Int. J. Adv. Biochem. Res.* **8**, 110–114 (2024).

85. DeBruyn, J. M., Keenan, S. W. & Taylor, L. S. From carrion to soil: microbial recycling of animal carcasses. *Trends Microbiol.* (2024) doi:10.1016/j.tim.2024.09.003.

86. Goulson, D. The insect apocalypse, and why it matters. *Curr. Biol.* **29**, R967–R971 (2019).

87. New, T. R. Consequences of Insect Declines and Loss. in *Insect Conservation in Australia: Why and How* (ed. New, T. R.) 69–80 (Springer Nature Switzerland, Cham, 2024). doi:10.1007/978-3-031-66631-5_4.

88. Sheppard, W. & Anelli, C. Insect Products: Apiculture, Sericulture, and Other Insect-Derived Commodities. in 101–124 (2024). doi:10.5040/9781474203821.ch-4.

89. Guichard, M., Phocas, F., Neuditschko, M., Basso, B. & Dainat, B. An Overview of Selection Concepts Applied to Honey Bees. *Bee World* **100**, 2–8 (2023).

90. Wang, M. *et al.* Genetic Mapping of Climbing and Mimicry: Two Behavioral Traits Degraded During Silkworm Domestication. *Front. Genet.* **11**, 566961 (2020).

91. Eisner, T., Nowicki, S., Goetz, M. & Meinwald, J. Red cochineal dye (carminic Acid): its role in nature. *Science* **208**, 1039–1042 (1980).

92. Bashir, N. H. *et al.* Unraveling the Role of Lac Insects in Providing Natural Industrial Products. *Insects* **13**, 1117 (2022).

93. Balu, R., Dutta, N. K., Dutta, A. K. & Choudhury, N. R. Resilin-mimetics as a smart biomaterial platform for biomedical applications. *Nat. Commun.* **12**, 149 (2021).

94. Bakhshandeh, B., Nateghi, S. S., Gazani, M. M., Dehghani, Z. & Mohammadzadeh, F. A review on advances in the applications of spider silk in biomedical issues. *Int. J. Biol. Macromol.* **192**, 258–271 (2021).

95. Lange, K. W. & Nakamura, Y. Edible insects as future food: chances and challenges. *J. Future Foods* **1**, 38–46 (2021).

96. Yang, L. H. & Gratton, C. Insects as drivers of ecosystem processes. *Curr. Opin. Insect Sci.* **2**, 26–32 (2014).

97. Stern, D. Body-size control: how an insect knows it has grown enough. *Curr. Biol. CB* **13**, R267-269 (2003).

98. Grunert, L. W., Clarke, J. W., Ahuja, C., Eswaran, H. & Nijhout, H. F. A Quantitative Analysis of Growth and Size Regulation in Manduca sexta: The Physiological Basis of Variation in Size and Age at Metamorphosis. *PLOS ONE* **10**, e0127988 (2015).

99. Čičková, H., Newton, G. L., Lacy, R. C. & Kozánek, M. The use of fly larvae for organic waste treatment. *Waste Manag.* **35**, 68–80 (2015).

100. Fowles, T. M. & Nansen, C. Insect-Based Bioconversion: Value from Food Waste. in *Food Waste Management: Solving the Wicked Problem* (eds. Närvänen, E., Mesiranta, N., Mattila, M. & Heikkinen, A.) 321–346 (Springer International Publishing, Cham, 2020). doi:10.1007/978-3-030-20561-4_12.

101. van Huis, A. Potential of insects as food and feed in assuring food security. *Annu. Rev. Entomol.* **58**, 563–583 (2013).

102. Defoliart, G. R. Edible insects as minilivestock. *Biodivers. Conserv.* **4**, 306–321 (1995).

103. Huis, A. van. *Edible Insects: Future Prospects for Food and Feed Security*. (Food and Agriculture Organization of the United Nations, Rome, 2013).

104. Mlcek, J., Rop, O., Borkovcova, M. & Bednarova, M. A Comprehensive Look at the Possibilities of Edible Insects as Food in Europe – A Review. *Pol. J. Food Nutr. Sci.* **64**, 147–157 (2014).

105. Ramos-Elorduy, J. Anthropo-entomophagy: Cultures, evolution and sustainability. *Entomol. Res.* **39**, 271–288 (2009).

106. Acosta-Estrada, B. A., Reyes, A., Rosell, C. M., Rodrigo, D. & Ibarra-Herrera, C. C. Benefits and Challenges in the Incorporation of Insects in Food Products. *Front. Nutr.* **8**, 687712 (2021).

107. Govorushko, S. Global status of insects as food and feed source: A review. *Trends Food Sci. Technol.* **91**, 436–445 (2019).

108. van Huis, A. Edible insects contributing to food security? *Agric. Food Secur.* **4**, 20 (2015).

109. Dzepe, D. *et al.* Adoption potential of black soldier fly (*Hermetia illucens* (L.), Diptera: Stratiomyidae) larvae composting technology among smallholder farmers in Greater Ahafo-Ano, Ashanti region of Ghana. *Green Technol. Sustain.* **2**, 100112 (2024).

110. Barragán-Fonseca, K. B., Cortés-Urquijo, J., Pineda-Mejía, J., Lagos-Sierra, D. & Dicke, M. Small-scale Black Soldier Fly-fish farming: a model with socioeconomic benefits. *Anim. Front.* **13**, 91–101 (2023).

111. Siddiqui, S. A. *et al.* Future opportunities for products derived from black soldier fly (BSF) treatment as animal feed and fertilizer - A systematic review. *Environ. Dev. Sustain.* (2024) doi:10.1007/s10668-024-04673-8.

112. Food and Agriculture Organization. *Zimbabwe Idai Recovery Project BSF Manual*. https://blogs.worldbank.org/en/africacan/green-protein-farming-black-soldier-fly-larvae-afe-zimbabwe (2024).

113. Tanga, C. & Dorothy, N. Black soldier fly production : farmers guide. (2022).

114. Bulinda, C. M., Gido, E. O., Kirscht, H. & Tanga, C. M. Gendered Awareness of Pig and Poultry Farmers on the Potential of Black Soldier Fly (Hermetia illucens) Farming in Kenya. *Sustainability* **15**, 3613 (2023).

115. Duquennoi, C. & Martinez, J. European Union's policymaking on sustainable waste management and circularity in agroecosystems: The potential for innovative interactions between science and decision-making. *Front. Sustain. Food Syst.* **6**, (2022).

116. Anuardo, R. G., Espuny, M., Costa, A. C. F. & Oliveira, O. J. Toward a cleaner and more sustainable world: A framework to develop and improve waste management through organizations, governments and academia. *Heliyon* **8**, e09225 (2022).

117. Tiwasing, P. *From Yuck to Yum: Unlocking the Potential of Edible Insects*. (SRUC's Rural Policy Centre, 2022).

118. Tiwasing, P. & Pate, L. Exploring edible insects as feed in the UK: current challenges and future prospects. (2024) doi:10.1163/23524588-00001264.

119. van Huis, A. Edible insects: Challenges and prospects. *Entomol. Res.* **52**, 161–177 (2022).

120. Chia, S. Y. *et al.* Threshold temperatures and thermal requirements of black soldier fly Hermetia illucens: Implications for mass production. *PLoS ONE* **13**, e0206097 (2018).

121. Schiemer, C., Halloran, A., Jespersen, K. & Kaukua, P. Marketing Insects: Superfood or Solution-Food? in *Edible Insects in Sustainable Food Systems* 213–236 (2018). doi:10.1007/978-3-319-74011-9_14.

122. Marshall, S., Woodley, N. & Hauser, M. The historical spread of the Black Soldier Fly, Hermetia illucens (L.) (Diptera, Stratiomyidae, Hermetiinae), and its establishment in Canada. in (2015).

123. Ståhls, G. *et al.* The puzzling mitochondrial phylogeography of the black soldier fly (Hermetia illucens), the commercially most important insect protein species. *BMC Evol. Biol.* **20**, 60 (2020).

124. Tomberlin, J. K. & Huis, A. van. Black soldier fly from pest to 'crown jewel' of the insects as feed industry: an historical perspective. (2020) doi:10.3920/JIFF2020.0003.

125. Furman, D. P., Young, R. D. & Catts, Paul. E. Hermetia illucens (Linnaeus) as a Factor in the Natural Control of Musca domestica Linnaeus. *J. Econ. Entomol.* **52**, 917–921 (1959).

126. Lord, W. D., Goff, M. L., Adkins, T. R. & Haskell, N. H. The black soldier fly Hermetia illucens (Diptera: Stratiomyidae) as a potential measure of human postmortem interval: observations and case histories. *J. Forensic Sci.* **39**, 215–222 (1994).

127. Oliveira, F. R., Doelle, K. & Smith, R. P. External Morphology of Hermetia illucens Stratiomyidae: Diptera (L.1758) Based on Electron Microscopy. *Annu. Res. Rev. Biol.* **9**, 1–10 (2016).

128. De Smet, J., Wynants, E., Cos, P. & Van Campenhout, L. Microbial Community Dynamics during Rearing of Black Soldier Fly Larvae (Hermetia illucens) and Impact on Exploitation Potential. *Appl. Environ. Microbiol.* **84**, e02722-17 (2018).

129. Munsch-Masset, P., Labrousse, C., Beaugeard, L. & Bressac, C. The reproductive tract of the black soldier fly (Hermetia illucens) is highly differentiated and suggests adaptations to sexual selection. *Entomol. Exp. Appl.* **171**, 857–866 (2023).

130. Alagappan, S. *et al.* Near Infrared Spectroscopy as a Traceability Tool to Monitor Black Soldier Fly Larvae (Hermetia illucens) Intended as Animal Feed. *Appl. Sci.* **12**, 8168 (2022).

131. Bonelli, M. *et al.* Black Soldier Fly Larvae Adapt to Different Food Substrates through Morphological and Functional Responses of the Midgut. *Int. J. Mol. Sci.* **21**, E4955 (2020).

132. Gligorescu, A., Toft, S., Hauggaard-Nielsen, H., Axelsen, J. A. & Nielsen, S. A. Development, growth and metabolic rate of Hermetia illucens larvae. *J. Appl. Entomol.* **143**, 875–881 (2019).

133. Shit, N. Black Soldier Fly (BSF): A Cost Effective Alternate Protein Source for Sustainable Livestock Production. *Res. Biot.* **3**, 128–132 (2021).

134. Sheppard, D. C., Tomberlin, J. K., Joyce, J. A., Kiser, B. C. & Sumner, S. M. Rearing methods for the black soldier fly (Diptera: Stratiomyidae). *J. Med. Entomol.* **39**, 695–698 (2002).

135. Gligorescu, A., Fischer, C. H., Larsen, P. F., Nørgaard, J. V. & Heckman, L.-H. L. Production and Optimization of Hermetia illucens (L.) Larvae Reared on Food Waste and Utilized as Feed Ingredient. *Sustainability* **12**, 9864 (2020).

136. Barragan-Fonseca, K. B., Dicke, M. & van Loon, J. J. A. Nutritional value of the black soldier fly (*Hermetia illucens* L.) and its suitability as animal feed – a review. *J. Insects Food Feed* **3**, 105–120 (2017).

137. Julita, U., Lusianti Fitri, L., Eka Putra, R. & Dana Permana, A. Mating Success and Reproductive Behavior of Black Soldier Fly Hermetia illucens L. (Diptera, Stratiomyidae) in Tropics. *J. Entomol.* **17**, 117–127 (2020).

138. Singh, N. K., Pandey, S., Sharma, H. & Goel, S. *Green Innovation, Sustainable Development, and Circular Economy*. (CRC Press, 2020).

139. Siva Raman, S., Stringer, L. C., Bruce, N. C. & Chong, C. S. Opportunities, challenges and solutions for black soldier fly larvae-based animal feed production. *J. Clean. Prod.* **373**, 133802 (2022).

140. Jannathulla, R. *et al.* Fishmeal availability in the scenarios of climate change: Inevitability of fishmeal replacement in aquafeeds and approaches for the utilization of plant protein sources. *Aquac. Res.* **50**, 3493–3506 (2019).

141. PubMed. Search for 'black soldier fly' OR 'Hermetia illucens' in title/abstract. (2024).

142. Scopus. Search for 'black soldier fly' OR 'Hermetia illucens' in title/abstract. (2024).

143. Lens.org. Search for 'Hermetia illucens' in patents and patent applications. (2024).

144. Meticulous Market Research Pvt. Ltd. *Global Black Soldier Fly Market Size, Share, Forecast, & Trends Analysis by Product, Application, and Geography - Forecast to 2033*. https://www.researchandmarkets.com/reports/4986109/global-black-soldier-fly-market-size-share (2024).

145. Makkar, H. P. S., Tran, G., Heuzé, V. & Ankers, P. State-of-the-art on use of insects as animal feed. *Anim. Feed Sci. Technol.* **197**, 1–33 (2014).

146. Suryati, T., Julaeha, E., Farabi, K., Ambarsari, H. & Hidayat, A. T. Lauric Acid from the Black Soldier Fly (Hermetia illucens) and Its Potential Applications. *Sustainability* **15**, 10383 (2023).

147. Dortmans, B., Diener, S., Verstappen, B. & Zurbrügg, C. *Black Soldier Fly Biowaste Processing - A Step-by-Step Guide*. (2017).

148. Diener, S., Studt Solano, N. M., Roa Gutiérrez, F., Zurbrügg, C. & Tockner, K. Biological Treatment of Municipal Organic Waste using Black Soldier Fly Larvae. *Waste Biomass Valorization* **2**, 357–363 (2011).

149. Du, S., Zhang, M., Zhang, K. & Wu, D. A Novel Integrated Continuous Process Was Developed for Treating Food Waste by Integrating Black Soldier Fly Larvae with Microorganisms. SSRN Scholarly Paper at https://doi.org/10.2139/ssrn.4793681 (2024).

150. Oonincx, D. G. A. B., Broekhoven, S. van, Huis, A. van & Loon, J. J. A. van. Feed Conversion, Survival and Development, and Composition of Four Insect Species on Diets Composed of Food By-Products. *PLOS ONE* **10**, e0144601 (2015).

151. Amrul, N. F. *et al.* A Review of Organic Waste Treatment Using Black Soldier Fly (Hermetia illucens). *Sustainability* **14**, 4565 (2022).

152. Silva, L. F. P., Brown, G. W. & James, P. J. Nutritional Value of Black Soldier Fly Grown on Abattoir Waste. *Proceedings* **36**, 216 (2020).

153. St-Hilaire, S. *et al.* Fish Offal Recycling by the Black Soldier Fly Produces a Foodstuff High in Omega-3 Fatty Acids. *J. World Aquac. Soc.* **38**, 309–313 (2007).

154. Liu, T. *et al.* Black soldier fly larvae for organic manure recycling and its potential for a circular bioeconomy: A review. *Sci. Total Environ.* **833**, 155122 (2022).

155. Arnone, S. *et al.* Black soldier fly (Hermetia illucens L.) as a high-potential agent for bioconversion of municipal primary sewage sludge. *Environ. Sci. Pollut. Res.* **29**, 64886–64901 (2022).

156. Kim, C.-H. *et al.* Use of Black Soldier Fly Larvae for Food Waste Treatment and Energy Production in Asian Countries: A Review. *Processes* **9**, 161 (2021).

157. Kim, W. *et al.* Biochemical characterization of digestive enzymes in the black soldier fly, *Hermetia illucens* (Diptera: Stratiomyidae). *J. Asia-Pac. Entomol.* **14**, 11–14 (2011).

158. Seyedalmoosavi, M. M., Mielenz, M., Veldkamp, T., Daş, G. & Metges, C. C. Growth efficiency, intestinal biology, and nutrient utilization and requirements of black soldier fly (Hermetia illucens) larvae compared to monogastric livestock species: a review. *J. Anim. Sci. Biotechnol.* **13**, 31 (2022).

159. Gold, M. *et al.* Novel Experimental Methods for the Investigation of Hermetia illucens (Diptera: Stratiomyidae) Larvae. *J. Insect Sci.* **20**, 21 (2020).

160. Eke, M. *et al.* Deciphering the functional diversity of the gut microbiota of the black soldier fly (Hermetia illucens): recent advances and future challenges. *Anim. Microbiome* **5**, 40 (2023).

161. Klammsteiner, T. *et al.* The Core Gut Microbiome of Black Soldier Fly (Hermetia illucens) Larvae Raised on Low-Bioburden Diets. *Front. Microbiol.* **11**, (2020).

162. Mannaa, M., Mansour, A., Park, I., Lee, D.-W. & Seo, Y.-S. Insect-based agri-food waste valorization: Agricultural applications and roles of insect gut microbiota. *Environ. Sci. Ecotechnology* **17**, 100287 (2024).

163. Ao, Y. *et al.* Characteristics and nutrient function of intestinal bacterial communities in black soldier fly (Hermetia illucens L.) larvae in livestock manure conversion. *Microb. Biotechnol.* **14**, 886–896 (2021).

164. Lee, K.-S., Yun, E.-Y. & Goo, T.-W. Antimicrobial Activity of an Extract of Hermetia illucens Larvae Immunized with Lactobacillus casei against Salmonella Species. *Insects* **11**, 704 (2020).

165. H, N. B. *et al.* Bioconversion and Performance of Black Soldier Fly (Hermetia illucens) in Converting Organic Poultry Waste Materials into High Value Products and Fertilizers. *Waste Biomass Valorization* **15**, 5559–5572 (2024).

166. Wilkinson, J. M. Re-defining efficiency of feed use by livestock. *animal* **5**, 1014–1022 (2011).

167. Lalander, C., Diener, S., Zurbrügg, C. & Vinnerås, B. Effects of feedstock on larval development and process efficiency in waste treatment with black soldier fly (*Hermetia illucens*). *J. Clean. Prod.* **208**, 211–219 (2019).

168. Eriksen, N. T. Metabolic performance and feed efficiency of black soldier fly larvae. *Front. Bioeng. Biotechnol.* **12**, 1397108 (2024).

169. Laganaro, M., Bahrndorff, S. & Eriksen, N. T. Growth and metabolic performance of black soldier fly larvae grown on low and high-quality substrates. *Waste Manag.* **121**, 198–205 (2021).

170. Lopez, G. & Leeson, S. Review: Energy partitioning in broiler chickens. *Can. J. Anim. Sci.* **88**, 205–212 (2008).

171. Lange, K. W. & Nakamura, Y. Potential contribution of edible insects to sustainable consumption and production. *Front. Sustain.* **4**, (2023).

172. da-Silva, W. C. *et al.* Nutritional Value of the Larvae of the Black Soldier Fly (Hermetia illucens) and the House Fly (Musca domestica) as a Food Alternative for Farm Animals—A Systematic Review. *Insects* **15**, 619 (2024).

173. Smil, V. Worldwide transformation of diets, burdens of meat production and opportunities for novel food proteins. *Enzyme Microb. Technol.* **30**, 305–311 (2002).

174. Lamichhane, J. R., Varaillas, C. & Debaeke, P. Seedling emergence and biomass production of soybean cultivars under wheat-soybean relay cropping. *PLOS ONE* **18**, e0293671 (2023).

175. Hadj Saadoun, J. *et al.* Lipid profile and growth of black soldier flies (Hermetia illucens, Stratiomyidae) reared on by-products from different food chains. *J. Sci. Food Agric.* **100**, 3648–3657 (2020).

176. Shumo, M. *et al.* The nutritive value of black soldier fly larvae reared on common organic waste streams in Kenya. *Sci. Rep.* **9**, 10110 (2019).

177. Newton, G. L. *et al.* Using the black soldier fly, Hermetia illucens, as a value-added tool for the management of swine manure. (2005).

178. Belhadj Slimen, I., Yerou, H., Ben Larbi, M., M'Hamdi, N. & Najar, T. Insects as an alternative protein source for poultry nutrition: a review. *Front. Vet. Sci.* **10**, (2023).

179. Li, X. *et al.* Growth and Fatty Acid Composition of Black Soldier Fly Hermetia illucens (Diptera: Stratiomyidae) Larvae Are Influenced by Dietary Fat Sources and Levels. *Anim. Open Access J. MDPI* **12**, 486 (2022).

180. Benzertiha, A. *et al.* Insect Oil as An Alternative to Palm Oil and Poultry Fat in Broiler Chicken Nutrition. *Anim. Open Access J. MDPI* **9**, 116 (2019).

181. Hender, A., Siddik, M. A., Howieson, J. & Fotedar, R. Black Soldier Fly, Hermetia illucens as an Alternative to Fishmeal Protein and Fish Oil: Impact on Growth, Immune Response, Mucosal Barrier Status, and Flesh Quality of Juvenile Barramundi, Lates calcarifer (Bloch, 1790). *Biology* **10**, 505 (2021).

182. Erbland, P., Alyokhin, A., Perkins, L. B. & Peterson, M. Dose-Dependent Retention of Omega-3 Fatty Acids by Black Soldier Fly Larvae (Diptera: Stratiomyidae). *J. Econ. Entomol.* **113**, 1221–1226 (2020).

183. Zulkifli, N. F. N. M. *et al.* Nutritional value of black soldier fly (Hermetia illucens) larvae processed by different methods. *PLoS ONE* **17**, e0263924 (2022).

184. Borland, M. *et al.* Growth performance, hepatic gene expression, and plasma biochemistry of rainbow trout fed full-fat meal, defatted meal, oil and chitin from black soldier flies. *Comp. Immunol. Rep.* **6**, 200149 (2024).

185. Lu, S. *et al.* Nutritional Composition of Black Soldier Fly Larvae (Hermetia illucens L.) and Its Potential Uses as Alternative Protein Sources in Animal Diets: A Review. *Insects* **13**, 831 (2022).

186. Zhao, J. *et al.* Egg quality and laying performance of Julia laying hens fed with black soldier fly (*Hermetia illucens*) larvae meal as a long-term substitute for fish meal. *Poult. Sci.* **101**, 101986 (2022).

187. Salahuddin, M. *et al.* Flight toward Sustainability in Poultry Nutrition with Black Soldier Fly Larvae. *Animals* **14**, 510 (2024).

188. Heuel, M. *et al.* Black soldier fly larvae meal and fat can completely replace soybean cake and oil in diets for laying hens. *Poult. Sci.* **100**, 101034 (2021).

189. Nekrasov, R. V. *et al.* Effect of Black Soldier Fly (Hermetia illucens L.) Fat on Health and Productivity Performance of Dairy Cows. *Animals* **12**, 2118 (2022).

190. Bruni, L. *et al.* Total replacement of dietary fish meal with black soldier fly (Hermetia illucens) larvae does not impair physical, chemical or volatile composition of farmed Atlantic salmon (Salmo salar L.). *J. Sci. Food Agric.* **100**, 1038–1047 (2020).

191. Chia, S. Y. *et al.* Effect of Dietary Replacement of Fishmeal by Insect Meal on Growth Performance, Blood Profiles and Economics of Growing Pigs in Kenya. *Animals* **9**, 705 (2019).

192. Nairuti, R. N., Musyoka, S. N., Yegon, M. J. & Opiyo, M. A. Utilization of Black Soldier Fly (Hermetia illucens Linnaeus) Larvae as a Protein Source for Fish Feed – a Review. *Aquac. Stud.* **22**, (2021).

193. Eide, L. H. *et al.* Black soldier fly larvae (*Hermetia illucens*) meal is a viable protein source for Atlantic salmon (*Salmo salar*) during a large-scale controlled field trial under commercial-like conditions. *Aquaculture* **579**, 740194 (2024).

194. Fiorilla, E. *et al.* Dehydrated and live black soldier fly larvae as environmental enrichment in indigenous slow-growing chickens: performance, gut health, and chitinolytic enzyme activity. *animal* **18**, 101239 (2024).

195. Biasato, I. *et al.* Welfare implications for broiler chickens reared in an insect larvae-enriched environment: Focus on bird behaviour, plumage status, leg health, and excreta corticosterone. *Front. Physiol.* **13**, 930158 (2022).

196. Gariglio, M. *et al.* Does the provision of live black soldier fly and yellow mealworm larvae improve Muscovy duck welfare? *J. Anim. Sci. Biotechnol.* **14**, 153 (2023).

197. Ipema, A. F., Bokkers, E. A. M., Gerrits, W. J. J., Kemp, B. & Bolhuis, J. E. Providing live black soldier fly larvae (Hermetia illucens) improves welfare while maintaining performance of piglets post-weaning. *Sci. Rep.* **11**, 7371 (2021).

198. Devic, E., Leschen, W., Murray, F. & Little, D. c. Growth performance, feed utilization and body composition of advanced nursing Nile tilapia (Oreochromis niloticus) fed diets containing Black Soldier Fly (Hermetia illucens) larvae meal. *Aquac. Nutr.* **24**, 416–423 (2018).

199. Koutsos, E., Modica, B. & Freel, T. Immunomodulatory potential of black soldier fly larvae: applications beyond nutrition in animal feeding programs. *Transl. Anim. Sci.* **6**, txac084 (2022).

200. Nakagawa, A., Sakamoto, T., Kanost, M. R. & Tabunoki, H. The Development of New Methods to Stimulate the Production of Antimicrobial Peptides in the Larvae of the Black Soldier Fly Hermetia illucens. *Int. J. Mol. Sci.* **24**, 15765 (2023).

201. Vogel, H., Müller, A., Heckel, D. G., Gutzeit, H. & Vilcinskas, A. Nutritional immunology: Diversification and diet-dependent expression of antimicrobial peptides in the black soldier fly Hermetia illucens. *Dev. Comp. Immunol.* **78**, 141–148 (2018).

202. Van Moll, L. *et al.* In Vitro Evaluation of Antimicrobial Peptides from the Black Soldier Fly (Hermetia Illucens) against a Selection of Human Pathogens. *Microbiol. Spectr.* **10**, e0166421 (2022).

203. Xu, J. *et al.* Transgenic expression of antimicrobial peptides from black soldier fly enhance resistance against entomopathogenic bacteria in the silkworm, Bombyx mori. *Insect Biochem. Mol. Biol.* **127**, 103487 (2020).

204. Komi, D. E. A., Sharma, L. & Cruz, C. S. D. Chitin and Its Effects on Inflammatory and Immune Responses. *Clin. Rev. Allergy Immunol.* **54**, 213 (2018).

205. Dörper, A. *et al.* Effects of different black soldier fly larvae products on slow-growing broiler performance and carcass characteristics. *Poult. Sci.* **103**, 103481 (2024).

206. Abdel-Latif, H. M. R. *et al.* Black soldier fly (*Hermetia illucens*) larvae meal in diets of European seabass: Effects on antioxidative capacity, non-specific immunity, transcriptomic responses, and resistance to the challenge with *Vibrio alginolyticus*. *Fish Shellfish Immunol.* **111**, 111–118 (2021).

207. Yu, M. *et al.* Hermetia illucens larvae as a potential dietary protein source altered the microbiota and modulated mucosal immune status in the colon of finishing pigs. *J. Anim. Sci. Biotechnol.* **10**, 50 (2019).

208. Schlüter, O. *et al.* Safety aspects of the production of foods and food ingredients from insects. *Mol. Nutr. Food Res.* **61**, (2017).

209. Elhag, O. *et al.* Inhibition of Zoonotic Pathogens Naturally Found in Pig Manure by Black Soldier Fly Larvae and Their Intestine Bacteria. *Insects* **13**, 66 (2022).

210. Shi, C. *et al.* Inhibition of pathogenic microorganisms in solid organic waste via black soldier fly larvae-mediated management. *Sci. Total Environ.* **913**, 169767 (2024).

211. Müller, A., Wiedmer, S. & Kurth, M. Risk Evaluation of Passive Transmission of Animal Parasites by Feeding of Black Soldier Fly (*Hermetia illucens*) Larvae and Prepupae. *J. Food Prot.* **82**, 948–954 (2019).

212. Abd Manan, F., Yeoh, Y.-K., Chai, T.-T. & Wong, F.-C. Unlocking the potential of black soldier fly frass as a sustainable organic fertilizer: A review of recent studies. *J. Environ. Manage.* **367**, 121997 (2024).

213. Quilliam, R. S. *et al.* Integrating insect frass biofertilisers into sustainable peri-urban agro-food systems. (2020) doi:10.3920/JIFF2019.0049.

214. Lopes, I. G., Yong, J. W. & Lalander, C. Frass derived from black soldier fly larvae treatment of biodegradable wastes. A critical review and future perspectives. *Waste Manag.* **142**, 65–76 (2022).

215. Saberi Riseh, R., Gholizadeh Vazvani, M., Vatankhah, M. & Kennedy, J. F. Chitin-induced disease resistance in plants: A review. *Int. J. Biol. Macromol.* **266**, 131105 (2024).

216. Basri, N. E. A. *et al.* Potential Applications of Frass Derived from Black Soldier Fly Larvae Treatment of Food Waste: A Review. *Foods* **11**, 2664 (2022).

217. Arabzadeh, G. *et al.* Suppressive Effect of Black Soldier Fly Larvae Frass on Fusarium Wilt Disease in Tomato Plants. *Insects* **15**, 613 (2024).

218. Lomonaco, G. *et al.* Larval Frass of Hermetia illucens as Organic Fertilizer: Composition and Beneficial Effects on Different Crops. *Insects* **15**, 293 (2024).

219. VICKERSON, A., Radley, R., MARCHANT, B., KAULFUSS, O. & KABALUK, T. Hermetia illucens frass production and use in plant nutrition and pest management. (2015).

220. Kisaakye, J. *et al.* Chitin-enriched insect frass fertilizer as a biorational alternative for root-knot nematode (Meloidogyne incognita) management. *Front. Plant Sci.* **15**, (2024).

221. Gebremikael, M. T., Wickeren, N. van, Hosseini, P. S. & De Neve, S. The Impacts of Black Soldier Fly Frass on Nitrogen Availability, Microbial Activities, C Sequestration, and Plant Growth. *Front. Sustain. Food Syst.* **6**, (2022).

222. Mertenat, A., Diener, S. & Zurbrügg, C. Black Soldier Fly biowaste treatment – Assessment of global warming potential. *Waste Manag.* **84**, 173–181 (2019).

223. Beesigamukama, D. *et al.* Low-cost technology for recycling agro-industrial waste into nutrient-rich organic fertilizer using black soldier fly. *Waste Manag.* **119**, 183–194 (2021).

224. Chavez, M. & Uchanski, M. Insect left-over substrate as plant fertiliser. (2021) doi:10.3920/JIFF2020.0063.

225. Setti, L. *et al.* Use of black soldier fly (Hermetia illucens (L.), Diptera: Stratiomyidae) larvae processing residue in peat-based growing media. *Waste Manag.* **95**, 278–288 (2019).

226. Bose, I., Singh, R., Negi, P. & Singh, Y. Chitin as bio-based nanomaterial in packaging: A review. *Mater. Today Proc.* **46**, 11254–11257 (2021).

227. Amiri, H. *et al.* Chitin and chitosan derived from crustacean waste valorization streams can support food systems and the UN Sustainable Development Goals. *Nat. Food* **3**, 822–828 (2022).

228. Siddiqui, S. A. *et al.* The potential of chitin and chitosan from dead black soldier fly (BSF) (Hermetia illucens) for biodegradable packaging material – A critical review. *Process Saf. Environ. Prot.* **189**, 1342–1367 (2024).

229. Hu, Z. & Gänzle, M. G. Challenges and opportunities related to the use of chitosan as a food preservative. *J. Appl. Microbiol.* **126**, 1318–1331 (2019).

230. Zhou, Y., He, Y., Lin, X., Feng, Y. & Liu, M. Sustainable, High-Performance, and Biodegradable Plastics Made from Chitin. *ACS Appl. Mater. Interfaces* **14**, 46980–46993 (2022).

231. Bhavsar, P. S., Dalla Fontana, G. & Zoccola, M. Sustainable Superheated Water Hydrolysis of Black Soldier Fly Exuviae for Chitin Extraction and Use of the Obtained Chitosan in the Textile Field. *ACS Omega* **6**, 8884–8893 (2021).

232. Baharlouei, P. & Rahman, A. Chitin and Chitosan: Prospective Biomedical Applications in Drug Delivery, Cancer Treatment, and Wound Healing. *Mar. Drugs* **20**, 460 (2022).

233. Islam, Md. M., Shahruzzaman, Md., Biswas, S., Nurus Sakib, Md. & Rashid, T. U. Chitosan based bioactive materials in tissue engineering applications-A review. *Bioact. Mater.* **5**, 164–183 (2020).

234. Gassara, F. *et al.* Chitin and chitosan as natural flocculants for beer clarification. *J. Food Eng.* **166**, 80–85 (2015).

235. Al-Manhel, A. J., Al-Hilphy, A. R. S. & Niamah, A. K. Extraction of chitosan, characterisation and its use for water purification. *J. Saudi Soc. Agric. Sci.* **17**, 186–190 (2018).

236. Atheena, P. V., Basawa, R. & Raval, R. Advancing wastewater treatment: chitin and derivatives for PPCP contaminant mitigation. *Polym. Bull.* **81**, 14307–14336 (2024).

237. Richter, H., Gover, O. & Schwartz, B. Anti-Inflammatory Activity of Black Soldier Fly Oil Associated with Modulation of TLR Signaling: A Metabolomic Approach. *Int. J. Mol. Sci.* **24**, 10634 (2023).

238. Borrelli, L. *et al.* Insect Derived Lauric Acid as Promising Alternative Strategy to Antibiotics in the Antimicrobial Resistance Scenario. *Front. Microbiol.* **12**, (2021).

239. Leni, G. *et al.* Black Soldier Fly Larvae Grown on Hemp Fiber: Nutritional Composition and Production of Potential Bioactive Peptides. *Macromol* **4**, 135–149 (2024).

240. Matos, F. M. de, Boscariol Rasera, G. & Castro, R. J. S. de. Insects as a sustainable source of emerging proteins and their processing to obtain bioactive compounds: an updated review. *Sustain. Food Technol.* **2**, 19–31 (2024).

241. Riolo, K. *et al.* Cytoprotective and Antioxidant Effects of Hydrolysates from Black Soldier Fly (Hermetia illucens). *Antioxidants* **12**, 519 (2023).

242. Almeida, C. *et al.* Characterization of lipid extracts from the *Hermetia illucens* larvae and their bioactivities for potential use as pharmaceutical and cosmetic ingredients. *Heliyon* **8**, e09455 (2022).

243. Phongpradist, R. *et al.* The in vitro effects of black soldier fly larvae (Hermitia illucens) oil as a high-functional active ingredient for inhibiting hyaluronidase, anti-oxidation benefits, whitening, and UVB protection. *Front. Pharmacol.* **14**, 1243961 (2023).

244. Lai-Foenander, A. S. *et al.* Black soldier fly (Hermetia illucens L.): A potential small mighty giant in the field of cosmeceuticals. *Health Sci. Rep.* **7**, e2120 (2024).

245. Xiong, J. *et al.* Refining and Sulfurization of Oil from Black Soldier Fly and Its Application as Biodegradable Lubricant Additive. *J. Am. Oil Chem. Soc.* **97**, 1243–1251 (2020).

246. Verheyen, G. R. *et al.* Glycine-acyl Surfactants Prepared from Black Soldier Fly Fat, Coconut Oil and Palm Kernel Oil. *Curr. Green Chem.* **7**, 239–248 (2020).

247. Gauto, M. A. *et al.* Hybrid vigor: Why hybrids with sustainable biofuels are better than pure electric vehicles. *Energy Sustain. Dev.* **76**, 101261 (2023).

248. Park, J.-Y. *et al.* Biodiesel production from the black soldier fly larvae grown on food waste and its fuel property characterization as a potential transportation fuel. *Environ. Eng. Res.* **27**, (2022).

249. Jung, S. *et al.* Biodiesel production from black soldier fly larvae derived from food waste by non-catalytic transesterification. *Energy* **238**, 121700 (2022).

250. Feng, W. *et al.* Energy consumption analysis of lipid extraction from black soldier fly biomass. *Energy* **185**, 1076–1085 (2019).

251. Mohan, K., Sathishkumar, P., Rajan, D. K., Rajarajeswaran, J. & Ganesan, A. R. Black soldier fly (Hermetia illucens) larvae as potential feedstock for the biodiesel production: Recent advances and challenges. *Sci. Total Environ.* **859**, 160235 (2023).

252. Aguilar-Murguía, D. M., Lara-Montaño, O. D., Gutiérrez-Antonio, C., Ponce-Ortega, J. M. & Martínez-Guido, S. I. Optimal supply chain design of biodiesel production using black soldier fly larvae oil as feedstock. *Int. J. Green Energy* **21**, 3252–3269 (2024).

253. Zhan, S. *et al.* Genomic landscape and genetic manipulation of the black soldier fly Hermetia illucens, a natural waste recycler. *Cell Res.* **30**, 50–60 (2020).

254. Meijer, N. *et al.* Aflatoxin B1 Conversion by Black Soldier Fly (Hermetia illucens) Larval Enzyme Extracts. *Toxins* **11**, 532 (2019).

255. Meijer, N., de Rijk, T., van Loon, J. J. A., Zoet, L. & van der Fels-Klerx, H. J. Effects of insecticides on mortality, growth and bioaccumulation in black soldier fly (Hermetia illucens) larvae. *PloS One* **16**, e0249362 (2021).

256. Cai, M. *et al.* Systematic characterization and proposed pathway of tetracycline degradation in solid waste treatment by *Hermetia illucens* with intestinal microbiota. *Environ. Pollut.* **242**, 634–642 (2018).

257. Yang, C. *et al.* Characteristics and mechanisms of ciprofloxacin degradation by black soldier fly larvae combined with associated intestinal microorganisms. *Sci. Total Environ.* **811**, 151371 (2022).

258. Lalander, C. *et al.* Fate of pharmaceuticals and pesticides in fly larvae composting. *Sci. Total Environ.* **565**, 279–286 (2016).

259. Cai, T., Xia, M., Yuan, W., Ming, Y. & Zhang, Q. Bioconversion of sulfamethazine-contaminated chicken manure by black soldier fly larvae: Effects on antibiotic resistance genes and microbial communities. *J. Environ. Manage.* **371**, 123206 (2024).

260. De Filippis, F. *et al.* Plastics shape the black soldier fly larvae gut microbiome and select for biodegrading functions. *Microbiome* **11**, 205 (2023).

261. Yang, S.-S. *et al.* Impacts of physical-chemical property of polyethylene on depolymerization and biodegradation in yellow and dark mealworms with high purity microplastics. *Sci. Total Environ.* **828**, 154458 (2022).

262. Yang, Y. *et al.* Biodegradation and Mineralization of Polystyrene by Plastic-Eating Mealworms: Part 1. Chemical and Physical Characterization and Isotopic Tests. *Environ. Sci. Technol.* **49**, 12080–12086 (2015).

263. Liu, J. *et al.* Biodegradation of polyether-polyurethane foam in yellow mealworms (Tenebrio molitor) and effects on the gut microbiome. *Chemosphere* **304**, 135263 (2022).

264. Tepper, K., Edwards, O., Sunna, A., Paulsen, I. T. & Maselko, M. Diverting organic waste from landfills via insect biomanufacturing using engineered black soldier flies (Hermetia illucens). *Commun. Biol.* **7**, 1–14 (2024).

265. Fan, M., Liu, N., Wu, X., Zhang, J. & Cai, M. Tolerance and Removal of Four Polycyclic Aromatic Hydrocarbon Compounds (PAHs) by Black Soldier Fly (Diptera: Stratiomyidae). *Environ. Entomol.* **49**, 667–672 (2020).

266. Fels-Klerx, H. J. van der, Meijer, N., Nijkamp, M. M., Schmitt, E. & Loon, J. J. A. van. Chemical food safety of using former foodstuffs for rearing black soldier fly larvae (Hermetia illucens) for feed and food use. (2020) doi:10.3920/JIFF2020.0024.

267. Siddiqui, S. A. *et al.* Effects of undesired substances and their bioaccumulation on the black soldier fly larvae, Hermetia illucens (Diptera: Stratiomyidae)-a literature review. *Environ. Monit. Assess.* **195**, 823 (2023).

268. Biancarosa, I. *et al.* Uptake of heavy metals and arsenic in black soldier fly (Hermetia illucens) larvae grown on seaweed-enriched media. *J. Sci. Food Agric.* **98**, 2176–2183 (2018).

269. Wu, N., Wang, X., Xu, X., Cai, R. & Xie, S. Effects of heavy metals on the bioaccumulation, excretion and gut microbiome of black soldier fly larvae (*Hermetia illucens*). *Ecotoxicol. Environ. Saf.* **192**, 110323 (2020).

270. McDonald, S. C. *et al.* Selecting recombinants to stack high protein with high oleic acid and low linoleic acid in soybean (). *Plant Breed.* **142**, 477–488 (2023).

271. Pope, M. *et al.* Quantifying the value of soybean meal in poultry and swine diets. *J. Appl. Poult. Res.* **32**, 100337 (2023).

272. Cashion, T., Le Manach, F., Zeller, D. & Pauly, D. Most fish destined for fishmeal production are food-grade fish. *Fish Fish.* **18**, 837–844 (2017).

273. Bell, M. & Davies, S. *Nutritional Value of Extracted Soybean, Full Fat Soya and Maize Gluten Meal Proteins in Comparison to a Fishmeal-Based Reference Diet for Fingerling Tilapia (Oreochromis Mossambicus)*. (2023). doi:10.1101/2023.11.12.566624.

274. Galán-Díaz, J. J., Pena-Mosquera, L., Puertas-Agudo, J. & Rodríguez, J. Carbon and water footprint assessment of the production cycle of the black soldier fly (Hermetia illucens) on a farm in Spain. *Environ. Dev.* **51**, 101038 (2024).

275. Lee, N.-H., Song, S.-H., Jung, M.-J., Ran Hui, K. & Park, J.-K. Dynamic emissions of N2O from solid waste landfills: A review. *Environ. Eng. Res.* **28**, (2023).

276. Oonincx, D. G. A. B. *et al.* An Exploration on Greenhouse Gas and Ammonia Production by Insect Species Suitable for Animal or Human Consumption. *PLOS ONE* **5**, e14445 (2010).

277. Tabassum-Abbasi, Abbasi, T. & Abbasi, S. A. Reducing the global environmental impact of livestock production: the minilivestock option. *J. Clean. Prod.* **112**, 1754–1766 (2016).

278. Rehman, K. ur *et al.* Black soldier fly, Hermetia illucens as a potential innovative and environmentally friendly tool for organic waste management: A mini-review. *Waste Manag. Res.* **41**, 81–97 (2023).

279. Salomone, R. *et al.* Environmental impact of food waste bioconversion by insects: Application of Life Cycle Assessment to process using Hermetia illucens. *J. Clean. Prod.* **140**, 890–905 (2017).

280. Food and Agriculture Organization. Production: Crops and livestock products (2023). (2023).

281. Food and Agriculture Organization. Land, Inputs and Sustainability: Land Use (2024). http://www.fao.org/faostat/en/#data/RL (2024).

282. Giannetti, D. *et al.* Do It by Yourself: Larval Locomotion in the Black Soldier Fly Hermetia illucens, with a Novel “Self-Harvesting” Method to Separate Prepupae. *Insects* **13**, 127 (2022).

283. Miranda, C. D., Cammack, J. A. & Tomberlin, J. K. Interspecific Competition between the House Fly, Musca domestica L. (Diptera: Muscidae) and Black Soldier Fly, Hermetia illucens (L.) (Diptera: Stratiomyidae) When Reared on Poultry Manure. *Insects* **10**, E440 (2019).

284. Erickson, M. C., ISLAM, M., SHEPPARD, C., LIAO, J. & DOYLE, M. P. Reduction of Escherichia coli O157:H7 and Salmonella enterica Serovar Enteritidis in Chicken Manure by Larvae of the Black Soldier Fly. *J. Food Prot.* **67**, 685–690 (2004).

285. Beskin, K. V. *et al.* Larval digestion of different manure types by the black soldier fly (Diptera: Stratiomyidae) impacts associated volatile emissions. *Waste Manag.* **74**, 213–220 (2018).

286. Singh, A. & Kumari, K. An inclusive approach for organic waste treatment and valorisation using Black Soldier Fly larvae: A review. *J. Environ. Manage.* **251**, 109569 (2019).

287. Khairuddin, D., Hassan, S. N. F. & Ghafar, S. N. A. Development of a small scale BSF rearing system and firsthand experience of the process and its lifecycle. *IOP Conf. Ser. Earth Environ. Sci.* **1019**, 012033 (2022).

288. Law, Y. & Wein, L. Reversing the nutrient drain through urban insect farming—opportunities and challenges. **5**, 226–237 (2018).

289. Grau, M. g. p., Dortmans, B. m. a., Egger, J., Virard, G. & Zurbrügg, C. Modelling the financial viability of centralised and decentralised black soldier fly larvae waste processing units in Surabaya, Indonesia. *J. Insects Food Feed* **9**, 303–316 (2023).

290. Abro, Z., Kassie, M., Tanga, C., Beesigamukama, D. & Diiro, G. Socio-economic and environmental implications of replacing conventional poultry feed with insect-based feed in Kenya. *J. Clean. Prod.* **265**, 121871 (2020).

291. Vo, V. Development of insect production automation: Automated processes for the production of Black Soldier Fly (Hermetia illucens). (2020).

292. Ko, H. *et al.* Air-Fluidized Aggregates of Black Soldier fly Larvae. *Front. Phys.* **9**, (2021).

293. Pahmeyer, M. J. *et al.* An automated, modular system for organic waste utilization using *Hermetia illucens* larvae: Design, sustainability, and economics. *J. Clean. Prod.* **379**, 134727 (2022).

294. Amoshie, D. A., Nyazenga, M. D. & Rosca, E. V. Automated Black Soldier Fly Incubator using Internet-of-Things and Computer Vision. in *2024 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA)* 1–6 (2024). doi:10.1109/ACDSA59508.2024.10467464.

295. Suckling, J., Druckman, A., Small, R., Cecelja, F. & Bussemaker, M. Supply chain optimization and analysis of *Hermetia illucens* (black soldier fly) bioconversion of surplus foodstuffs. *J. Clean. Prod.* **321**, 128711 (2021).

296. Avila, K. U., Campbell, M., Mauck, K., Gebiola, M. & Karydis, K. Development and Testing of a Smart Bin toward Automated Rearing of Black Soldier Fly Larvae. Preprint at https://doi.org/10.48550/arXiv.2207.04335 (2022).

297. Żuk-Gołaszewska, K. *et al.* Edible Insect Farming in the Context of the EU Regulations and Marketing—An Overview. *Insects* **13**, 446 (2022).

298. Tim Black, Rachel Keyserlingk, & Ludovico Lattanzi. *Delivering Insect-Based Feed at Scale*. https://www.frontier-economics.com/uk/en/news-and-insights/news/news-article-i20624-delivering-insect-based-feed-at-scale/ (2023).

299. Beyhan de Jong & Gorjan Nikolik. *No Longer Crawling: Insect Protein to Come of Age in the 2020s*. https://www.rabobank.com/knowledge/q011329132-no-longer-crawling-insect-protein-to-come-of-age-in-the-2020s (2021).

300. Meijer, N., Van Raamsdonk, L. W. D., Gerrits, E. W. J. & Appel, M. J. The use of animal by-products in a circular bioeconomy: Time for a TSE road map 3? *Heliyon* **9**, e14021 (2023).

301. Linh, N. V. *et al.* Black Soldier Fly (Hermetia illucens) Larvae Meal: A Sustainable Alternative to Fish Meal Proven to Promote Growth and Immunity in Koi Carp (Cyprinus carpio var. koi). *Fishes* **9**, 53 (2024).

302. Takakuwa, F. *et al.* Availability of black soldier fly meal as an alternative protein source to fish meal in red sea bream (Pagrus major, Temminck & Schlegel) fingerling diets. *Aquac. Res.* **53**, 36–49 (2022).

303. Richardson, A., Dantas-Lima, J., Lefranc, M. & Walraven, M. Effect of a Black Soldier Fly Ingredient on the Growth Performance and Disease Resistance of Juvenile Pacific White Shrimp (Litopenaeus vannamei). *Anim. Open Access J. MDPI* **11**, 1450 (2021).

304. Kotob, G. *et al.* Potential application of black soldier fly fats in canine and feline diet formulations: A review of literature. *J. Asia-Pac. Entomol.* **25**, 101994 (2022).

305. Dalle Zotte, A. *et al.* Incorporation of Black Soldier Fly (*Hermetia illucens* L.) larvae fat or extruded linseed in diets of growing rabbits and their effects on meat quality traits including detailed fatty acid composition. *Meat Sci.* **146**, 50–58 (2018).

306. Boykin, K. L. *et al.* Digestibility of black soldier fly larvae (Hermetia illucens) fed to leopard geckos (Eublepharis macularius). *PLoS ONE* **15**, e0232496 (2020).

307. Olivadese, M. & Dindo, M. L. Edible Insects: A Historical and Cultural Perspective on Entomophagy with a Focus on Western Societies. *Insects* **14**, 690 (2023).

308. Mancini, S. *et al.* Exploring the Future of Edible Insects in Europe. *Foods Basel Switz.* **11**, 455 (2022).

309. Bessa, L. w., Pieterse, E., Marais, J. & Hoffman, L. c. Black soldier fly larvae (Hermetia illucens) as a meat replacer in a burger patty. *J. Insects Food Feed* **9**, 1211–1222 (2023).

310. Bessa, L. W., Pieterse, E., Sigge, G. & Hoffman, L. C. An Exploratory Study into the Use of Black Soldier Fly (Hermetia illucens) Larvae in the Production of a Vienna-Style Sausage. *Meat Muscle Biol.* **3**, (2019).

311. Bußler, S., Rumpold, B. A., Jander, E., Rawel, H. M. & Schlüter, O. K. Recovery and techno-functionality of flours and proteins from two edible insect species: Meal worm (Tenebrio molitor) and black soldier fly (Hermetia illucens) larvae. *Heliyon* **2**, (2016).

312. Delicato, C., Schouteten, J. J., Dewettinck, K., Gellynck, X. & Tzompa-Sosa, D. A. Consumers' perception of bakery products with insect fat as partial butter replacement. *Food Qual. Prefer.* **79**, 103755 (2020).

313. Delfino, D. *et al.* Allergenicity of tropomyosin variants identified in the edible insect Hermetia illucens (black soldier fly). *Food Chem.* **437**, 137849 (2023).

314. de Gier, S. & Verhoeckx, K. Insect (food) allergy and allergens. *Mol. Immunol.* **100**, 82–106 (2018).

315. Adli, D. Uses insects in poultry feed as replacement soya bean meal and fish meal in development countries: a systematic review. *Livest. Res. Rural Dev.* **33**, 128 (2021).

316. Dreoni, I., Matthews, Z. & Schaafsma, M. The impacts of soy production on multi-dimensional well-being and ecosystem services: A systematic review. *J. Clean. Prod.* **335**, 130182 (2022).

317. Nagappan, S. *et al.* Potential of microalgae as a sustainable feed ingredient for aquaculture. *J. Biotechnol.* **341**, (2021).

318. Barboza Martignone, G., Behrendt, K. & Paparas, D. Price Transmission Analysis of the International Soybean Market in a Trade War Context. *Economies* **10**, 203 (2022).

319. Chekmarev, O. P., Lukichev, P. M. & Manilov, A. N. Agflation as a Threat to Food Security: Analysis of Inflationary Factors. in *Sustainable Agriculture: Circular to Reconstructive, Volume 1* (eds. Popkova, E. G. & Sergi, B. S.) 83–91 (Springer, Singapore, 2022). doi:10.1007/978-981-16-8731-0_9.

320. Soliman, A. M., Lau, C. K., Cai, Y., Sarker, P. K. & Dastgir, S. Asymmetric Effects of Energy Inflation, Agri-inflation and CPI on Agricultural Output: Evidence from NARDL and SVAR Models for the UK. *Energy Econ.* **126**, 106920 (2023).

321. Wang, L., Wu, R., Ma, W. & Xu, W. Examining the volatility of soybean market in the MIDAS framework: The importance of bagging-based weather information. *Int. Rev. Financ. Anal.* **89**, 102720 (2023).

322. Waithaka, M. K. *et al.* Evaluating the growth and cost–benefit analysis of feeding improved indigenous chicken with diets containing black soldier fly larva meal. *Front. Insect Sci.* **2**, (2022).

323. *The Economic Cost-Benefit Analysis of Black Soldier Fly as an Alternative Animal and Fish Feed Ingredient in Malawi*. (2023). doi:10.22004/ag.econ.338590.

324. Abro, Z. *et al.* The potential economic benefits of insect-based feed in Uganda. *Front. Insect Sci.* **2**, (2022).

325. International Monetary Fund (IMF). Global price of Fish Meal. (2024).

326. Olsen, R. L. & Hasan, M. R. A limited supply of fishmeal: Impact on future increases in global aquaculture production. *Trends Food Sci. Technol.* **27**, 120–128 (2012).

327. Siddiqui, S. A. *et al.* Replacement of soy by mealworms for livestock feed - A comparative review between soy and mealworms considering environmental aspects. *Environ. Dev. Sustain.* (2024) doi:10.1007/s10668-024-04874-1.

328. Deruytter, D. *et al.* Standardising black soldier fly larvae feeding experiments: an initial protocol and variability estimates. (2023) doi:10.1163/23524588-20230008.

329. Dossey, A., Ramos & Rojas. *Insects as Sustainable Food Ingredients: Production, Processing and Food Applications*. (2016).

330. Hoc, B., Noël, G., Carpentier, J., Francis, F. & Megido, R. C. Optimization of black soldier fly (Hermetia illucens) artificial reproduction. *PLOS ONE* **14**, e0216160 (2019).

331. Hoffmann, L. *et al.* Patterns of Genetic Diversity and Mating Systems in a Mass-Reared Black Soldier Fly Colony. *Insects* **12**, 480 (2021).

332. Hill, W. G. Selective Breeding. in *Brenner's Encyclopedia of Genetics (Second Edition)* (eds. Maloy, S. & Hughes, K.) 371–373 (Academic Press, San Diego, 2013). doi:10.1016/B978-0-12-374984-0.01390-5.

333. Facchini, E. *et al.* Long-Term Artificial Selection for Increased Larval Body Weight of Hermetia illucens in Industrial Settings. *Front. Genet.* **13**, (2022).

334. Ma, C. *et al.* Selective breeding of cold-tolerant black soldier fly (*Hermetia illucens*) larvae: Gut microbial shifts and transcriptional patterns. *Waste Manag.* **177**, 252–265 (2024).

335. Akey, J. M. *et al.* Tracking footprints of artificial selection in the dog genome. *Proc. Natl. Acad. Sci.* **107**, 1160–1165 (2010).

336. Andersson, L. Genetic dissection of phenotypic diversity in farm animals. *Nat. Rev. Genet.* **2**, 130–138 (2001).

337. Kerridge, E. *The Agricultural Revolution*. (Taylor & Francis, 2006).

338. Nelson, R. M., Pettersson, M. E. & Carlborg, Ö. A century after Fisher: time for a new paradigm in quantitative genetics. *Trends Genet.* **29**, 669–676 (2013).

339. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

340. Powder, K. E. Quantitative Trait Loci (QTL) Mapping. *Methods Mol. Biol. Clifton NJ* **2082**, 211–229 (2020).

341. Nakano, Y. & Kobayashi, Y. Genome-wide Association Studies of Agronomic Traits Consisting of Field- and Molecular-based Phenotypes. *Rev. Agric. Sci.* **8**, 28–45 (2020).

342. Andersson, L. & Georges, M. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat. Rev. Genet.* **5**, 202–212 (2004).

343. Hume, D. A., Whitelaw, C. B. A. & Archibald, A. L. The future of animal production: improving productivity and sustainability. *J. Agric. Sci.* **149**, 9–16 (2011).

344. Bouwman, A. c. *et al.* 613. Genetic parameters of black soldier flies estimated in full sib design. in *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)* 2537–2540 (Wageningen Academic Publishers, 2022). doi:10.3920/978-90-8686-940-4_613.

345. Hoppe, A. *et al.* Substantial Genetic Progress in the International Apis mellifera carnica Population Since the Implementation of Genetic Evaluation. *Insects* **11**, 768 (2020).

346. Zambrano-Gonzalez, G., Almanza-Pinzon, M. I. & Vélez-T, M. Genetic parameters in traits of productive importance in lines of Bombyx mori L. *J. Anim. Breed. Genet.* **139**, 136–144 (2022).

347. Morales-Ramos, J. A., Kelstrup, H. C., Rojas, M. G. & Emery, V. Body Mass Increase Induced by Eight Years of Artificial Selection in the Yellow Mealworm (Coleoptera: Tenebrionidae) and Life History Trade-offs. *J. Insect Sci. Online* **19**, 4 (2019).

348. Reznick, D., Bryant, M. J. & Bashey, F. r- AND K-SELECTION REVISITED: THE ROLE OF POPULATION REGULATION IN LIFE-HISTORY EVOLUTION. *Ecology* **83**, 1509–1520 (2002).

349. Eriksson, T. & Picard, C. j. Genetic and genomic selection in insects as food and feed. *J. Insects Food Feed* **7**, 661–682 (2021).

350. Rhode, C. *et al.* Genetic and phenotypic consequences of early domestication in black soldier flies (Hermetia illucens). *Anim. Genet.* **51**, 752–762 (2020).

351. Briscoe, D. a. *et al.* Rapid Loss of Genetic Variation in Large Captive Populations of Drosophila Flies: Implications for the Genetic Management of Captive Populations. *Conserv. Biol.* **6**, 416–425 (1992).

352. Smetana, S., Schmitt, E. & Mathys, A. Sustainable use of Hermetia illucens insect biomass for feed and food: Attributional and consequential life cycle assessment. *Resour. Conserv. Recycl.* **144**, 285–296 (2019).

353. Haas, Y. de, Bink, M. C. A. M., Koenen, E. P. C., Verschuren, L. M. G. & Mollenhorst, H. The contribution of animal breeding to reducing the environmental impact of livestock production. in *Reducing greenhouse gas emissions from livestock production* (Burleigh Dodds Science Publishing, 2020).

354. Zaalberg, R. M. *et al.* A bio-economic model for estimating economic values of important production traits in the black soldier fly (Hermetia illucens). (2024) doi:10.1163/23524588-00001126.

355. Slagboom, M., Nielsen, H. M., Kargo, M., Henryon, M. & Hansen, L. S. The effect of phenotyping, adult selection, and mating strategies on genetic gain and rate of inbreeding in black soldier fly breeding programs. *Genet. Sel. Evol.* **56**, 71 (2024).

356. Hansen, L. S. *et al.* Estimation of genetic parameters for the implementation of selective breeding in commercial insect production. *Genet. Sel. Evol.* **56**, 21 (2024).

357. Bhat, S. A. *et al.* Advances in genome editing for improved animal breeding: A review. *Vet. World* **10**, 1361 (2017).

358. Kaya, C. *et al.* Global population genetic structure and demographic trajectories of the black soldier fly, Hermetia illucens. *BMC Biol.* **19**, 94 (2021).

359. Ares, E. & Coe, S. Genetic Technology (Precision Breeding) Bill 2022-23. (2023).

360. Caccamo, M. New precision-breeding law unlocks gene editing in England. *Nat. Biotechnol.* **41**, 752–753 (2023).

361. Sax, J. The GMO/GE Debate. *Tex. AM Law Rev.* **4**, 345–371 (2017).

362. Brookes, G. & Smyth, S. J. Risk-appropriate regulations for gene-editing technologies. *GM Crops Food* **15**, 1–14 (2024).

363. Watson, O. & Hayta, S. Precision breeding in agriculture and food systems in the United Kingdom. *Transgenic Res.* (2024) doi:10.1007/s11248-024-00397-7.

364. Johnsson, M. *et al.* Removal of alleles by genome editing (RAGE) against deleterious load. *Genet. Sel. Evol.* **51**, (2019).

365. C, G. Genome engineering for crop improvement and future agriculture. *Cell* **184**, (2021).

366. Kim, Y. W. & Kim, A. Deletion of transcription factor binding motifs using the CRISPR/spCas9 system in the β-globin LCR. *Biosci. Rep.* **37**, BSR20170976 (2017).

367. Mou, H. *et al.* CRISPR/Cas9-mediated genome editing induces exon skipping by alternative splicing or exon deletion. *Genome Biol.* **18**, 108 (2017).

368. Zeng, X. *et al.* CRISPR/Cas9-mediated mutation of OsSWEET14 in rice cv. Zhonghua11 confers resistance to Xanthomonas oryzae pv. oryzae without yield penalty. *BMC Plant Biol.* **20**, 313 (2020).

369. Singh, P. & Ali, S. A. Impact of CRISPR-Cas9-Based Genome Engineering in Farm Animals. *Vet. Sci.* **8**, 122 (2021).

370. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).

371. Perisse, I. V., Fan, Z., Singina, G. N., White, K. L. & Polejaeva, I. A. Improvements in Gene Editing Technology Boost Its Applications in Livestock. *Front. Genet.* **11**, (2021).

372. Proudfoot, C. *et al.* Genome edited sheep and cattle. *Transgenic Res.* **24**, 147–153 (2015).

373. Kang, J.-D. *et al.* Generation of cloned adult muscular pigs with myostatin gene mutation by genetic engineering. *RSC Adv.* **7**, 12541–12549 (2017).

374. Wang, X. *et al.* CRISPR/Cas9-mediated MSTN disruption and heritable mutagenesis in goats causes increased body mass. *Anim. Genet.* **49**, 43–51 (2018).

375. Pursel, V. G. *et al.* Growth and tissue accretion rates of swine expressing an insulin-like growth factor I transgene. *Anim. Biotechnol.* **15**, 33–45 (2004).

376. Xiang, G. *et al.* Editing porcine IGF2 regulatory element improved meat production in Chinese Bama pigs. *Cell. Mol. Life Sci. CMLS* **75**, 4619–4628 (2018).

377. Tait-Burkard, C. *et al.* Livestock 2.0 – genome editing for fitter, healthier, and more productive farmed animals. *Genome Biol.* **19**, 204 (2018).

378. Zaidi, S. S.-A., Mahas, A., Vanderschuren, H. & Mahfouz, M. M. Engineering crops of the future: CRISPR approaches to develop climate-resilient and disease-resistant plants. *Genome Biol.* **21**, 289 (2020).

379. A route to de novo domestication of wild allotetraploid rice: Cell. https://www.cell.com/cell/fulltext/S0092-8674(21)00013-1?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867421000131%3Fshowall%3Dtrue.

380. Thomson, M. J., Biswas, S., Tsakirpaloglou, N. & Septiningsih, E. M. Functional Allele Validation by Gene Editing to Leverage the Wealth of Genetic Resources for Crop Improvement. *Int. J. Mol. Sci.* **23**, 6565 (2022).

381. Guo, Y. *et al.* QTL mapping for quality traits using a high-density genetic map of wheat. *PLOS ONE* **15**, e0230601 (2020).

382. Liu, X., Tu, X., He, H., Chen, C. & Xue, F. Evidence for Inbreeding Depression and Pre-Copulatory, but Not Post Copulatory Inbreeding Avoidance in the Cabbage Beetle Colaphellus bowringi. *PLOS ONE* **9**, e94389 (2014).

383. Uzunov, A., Brascamp, P., Du, M. & Ralph, B. Initiation and Implementation of Honey Bee Breeding Programs. *Bee World* 1 (2022) doi:10.1080/0005772X.2022.2031545.

384. Dhugga, K. S. Gene Editing to Accelerate Crop Breeding. *Front. Plant Sci.* **13**, 889995 (2022).

385. Port, F., Chen, H.-M., Lee, T. & Bullock, S. L. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc. Natl. Acad. Sci.* **111**, E2967–E2976 (2014).

386. Ma, S. *et al.* An integrated CRISPR Bombyx mori genome editing system with improved efficiency and expanded target sites. *Insect Biochem. Mol. Biol.* **83**, 13–20 (2017).

387. Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat. Biotechnol.* **34**, 78–83 (2016).

388. Sun, D., Guo, Z., Liu, Y. & Zhang, Y. Progress and Prospects of CRISPR/Cas Systems in Insects and Other Arthropods. *Front. Physiol.* **8**, (2017).

389. Wang, J. Y. & Doudna, J. A. CRISPR technology: A decade of genome editing is only the beginning. *Science* **379**, eadd8643 (2023).

390. Song, B., Yang, S., Hwang, G.-H., Yu, J. & Bae, S. Analysis of NHEJ-Based DNA Repair after CRISPR-Mediated DNA Cleavage. *Int. J. Mol. Sci.* **22**, 6397 (2021).

391. Guo, T. *et al.* Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing. *Genome Biol.* **19**, 170 (2018).

392. Davis, A. J. & Chen, D. J. DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.* **2**, 130 (2013).

393. Sansbury, B. M., Hewes, A. M. & Kmiec, E. B. Understanding the diversity of genetic outcomes from CRISPR-Cas generated homology-directed repair. *Commun. Biol.* **2**, 1–10 (2019).

394. Asmamaw, M. & Zawdie, B. Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing. *Biol. Targets Ther.* **15**, 353–361 (2021).

395. Karvelis, T. *et al.* Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* **16**, 253 (2015).

396. Lieber, M. R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annu. Rev. Biochem.* **79**, 181 (2010).

397. Vogt, A. & He, Y. Structure and mechanism in non-homologous end joining. *DNA Repair* **130**, 103547 (2023).

398. Weterings, E. & Chen, D. J. The endless tale of non-homologous end-joining. *Cell Res.* **18**, 114–124 (2008).

399. Kosicki, M. *et al.* Cas9-induced large deletions and small indels are controlled in a convergent fashion. *Nat. Commun.* **13**, 3422 (2022).

400. Lèbre, S. & Michel, C. J. A stochastic evolution model for residue Insertion-Deletion Independent from Substitution. *Comput. Biol. Chem.* **34**, 259–267 (2010).

401. Bai, X. *et al.* CRISPR/Cas9-mediated knockout of the eye pigmentation gene white leads to alterations in colour of head spots in the oriental fruit fly, Bactrocera dorsalis. *Insect Mol. Biol.* **28**, 837–849 (2019).

402. McCarty, N. S., Graham, A. E., Studená, L. & Ledesma-Amaro, R. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat. Commun.* **11**, 1281 (2020).

403. Port, F., Starostecka, M. & Boutros, M. Multiplexed conditional genome editing with Cas12a in Drosophila. *Proc. Natl. Acad. Sci.* **117**, 22890–22899 (2020).

404. Rojek, J. *et al. Mad7: An IP Friendly CRISPR Enzyme*. https://www.authorea.com/users/430079/articles/533620-mad7-an-ip-friendly-crispr-enzyme?commit=ac834f37b4b1f5fd55c2be8816ff39c3eb8350c0 (2021) doi:10.22541/au.162863226.68733765/v1.

405. Zetsche, B. *et al.* Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).

406. Bennis, N. X., Anderson, J. P., Kok, S. M. C. & Daran, J.-M. G. Expanding the genome editing toolbox of Saccharomyces cerevisiae with the endonuclease ErCas12a. *FEMS Yeast Res.* **23**, foad043 (2023).

407. Pietralla, J., Capdeville, N., Schindele, P. & Puchta, H. Optimizing ErCas12a for efficient gene editing in Arabidopsis thaliana. *Plant Biotechnol. J.* **22**, 401–412 (2024).

408. Santos, L. *et al.* Comparison of Cas9 and Cas12a CRISPR editing methods to correct the W1282X-CFTR mutation. *J. Cyst. Fibros.* **21**, 181–187 (2022).

409. Smirnov, K. *et al.* Comparison of CRISPR-MAD7 and CRISPR-Cas9 for Gene Disruptions in Komagataella phaffii. *J. Fungi* **10**, 197 (2024).

410. Lin, Q. *et al.* Genome editing in plants with MAD7 nuclease. *J. Genet. Genomics Yi Chuan Xue Bao* **48**, 444–451 (2021).

411. M, M., W, W., D, D. & C, S. A MAD7-based genome editing system for Escherichia coli. *Microb. Biotechnol.* (2023) doi:10.1111/1751-7915.14234.

412. Liu, Z. *et al.* ErCas12a CRISPR-MAD7 for Model Generation in Human Cells, Mice, and Rats. *CRISPR J.* **3**, 97–108 (2020).

413. Wang, B. Improving CRISPR-Cas9's Rate of Homology-Directed Repair. *Genet. Eng. Biotechnol. News* **38**, 12–13 (2018).

414. Devkota, S. The road less traveled: strategies to enhance the frequency of homology-directed repair (HDR) for increased efficiency of CRISPR/Cas-mediated transgenesis. *BMB Rep.* **51**, 437 (2018).

415. Chen, Q. *et al.* Structural basis of seamless excision and specific targeting by piggyBac transposase. *Nat. Commun.* **11**, 3446 (2020).

416. Li, X. *et al.* piggyBac transposase tools for genome engineering. *Proc. Natl. Acad. Sci.* **110**, E2279–E2287 (2013).

417. Am, H. & Ra, H. Germline transformation of Drosophila melanogaster with the piggyBac transposon vector. *Insect Mol. Biol.* **8**, (1999).

418. Yoshida, J. *et al.* Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Sci. Rep.* **7**, 43613 (2017).

419. Galvan, D. L. *et al.* Genome-Wide Mapping of PiggyBac Transposon Integrations in Primary Human T Cells. *J. Immunother. Hagerstown Md 1997* **32**, 837–844 (2009).

420. Morris, A. C. Microinjection of mosquito embryos. in *The Molecular Biology of Insect Disease Vectors: A Methods Manual* (eds. Crampton, J. M., Beard, C. B. & Louis, C.) 423–429 (Springer Netherlands, Dordrecht, 1997). doi:10.1007/978-94-009-1535-0_35.

421. Spranghers, T., Noyez, A., Schildermans, K. & De Clercq, P. Cold Hardiness of the Black Soldier Fly (Diptera: Stratiomyidae). *J. Econ. Entomol.* **110**, 1501–1507 (2017).

422. Kou, Z. *et al.* Establishment of highly efficient transgenic system for black soldier fly (Hermetia illucens). *Insect Sci.* (2022) doi:10.1111/1744-7917.13147.

423. Campos-Ortega, J. A. & Hartenstein, V. Stages of Drosophila Embryogenesis. in *The Embryonic Development of Drosophila melanogaster* (eds. Campos-Ortega, J. A. & Hartenstein, V.) 9–102 (Springer, Berlin, Heidelberg, 1997). doi:10.1007/978-3-662-22489-2_2.

424. Irish, V., Lehmann, R. & Akam, M. The Drosophila posterior-group gene nanos functions by repressing hunchback activity. *Nature* **338**, 646–648 (1989).

425. Raz, E. The function and regulation of vasa-like genes in germ-cell development. *Genome Biol.* **1**, reviews1017.1 (2000).

426. Gilbert, S. F. Early Drosophila Development. in *Developmental Biology. 6th edition* (Sinauer Associates, 2000).

427. Mehravar, M., Shirazi, A., Nazari, M. & Banan, M. Mosaicism in CRISPR/Cas9-mediated genome editing. *Dev. Biol.* **445**, 156–162 (2019).

428. Sebo, Z. L., Lee, H. B., Peng, Y. & Guo, Y. A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. *Fly (Austin)* **8**, 52–57 (2014).

429. Generalovic, T. N. *et al.* A high-quality, chromosome-level genome assembly of the Black Soldier Fly (Hermetia illucens L.). *G3 GenesGenomesGenetics* **11**, jkab085 (2021).

430. Chen, B. *et al.* Intersex is required for female sexual development in Hermetia illucens. *Insect Sci.* **30**, 901–911 (2023).

431. Fellner, S. K., Rybczynski, R. & Gilbert, L. I. Ca2+ signaling in prothoracicotropic hormone-stimulated prothoracic gland cells of Manduca sexta: evidence for mobilization and entry mechanisms. *Insect Biochem. Mol. Biol.* **35**, 263–275 (2005).

432. Williams, J. A., Bell, J. B. & Carroll, S. B. Control of Drosophila wing and haltere development by the nuclear vestigial gene product. *Genes Dev.* **5**, 2481–2495 (1991).

433. Tomberlin, J. K. & Sheppard, D. C. Lekking Behavior of the Black Soldier Fly (Diptera: Stratiomyidae). *Fla. Entomol.* **84**, 729–730 (2001).

434. Garrett-Engele, C. M. *et al.* intersex, a gene required for female sexual development in Drosophila, is expressed in both sexes and functions together with doublesex to regulate terminal differentiation. *Development* **129**, 4661–4675 (2002).

435. Sui, Z., Wu, Q., Geng, J., Xiao, J. & Huang, D. CRISPR/Cas9-mediated efficient white genome editing in the black soldier fly Hermetia illucens. *Mol. Genet. Genomics MGG* **299**, 5 (2024).

436. Dong, Y. *et al.* Genome-wide identification of yellow gene family in Hermetia illucens and functional analysis of yellow-y by CRISPR/Cas9. *Insect Sci.* (2024) doi:10.1111/1744-7917.13371.

437. Xiao, C., Qiu, S. & Robertson, R. M. The white gene controls copulation success in Drosophila melanogaster. *Sci. Rep.* **7**, 7712 (2017).

438. Massey, J. H., Chung, D., Siwanowicz, I., Stern, D. L. & Wittkopp, P. J. The yellow gene influences Drosophila male mating success through sex comb melanization. *eLife* **8**, e49388 (2019).

439. Jiang, Y., Kou, Z., Chen, B., Xingyu, L. & Huang, Y. Functional characterization of Lipid storage droplets 1 (LSD1) in growth and lipolysis of Hermetia illucens. 2024.03.22.586280 Preprint at https://doi.org/10.1101/2024.03.22.586280 (2024).

440. Gunther, D. *et al.* Transgenic black soldier flies for production of carotenoids. *Insect Biochem. Mol. Biol.* **168**, 104110 (2024).

441. Pfitzner, C. *et al.* Rapid generation and screening of transgenic black soldier fly (Hermetia illucens). 2024.02.21.581498 Preprint at https://doi.org/10.1101/2024.02.21.581498 (2024).

442. Politi, Y. Unveiling the Enigmatic World of the BSF Larvae | FreezeM. https://www.freezem.com/unveiling-the-enigmatic-world-of-the-bsf-larvae/ (2024).

443. YASSAF, I. N., GOREN, A. & ALYAGOR, I. High-weight black soldier fly larvae, methods of producing same and use thereof. (2024).

444. Juarez-Carreño, S. *et al.* Body-fat sensor triggers ribosome maturation in the steroidogenic gland to initiate sexual maturation in Drosophila. *Cell Rep.* **37**, 109830 (2021).

445. Kalis, A. Bugera: Bio-Fuel of the Future. *Milltrust* https://www.milltrust.com/bugera-bio-fuel-of-the-future/ (2023).

446. Varol, B. Unveiling M-BSF Platform - The most advanced platform for rapid development of BSF strains. *Feed & Additive Magazine* https://www.feedandadditive.com/unveiling-m-bsf-platform-the-most-advanced-platform-for-rapid-development-of-bsf-strains/ (2024).

447. Watson, E. ByBug: Black soldier flies can deliver cheaper, more scalable recombinant protein production. *AgFunderNews* https://agfundernews.com/bybug-black-soldier-flies-can-deliver-cheaper-more-scalable-recombinant-protein-production (2024).

448. Watson, E. Insulin… from black soldier flies? FlyBlast makes the case for bugs vs microbes for recombinant protein production. *AgFunderNews* https://agfundernews.com/insulin-from-black-solider-flies-flyblast-makes-the-case-for-bugs-vs-microbes-for-recombinant-protein-production (2024).

449. van Gulik, W. M., van Winden, W. A. & Heijnen, J. J. Metabolic Flux Analysis, Modeling, and Engineering Solutions. in *Handbook of Industrial Cell Culture: Mammalian, Microbial, and Plant Cells* (eds. Vinci, V. A. & Parekh, S. R.) 349–392 (Humana Press, Totowa, NJ, 2003). doi:10.1007/978-1-59259-346-0_15.

450. Gòdia, F. & Cairó, J. J. Metabolic engineering of animal cells. *Bioprocess Biosyst. Eng.* **24**, 289–298 (2002).

451. Tang, X. & Zhang, Y. Beyond knockouts: fine-tuning regulation of gene expression in plants with CRISPR-Cas-based promoter editing. *New Phytol.* **239**, 868–874 (2023).

452. Ceccotti, C. *et al.* New value from food and industrial wastes - Bioaccumulation of omega-3 fatty acids from an oleaginous microbial biomass paired with a brewery by-product using black soldier fly (Hermetia illucens) larvae. *Waste Manag.* **143**, 95–104 (2022).

453. Peguero, D. A., Gold, M., Vandeweyer, D., Zurbrügg, C. & Mathys, A. A Review of Pretreatment Methods to Improve Agri-Food Waste Bioconversion by Black Soldier Fly Larvae. *Front. Sustain. Food Syst.* **5**, (2022).

454. Peguero, D. A. *et al.* Physical pretreatment of three biowastes to improve black soldier fly larvae bioconversion efficiency. *Waste Manag.* **178**, 280–291 (2024).

455. Xiang, F., Zhang, Q., Xu, X. & Zhang, Z. Black soldier fly larvae recruit functional microbiota into the intestines and residues to promote lignocellulosic degradation in domestic biodegradable waste. *Environ. Pollut.* **340**, 122676 (2024).

456. Sun, J. & Zhou, X. J. Utilization of Lignocellulose-feeding Insects for Viable Biofuels: an Emerging and Promising Area of Entomological Science. in *Recent Advances in Entomological Research: From Molecular Biology to Pest Management* (eds. Liu, T. & Kang, L.) 434–500 (Springer, Berlin, Heidelberg, 2011). doi:10.1007/978-3-642-17815-3_25.

457. Sanluis-Verdes, A. *et al.* Wax worm saliva and the enzymes therein are the key to polyethylene degradation by Galleria mellonella. *Nat. Commun.* **13**, 5568 (2022).

458. Tepper, K. *et al.* Methylmercury demethylation and volatilization by animals expressing microbial enzymes. 2023.12.11.571038 Preprint at https://doi.org/10.1101/2023.12.11.571038 (2023).

459. Xu, J. *et al.* Identification of a germline-expression promoter for genome editing in Bombyx mori. *Insect Sci.* **26**, 991–999 (2019).

460. Akbari, O. S., Papathanos, P. A., Sandler, J. E., Kennedy, K. & Hay, B. A. Identification of germline transcriptional regulatory elements in Aedes aegypti. *Sci. Rep.* **4**, 3954 (2014).

461. Hersperger, F. *et al.* Drosophila macrophages control systemic cytokine levels in response to oxidative stress via a non-canonical DNA damage repair signaling cascade. *eLife* **12**, (2023).

462. Eckermann, K. N. *et al.* Hyperactive piggyBac transposase improves transformation efficiency in diverse insect species. *Insect Biochem. Mol. Biol.* **98**, 16–24 (2018).

463. Buchholz, R., Miyazaki, W. & Dompe, N. 5 Fly Drosphila Genomic Prep for iPCR in 96-well Format. in (Exelixis, Inc, San Francisco, CA, 2001).

464. Dieffenbach, C. W., Lowe, T. M. & Dveksler, G. S. General concepts for PCR primer design. *PCR Methods Appl.* **3**, S30-37 (1993).

465. Davydova, S. *et al.* Next-generation genetic sexing strain establishment in the agricultural pest Ceratitis capitata. *Sci. Rep.* **13**, 19866 (2023).

466. Pfitzner, C. *et al.* Rapid generation and screening of transgenic black soldier fly (Hermetia illucens). *bioRxiv* (2024) doi:10.1101/2024.02.21.581498.

467. Khan, M. A. Fine-tuning Cas9 half-life for highly efficient gene drive in Anopheles stephensi (PhD thesis). (The Royal Veterinary College, The Pirbright Institute, London, 2024).

468. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* **47**, W171–W174 (2019).

469. Bloh, K. *et al.* Deconvolution of Complex DNA Repair (DECODR): Establishing a Novel Deconvolution Algorithm for Comprehensive Analysis of CRISPR-Edited Sanger Sequencing Data. *CRISPR J.* **4**, 120–131 (2021).

470. Pinello, L. *et al.* Analyzing CRISPR genome-editing experiments with CRISPResso. *Nat. Biotechnol.* **34**, 695–697 (2016).

471. Li, M. *et al.* Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti. *Proc. Natl. Acad. Sci.* **114**, E10540–E10549 (2017).

472. Morgens, D. W. *et al.* Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8**, 15178 (2017).

473. Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol. Ther. - Nucleic Acids* **4**, e264 (2015).

474. Kyrou, K. *et al.* A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nat. Biotechnol.* **36**, 1062–1066 (2018).

475. Vidaurre, V. & Chen, X. Epigenetic regulation of *drosophila* germline stem cell maintenance and differentiation. *Dev. Biol.* **473**, 105–118 (2021).

476. Terradas, G., Hermann, A., James, A. A., McGinnis, W. & Bier, E. High-resolution in situ analysis of Cas9 germline transcript distributions in gene-drive Anopheles mosquitoes. *G3 GenesGenomesGenetics* **12**, jkab369 (2022).

477. Meccariello, A. *et al.* Gene drive and genetic sex conversion in the global agricultural pest Ceratitis capitata. *Nat. Commun.* **15**, 372 (2024).

478. Deshpande, G., Spady, E., Goodhouse, J. & Schedl, P. Maintaining Sufficient Nanos Is a Critical Function for Polar Granule Component in the Specification of Primordial Germ Cells. *G3 GenesGenomesGenetics* **2**, 1397–1403 (2012).

479. Liu, N., Han, H. & Lasko, P. Vasa promotes Drosophila germline stem cell differentiation by activating mei-P26 translation by directly interacting with a (U)-rich motif in its 3′ UTR. *Genes Dev.* **23**, 2742 (2009).

480. Kor, S. D. *et al.* RNA Pol III promoters—key players in precisely targeted plant genome editing. *Front. Genet.* **13**, (2023).

481. Didychuk, A. L., Butcher, S. E. & Brow, D. A. The life of U6 small nuclear RNA, from cradle to grave. *RNA* **24**, 437 (2018).

482. Mefferd, A. L., Kornepati, A. V., Bogerd, H. P., Kennedy, E. M. & Cullen, B. R. Expression of CRISPR/Cas single guide RNAs using small tRNA promoters. *RNA* **21**, 1683 (2015).

483. Wittkopp, P. J., Williams, B. L., Selegue, J. E. & Carroll, S. B. Drosophila pigmentation evolution: Divergent genotypes underlying convergent phenotypes. *Proc. Natl. Acad. Sci.* **100**, 1808–1813 (2003).

484. O'Grady, P. M. & DeSalle, R. Insect evolution: How the fruit fly changed (some of) its spots. *Curr. Biol.* **10**, R75–R77 (2000).

485. Davidowitz, G., D'Amico, L. J. & Nijhout, H. F. Critical weight in the development of insect body size. *Evol. Dev.* **5**, 188–197 (2003).

486. Ward, C. M. *et al.* White pupae phenotype of tephritids is caused by parallel mutations of a MFS transporter. *Nat. Commun.* **12**, 491 (2021).

487. Khanna, A., Pradhan, A. & Curran, S. P. Emerging Roles for Maf1 beyond the Regulation of RNA Polymerase III Activity. *J. Mol. Biol.* **427**, 2577–2585 (2015).

488. Caldwell, P. E., Walkiewicz, M. & Stern, M. Ras activity in the Drosophila prothoracic gland regulates body size and developmental rate via ecdysone release. *Curr. Biol. CB* **15**, 1785–1795 (2005).

489. Honegger, B. *et al.* Imp-L2, a putative homolog of vertebrate IGF-binding protein 7, counteracts insulin signaling in Drosophila and is essential for starvation resistance. *J. Biol.* **7**, 10 (2008).

490. Walkiewicz, M. A. & Stern, M. Increased insulin/insulin growth factor signaling advances the onset of metamorphosis in Drosophila. *PloS One* **4**, e5072 (2009).

491. Choi, T.-Y. *et al.* Cereblon Maintains Synaptic and Cognitive Function by Regulating BK Channel. *J. Neurosci. Off. J. Soc. Neurosci.* **38**, 3571–3583 (2018).

492. Wakabayashi, S. *et al.* Ohgata, the Single Drosophila Ortholog of Human Cereblon, Regulates Insulin Signaling-dependent Organismic Growth. *J. Biol. Chem.* **291**, 25120–25132 (2016).

493. Taylor, J. R. *et al.* Sirt6 regulates lifespan in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2111176119 (2022).

494. Shukla, N. & Kolthur-Seetharam, U. Drosophila Sirtuin 6 mediates developmental diet-dependent programming of adult physiology and survival. *Aging Cell* **21**, e13576 (2022).

495. Panwar, V. *et al.* Multifaceted role of mTOR (mammalian target of rapamycin) signaling pathway in human health and disease. *Signal Transduct. Target. Ther.* **8**, 1–25 (2023).

496. Dibble, C. C. *et al.* TBC1D7 Is a Third Subunit of the TSC1-TSC2 Complex Upstream of mTORC1. *Mol. Cell* **47**, 535–546 (2012).

497. Ren, S., Huang, Z., Jiang, Y. & Wang, T. dTBC1D7 regulates systemic growth independently of TSC through insulin signaling. *J. Cell Biol.* **217**, 517–526 (2018).

498. Alcaino, C., Farrugia, G. & Beyder, A. Mechanosensitive Piezo Channels in the Gastrointestinal Tract. *Curr. Top. Membr.* **79**, 219 (2017).

499. Min, S. *et al.* Control of feeding by Piezo-mediated gut mechanosensation in Drosophila. *eLife* **10**, e63049 (2021).

500. Anderson, M. A. E. *et al.* Expanding the CRISPR Toolbox in Culicine Mosquitoes: In Vitro Validation of Pol III Promoters. *ACS Synth. Biol.* **9**, 678–681 (2020).

501. Chen, S. & Corces, V. G. The gypsy insulator of Drosophila affects chromatin structure in a directional manner. *Genetics* **159**, 1649 (2001).

502. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for Drosophila using germ-line-specific φC31 integrases. *Proc. Natl. Acad. Sci.* **104**, 3312–3317 (2007).

503. Tóth, E. *et al.* Restriction enzyme body doubles and PCR cloning: on the general use of type IIs restriction enzymes for cloning. *PloS One* **9**, e90896 (2014).

504. Liao, C. *et al.* Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nat. Commun.* **10**, 2948 (2019).

505. Ceriani, M. F. Basic Protocols for Drosophila S2 Cell Line. in *Circadian Rhythms: Methods and Protocols* (ed. Rosato, E.) 415–422 (Humana Press, Totowa, NJ, 2007). doi:10.1007/978-1-59745-257-1_33.

506. Ehrke-Schulz, E. *et al.* Quantification of designer nuclease induced mutation rates: a direct comparison of different methods. *Mol. Ther. Methods Clin. Dev.* **3**, 16047 (2016).

507. Che, L.-R. *et al.* Electroporation-mediated nucleic acid delivery during non-embryonic stages for gene-function analysis in *Anopheles sinensis*. *Insect Biochem. Mol. Biol.* **128**, 103500 (2021).

508. Gregory, M., Alphey, L., Morrison, N. I. & Shimeld, S. M. Insect transformation with piggyBac: getting the number of injections just right. *Insect Mol. Biol.* **25**, 259–271 (2016).

509. Chaverra-Rodriguez, D. *et al.* Targeted delivery of CRISPR-Cas9 ribonucleoprotein into arthropod ovaries for heritable germline gene editing. *Nat. Commun.* **9**, 3008 (2018).

510. Alegria, A. D. *et al.* High-throughput genetic manipulation of multicellular organisms using a machine-vision guided embryonic microinjection robot. *Genetics* **226**, iyae025 (2024).

511. Meza, J. S. *et al.* Fitness Cost Implications of PhiC31-Mediated Site-Specific Integrations in Target-Site Strains of the Mexican Fruit Fly, Anastrepha ludens (Diptera: Tephritidae). *PLOS ONE* **9**, e109690 (2014).

512. Zabelina, V. *et al.* The Exact Timing of Microinjection of Parthenogenetic Silkworm Embryos Is Crucial for Their Successful Transgenesis. *Front. Physiol.* **13**, (2022).

513. Gabrieli, P. & Scolari, F. Delivery of Nucleic Acids through Embryo Microinjection in the Worldwide Agricultural Pest Insect, Ceratitis capitata. *J. Vis. Exp. JoVE* 54528 (2016) doi:10.3791/54528.

514. Rand, M. D., Kearney, A. L., Dao, J. & Clason, T. Permeabilization of *Drosophila* embryos for introduction of small molecules. *Insect Biochem. Mol. Biol.* **40**, 792–804 (2010).

515. Bertinetti, C., Samayoa, A. C. & Hwang, S.-Y. Effects of Feeding Adults of Hermetia illucens (Diptera: Stratiomyidae) on Longevity, Oviposition, and Egg Hatchability: Insights Into Optimizing Egg Production. *J. Insect Sci.* **19**, 19 (2019).

516. Thomas, N. K., Karpati, Z., Schmitt, T. & Riabinina, O. A chemically defined oviposition attractant and repellent of Black Soldier Flies (Hermetia illucens). 2024.06.04.597456 Preprint at https://doi.org/10.1101/2024.06.04.597456 (2024).

517. Addo, P., Fosu-Gyasi, S., Oduro-Kwarteng, S., Duku, G. A. & Awuah, E. Assessing egg deposition behaviour of female black soldier fly (*Hermetia: Illucens)* in Kumasi, Ghana using MOSW as bait. *J. Asia-Pac. Entomol.* **27**, 102168 (2024).

518. Grainger, S. Beta Bugs Ltd. Personal communication. (2021).

519. Liu, Z., Najar-Rodriguez, A. J., Minor, M. A., Hedderley, D. I. & Morel, P. C. H. Mating success of the black soldier fly, Hermetia illucens (Diptera: Stratiomyidae), under four artificial light sources. *J. Photochem. Photobiol. B* **205**, 111815 (2020).

520. Barrett, M., Chia, S. Y., Fischer, B. & Tomberlin, J. Welfare considerations for farming black soldier flies, Hermetia illucens (Diptera: Stratiomyidae): a model for the insects as food and feed industry. *J. Insects Food Feed* **9**, 1–30 (2022).

521. Johnson, S. A. & Milner, M. J. The final stages of wing development in Drosophila melanogaster. *Tissue Cell* **19**, 505–513 (1987).

522. Zhang, J. *et al.* An Artificial Light Source Influences Mating and Oviposition of Black Soldier Flies, Hermetia illucens. *J. Insect Sci.* **10**, 202 (2010).

523. Allocca, M. *et al.* The Fruit Fly, Drosophila melanogaster: The Making of a Model (Part I). in *Drosophila melanogaster - Model for Recent Advances in Genetics and Therapeutics* (IntechOpen, 2018). doi:10.5772/intechopen.72832.

524. Sampath, K. S. & Puttaraju, H. P. Improvised microinjection technique for mosquito vectors. *Indian J. Med. Res.* **136**, 971–978 (2012).

525. Heryanto, C., Mazo-Vargas, A. & Martin, A. Efficient hyperactive piggyBac transgenesis in Plodia pantry moths. *Front. Genome Ed.* **4**, (2022).

526. Meccariello, A. *et al.* Highly efficient DNA-free gene disruption in the agricultural pest Ceratitis capitata by CRISPR-Cas9 ribonucleoprotein complexes. *Sci. Rep.* **7**, 10061 (2017).

527. Nellaiappan, K. & Sugumaran, M. Chapter Three - Hardening and tanning of insect ootheca, egg cases, egg sac, chorion, and silk. in *Advances in Insect Physiology* (ed. Sugumaran, M.) vol. 62 215–271 (Academic Press, 2022).

528. Tollenaar, W., Joosten, L. & Schmitt, E. The phases and effects of low temperature exposure on embryonic development in Hermetia illucens (L.) (Diptera: Stratiomyidae). *J. Insects Food Feed* **8**, 379–386 (2022).

529. Martin, A. C. The Physical Mechanisms of Drosophila Gastrulation: Mesoderm and Endoderm Invagination. *Genetics* **214**, 543 (2020).

530. Toivonen, J. M. *et al.* No Influence of Indy on Lifespan in Drosophila after Correction for Genetic and Cytoplasmic Background Effects. *PLOS Genet.* **3**, e95 (2007).

531. Meccariello, A. Imperial College London. Personal communication. (2022).

532. Huynh, C. Q. & Zieler, H. Construction of modular and versatile plasmid vectors for the high-level expression of single or multiple genes in insects and insect cell lines. *J. Mol. Biol.* **288**, 13–20 (1999).

533. Liu, J., Rayes, D. & Akbari, O. S. A fluorescent sex-sorting technique for insects with the demonstration in Drosophila melanogaster. *bioRxiv* 2023.08.11.553026 (2023) doi:10.1101/2023.08.11.553026.

534. C, H., B, J. & Ea, W. Highly sensitive, fluorescent transformation marker for Drosophila transgenesis. *Dev. Genes Evol.* **210**, (2000).

535. Kokoza, V., Ahmed, A., Wimmer, E. A. & Raikhel, A. S. Efficient transformation of the yellow fever mosquito Aedes aegypti using the piggyBac transposable element vector pBac[3xP3-EGFP afm]. *Insect Biochem. Mol. Biol.* **31**, 1137–1143 (2001).

536. Osanai-Futahashi, M. *et al.* A visible dominant marker for insect transgenesis. *Nat. Commun.* **3**, 1295 (2012).

537. Handler, A. M. & Harrell, R. A. Polyubiquitin-regulated DsRed marker for transgenic insects. *BioTechniques* **31**, 820, 824–828 (2001).

538. Handler, A. M. & Harrell, R. A. Transformation of the Caribbean fruit fly, *Anastrepha suspensa*, with a *piggyBac* vector marked with polyubiquitin-regulated GFP. *Insect Biochem. Mol. Biol.* **31**, 199–205 (2001).

539. Aladdin, A., Sahly, N., Faty, R., Youssef, M. & Salem, T. *The Baculovirus Promoter OpIE2 Sequence Has Inhibitory Effect on the Activity of the Cytomegalovirus (CMV) Promoter in HeLa and HEK-293T Cells*. (2020). doi:10.1101/2020.10.09.315515.

540. Linzhu, R. *et al.* Comparative analysis of the activity of two promoters in insect cells. *Afr. J. Biotechnol.* **10**, 8930–8941 (2011).

541. 黄勇平, 詹帅, 王四宝, 寇宗庆 & 许军. Marker gene applied to CRISPR/Cas9 gene editing system of hermetia illucens. (2020).

542. Ye, S., Enghiad, B., Zhao, H. & Takano, E. Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in Streptomyces. *J. Ind. Microbiol. Biotechnol.* **47**, 413–423 (2020).

543. Port, F. *et al.* A large-scale resource for tissue-specific CRISPR mutagenesis in Drosophila. *eLife* **9**, e53865 (2020).

544. Ablain, J., Durand, E. M., Yang, S., Zhou, Y. & Zon, L. I. A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Dev. Cell* **32**, 756–764 (2015).

545. Feng, Y.-Q., Lorincz, M. C., Fiering, S., Greally, J. M. & Bouhassira, E. E. Position Effects Are Influenced by the Orientation of a Transgene with Respect to Flanking Chromatin. *Mol. Cell. Biol.* **21**, 298–309 (2001).

546. Hacker, U., Nystedt, S., Barmchi, M. P., Horn, C. & Wimmer, E. A. piggyBac-based insertional mutagenesis in the presence of stably integrated P elements in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7720–7725 (2003).

547. Horn, C., Offen, N., Nystedt, S., Häcker, U. & Wimmer, E. A. piggyBac-based insertional mutagenesis and enhancer detection as a tool for functional insect genomics. *Genetics* **163**, 647–661 (2003).

548. Potter, C. J. & Luo, L. Splinkerette PCR for Mapping Transposable Elements in Drosophila. *PLOS ONE* **5**, e10168 (2010).

549. Aravamudan, B., Fergestad, T., Davis, W. S., Rodesch, C. K. & Broadie, K. Drosophila UNC-13 is essential for synaptic transmission. *Nat. Neurosci.* **2**, 965–971 (1999).

550. Arora, K. *et al.* The Drosophila schnurri gene acts in the Dpp/TGF beta signaling pathway and encodes a transcription factor homologous to the human MBP family. *Cell* **81**, 781–790 (1995).

551. Dolezal, T., Dolezelova, E., Zurovec, M. & Bryant, P. J. A Role for Adenosine Deaminase in Drosophila Larval Development. *PLoS Biol.* **3**, e201 (2005).

552. van Meyel, D. J., Thomas, J. B. & Agulnick, A. D. Ssdp proteins bind to LIM-interacting co-factors and regulate the activity of LIM-homeodomain protein complexes in vivo. *Dev. Camb. Engl.* **130**, 1915–1925 (2003).

553. Salim, S. *et al.* The ortholog of human ssDNA-binding protein SSBP3 influences neurodevelopment and autism-like behaviors in Drosophila melanogaster. *PLoS Biol.* **21**, e3002210 (2023).

554. Merritt, T. J. S. *et al.* Quantifying Interactions Within the NADP(H) Enzyme Network in Drosophila melanogaster. *Genetics* **182**, 565–574 (2009).

555. Stanton, R. C. Glucose-6-Phosphate Dehydrogenase, NADPH, and Cell Survival. *Iubmb Life* **64**, 362–369 (2012).

556. Ghani, M. U. *et al.* Comprehensive review on glucose 6 phosphate dehydrogenase: A critical immunometabolic and redox switch in insects. *Int. J. Biol. Macromol.* **273**, 132867 (2024).

557. Luckinbill, L. S., Riha, V., Rhine, S. & Grudzien, T. A. The role of glucose-6-phosphate dehydrogenase in the evolution of longevity in Drosophila melanogaster. *Heredity* **65 (Pt 1)**, 29–38 (1990).

558. Legan, S. K. *et al.* Overexpression of Glucose-6-phosphate Dehydrogenase Extends the Life Span of Drosophila melanogaster. *J. Biol. Chem.* **283**, 32492–32499 (2008).

559. Frank, J. E. Diagnosis and management of G6PD deficiency. *Am. Fam. Physician* **72**, 1277–1282 (2005).

560. Richardson, S. R. & O'Malley, G. F. *Glucose-6-Phosphate Dehydrogenase Deficiency*. (StatPearls Publishing, Treasure Island (FL), 2023).

561. Chen, P.-H. *et al.* Glucose-6-Phosphate Dehydrogenase, Redox Homeostasis and Embryogenesis. *Int. J. Mol. Sci.* **23**, 2017 (2022).

562. Yang, H.-C. *et al.* Glucose 6-phosphate dehydrogenase deficiency enhances germ cell apoptosis and causes defective embryogenesis in Caenorhabditis elegans. *Cell Death Dis.* **4**, e616 (2013).

563. Oonincx, D. G. A. B., Volk, N., Diehl, J. J. E., van Loon, J. J. A. & Belušič, G. Photoreceptor spectral sensitivity of the compound eyes of black soldier fly (*Hermetia illucens*) informing the design of LED-based illumination to enhance indoor reproduction. *J. Insect Physiol.* **95**, 133–139 (2016).

564. D'Amato, R. *et al.* Anti-CRISPR Anopheles mosquitoes inhibit gene drive spread under challenging behavioural conditions in large cages. *Nat. Commun.* **15**, 952 (2024).

565. Khan, M. Imperial College London. Personal communication. (2024).

566. Kou, Z. *et al.* Wingless strain created using binary transgenic CRISPR/Cas9 alleviates concerns about mass rearing of Hermetia illucens. *Commun. Biol.* **7**, 1652 (2024).

567. Ikonomou, L., Schneider, Y.-J. & Agathos, S. N. Insect cell culture for industrial production of recombinant proteins. *Appl. Microbiol. Biotechnol.* **62**, 1–20 (2003).

568. Wei, M., Mi, C.-L., Jing, C.-Q. & Wang, T.-Y. Progress of Transposon Vector System for Production of Recombinant Therapeutic Proteins in Mammalian Cells. *Front. Bioeng. Biotechnol.* **10**, (2022).

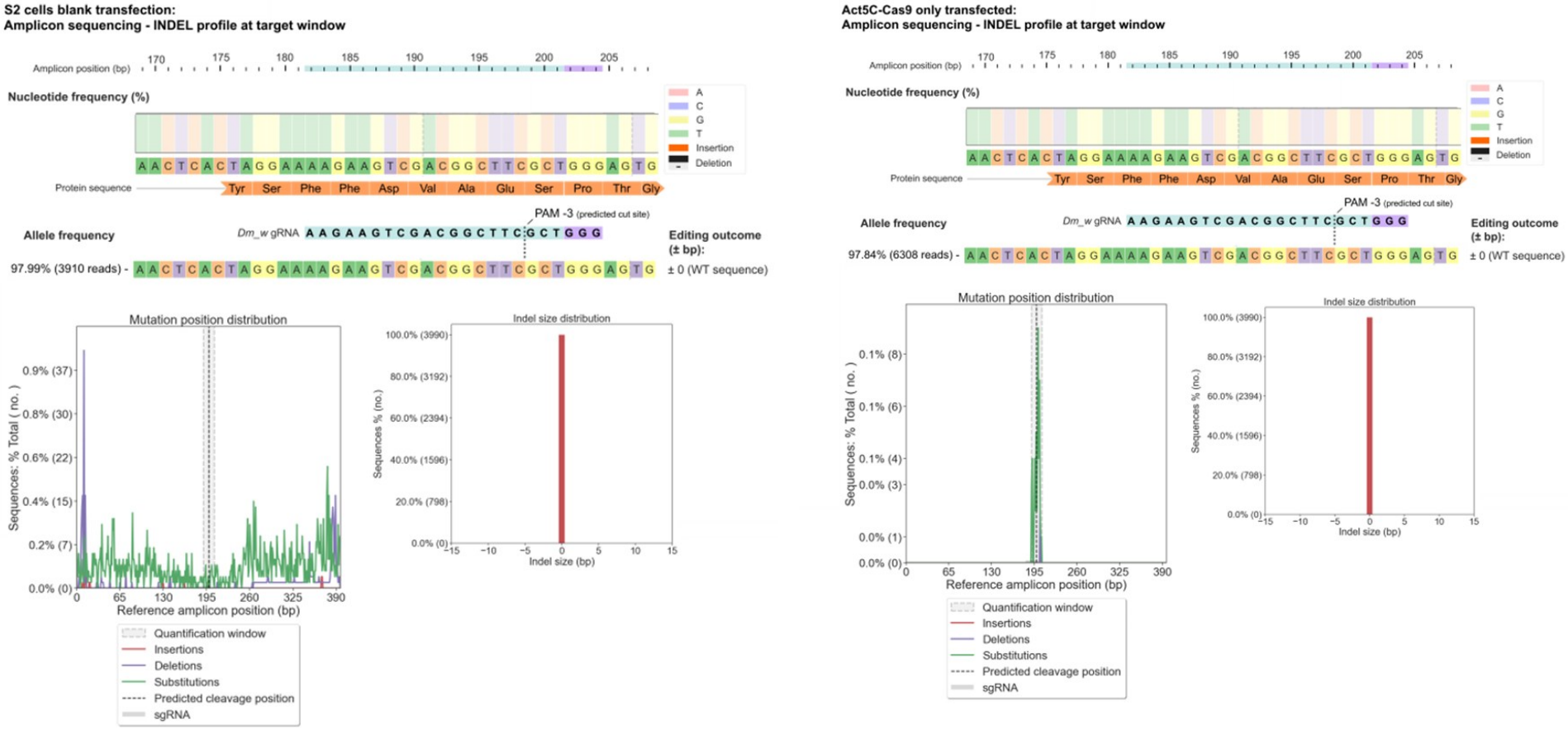
569. Häcker, I. *et al.* Cre/lox-Recombinase-Mediated Cassette Exchange for Reversible Site-Specific Genomic Targeting of the Disease Vector, Aedes aegypti. *Sci. Rep.* **7**, 43883 (2017).

570. Tang, Y.-H. *et al.* CRISPR-Cas13-mediated RNA editing in the silkworm Bombyx mori. *Zool. Res.* **45**, 1249–1260 (2024).

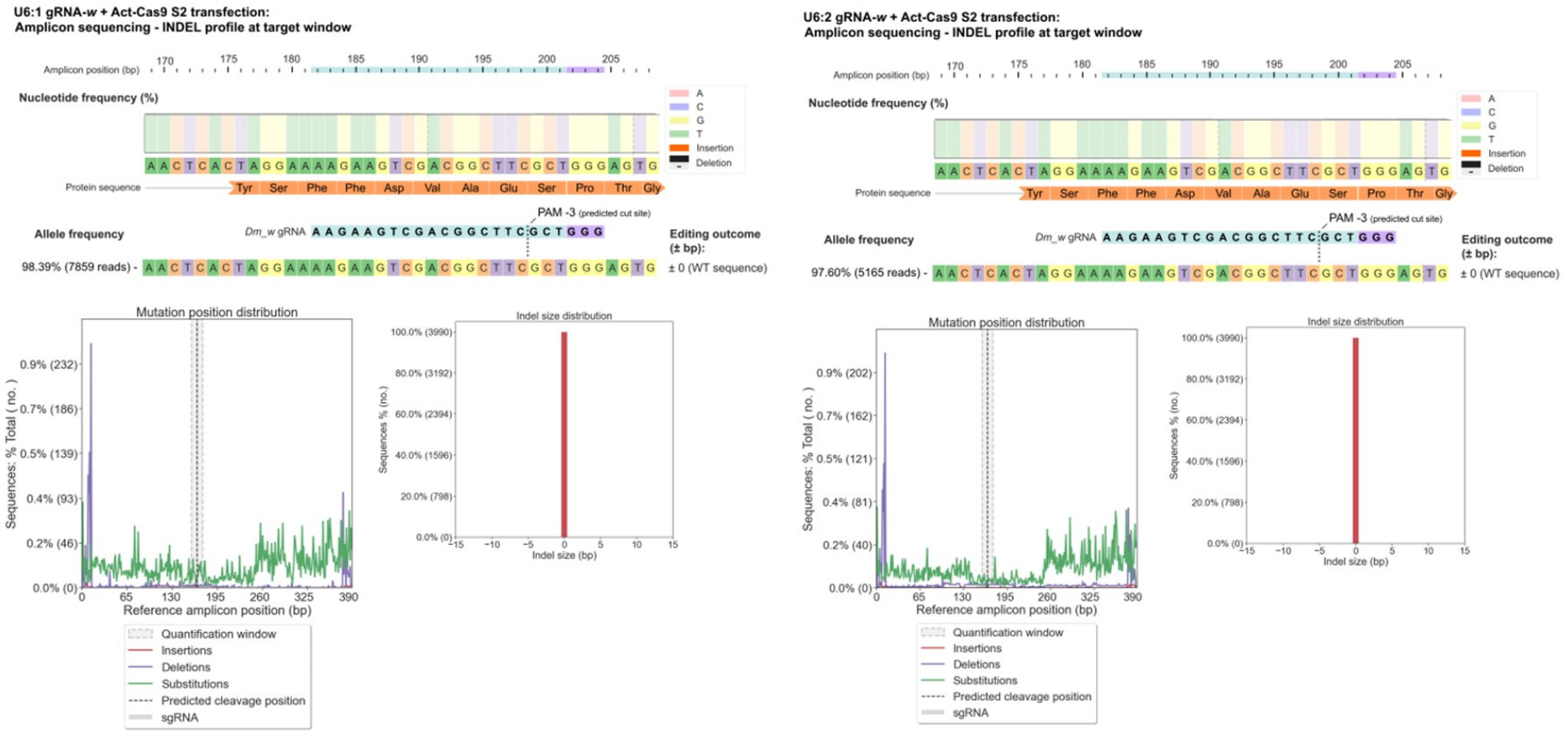
571. Clark, M. *et al.* Expanding the CRISPR base editing toolbox in Drosophila melanogaster. *Commun. Biol.* **7**, 1–11 (2024).

572. Bosch, J. A., Birchak, G. & Perrimon, N. Precise genome engineering in Drosophila using prime editing. *Proc. Natl. Acad. Sci.* **118**, e2021996118 (2021).

Appendix



Supplementary Figure 1: Amplicon sequencing analysis of blank transfected and *Act5C*-Cas9 only transfected S2 cells from section 3.2.7.



Supplementary Figure 2: Amplicon sequencing analysis of U6:1 gRNA-*w* and U6:2 gRNA-*w* co-transfected with *Act5C*-Cas9 in S2 cells from section 3.2.7.

Transgene inheritence statistics

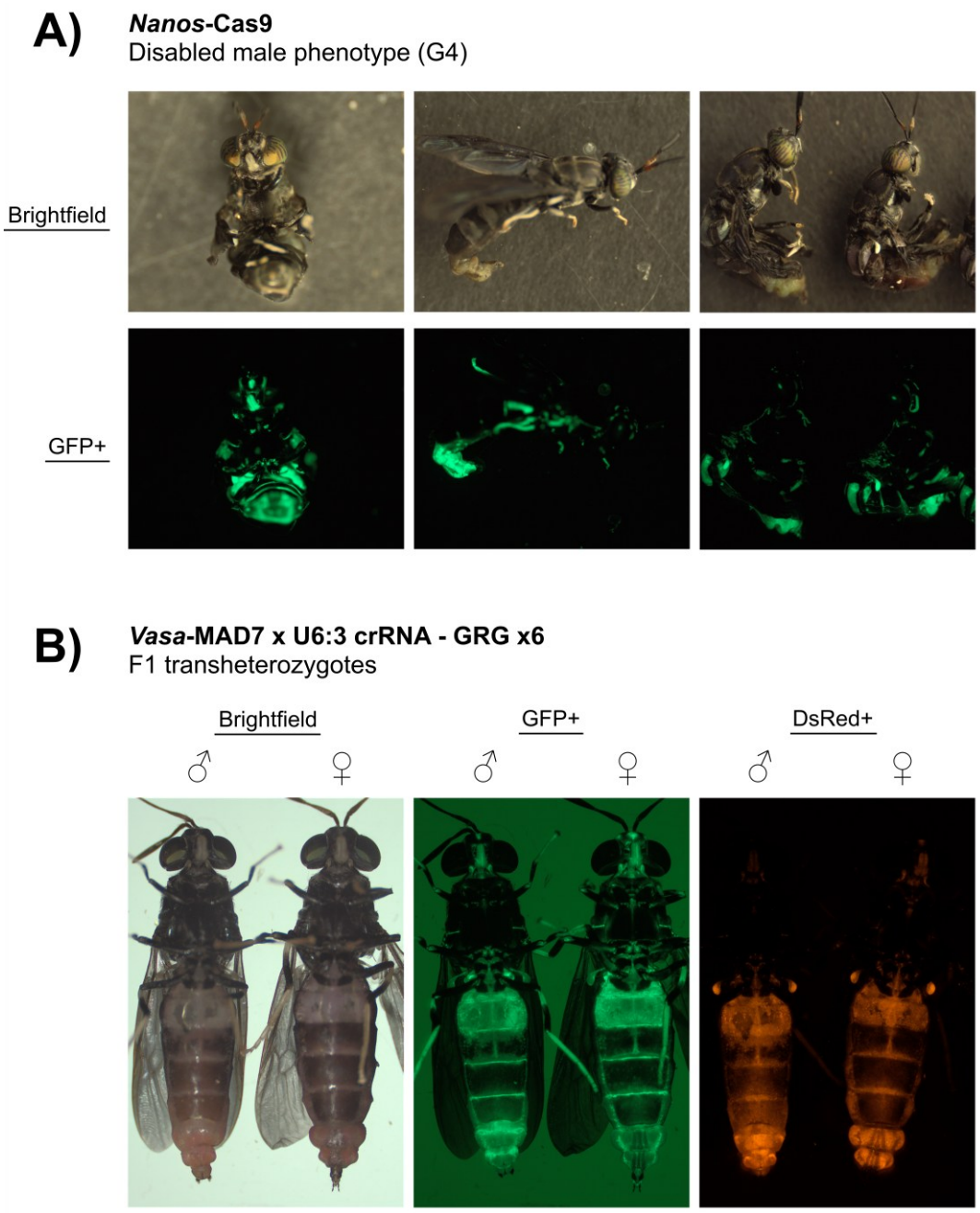
Inheritence by sex — **Exact binomial test**

Generation	Line	Sex	# Transgenic	# Non-transgenic	# Total	% Transgenic	p-value	Probability	95% CI - lower	95% CI - Upper
G3	Nanos-Cas9	M	85	158	243	34.98%	3.29E-06	0.35	0.290	0.413
G3	Nanos-Cas9	F	81	149	230	35.22%	8.65E-06	0.35	0.291	0.418
G5	Nanos-Cas9	M	22	67	89	24.72%	1.90E-06	0.25	0.162	0.350
G3	Nanos-Cas9	F	10	66	76	13.16%	2.96E-11	0.13	0.065	0.229
G3	Vasa-Cas9	M	241	224	465	51.83%	0.4581	0.52	0.472	0.565
G3	Vasa-Cas9	F	202	256	458	44.10%	0.01318	0.44	0.395	0.488
G3	Vasa-MAD7	M	100	106	206	48.54%	0.7277	0.49	0.415	0.556
G3	Vasa-MAD7	F	86	116	202	42.57%	0.04104	0.43	0.357	0.497
G3	Act-MAD7	M	156	198	354	44.07%	0.02918	0.44	0.388	0.494
G3	Act-MAD7	F	202	236	438	46.12%	0.1147	0.46	0.414	0.509
G3	U6-3 gRNA-*w*	M	114	105	219	52.05%	0.5889	0.52	0.452	0.588
G3	U6-3 gRNA-*w*	F	102	141	243	41.98%	0.01461	0.42	0.357	0.485
G3	U6-3 crRNA-*w* x2	M	175	137	312	56.09%	0.03603	0.56	0.504	0.617
G3	U6-3 crRNA-*w* x2	F	181	170	351	51.57%	0.5936	0.52	0.462	0.569
G3	U6-3 crRNA-GRG x6	M	123	159	282	43.62%	0.03695	0.44	0.377	0.496
G3	U6-3 crRNA-GRG x6	F	119	137	256	46.48%	0.288	0.46	0.403	0.528

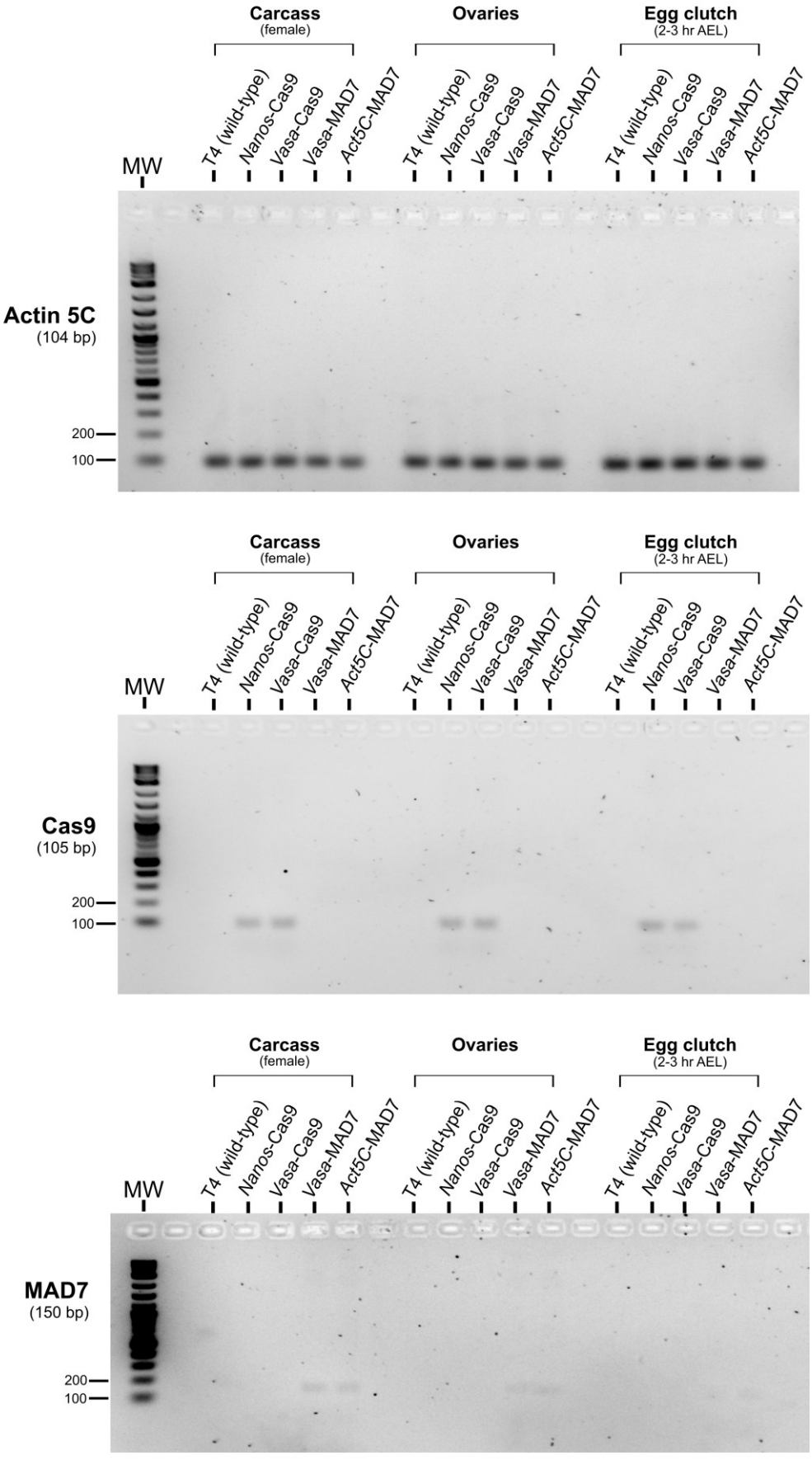
Inheritence - both sexes — **G-test goodness of fit**

Generation	Line	Sex	# Transgenic	# Non-transgenic	# Total	% Transgenic	p-value	G value	X^2 df
G3	Nanos-Cas9	Mixed	166	307	473	35.10%	6.5E-11	42.677	1
G5	Nanos-Cas9	Mixed	32	133	165	19.39%	5.9E-16	65.463	1
G3	Vasa-Cas9	Mixed	443	480	923	48.00%	2.2E-01	1.484	1
G3	Vasa-MAD7	Mixed	186	222	408	45.59%	7.5E-02	3.181	1
G3	Act-MAD7	Mixed	358	434	792	45.20%	6.9E-03	7.304	1
G3	U6-3 gRNA-*w*	Mixed	216	246	462	46.75%	1.6E-01	1.949	1
G3	U6-3 crRNA-*w* x2	Mixed	356	307	663	53.70%	5.7E-02	3.625	1
G3	U6-3 crRNA-GRG x6	Mixed	242	296	538	44.98%	2.0E-02	5.429	1

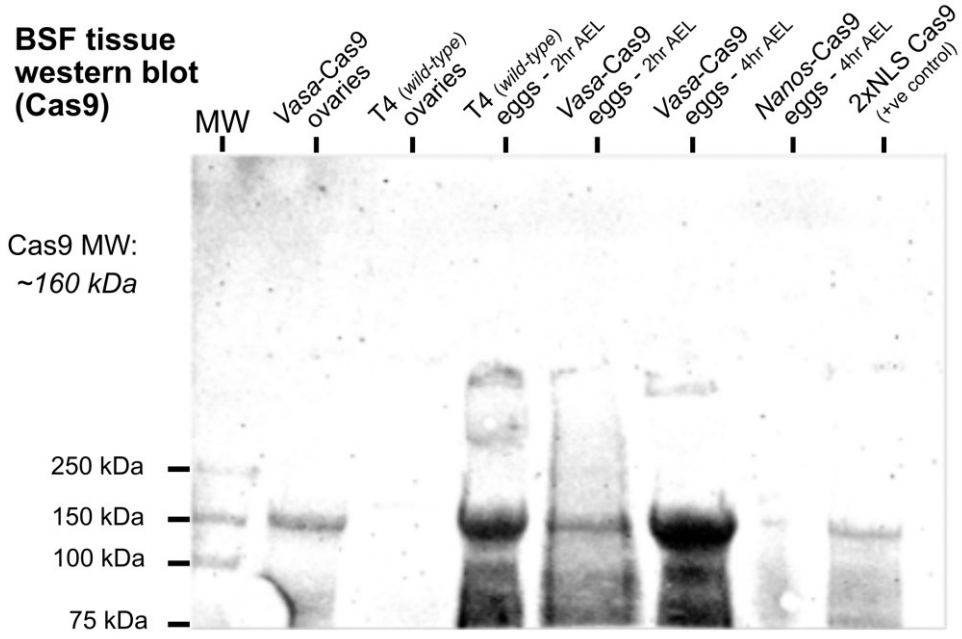
Supplementary Figure 3: Transgenic inheritance data for the CRISPR lines used in chapter 5.2.2.



Supplementary Figure 4. *Nanos*-Cas9 flies exhibited an apparent lack of fitness in later generations (>G4), while *Vasa*-MAD7 / U6:3 crRNA – GRG x6 transheterozygotes had no clear fitness cost.



Supplementary Figure 5: Unedited RT-PCR gels used in chapter 5.2.5



Supplementary Figure 6: Unedited Western blot used in chapter 5.2.5