ABSTRACT

GUTZMANN, NICOLE ELIZABETH. Making, Testing, and Debating Genetically Modified Insects (Under the direction of Dr. Marcé Lorenzen).

Modern biotechnology provides powerful tools that could potentially be used to increase food security and protect human lives. Insect pests vector human and plant diseases, causing billions in damages and negatively impacting lives. Recent advances in gene editing have broadened the horizon for the emerging field of genetic pest management. This type of pest management makes use of biotechnologies to manage pest populations through genetic rather than chemical means. The red flour beetle (*Tribolium castaneum*) is a global pest of stored grain and model organism for studying development and pest biology.

A better understanding of CRISPR/Cas9 in Tribolium will facilitate genetic studies in this model, as well as provide scaffolding for building management strategies for other pest insects. We created transgenic *Cas9* beetle lines that can be used for functional genomics studies pertinent to genetic pest management. As expected, we found gene editing efficiency to be higher when *Cas9* is expressed in the germline than when provided by injection of plasmid DNA.

We also used CRISPR/Cas9 to investigate regions of a pesticide receptor (cadherin; receptor of Cry3 toxin) found in the midgut of beetles. In the US, most corn production uses transgenic varieties which express multiple Cry toxin genes from the bacterium *Bacillus thuringiensis* (Bt). A better understanding of Bt mode of action and Cry3 resistance will facilitate resistance management in pests like Western corn rootworm which alone costs producers billions of dollars every year.

In addition to gene editing, gene drives have recently been of heightened interest to the pest management community. The goal of gene drives is to spread engineered genes through a wild population by releasing a relatively small number of transgenic individuals. Towards developing a gene drive system, we investigated the molecular function of a naturally occurring gene drive found in the red flour beetle. *Medea* factors are hypothesized to function via a

toxin/antidote mechanism. Our analysis supports this hypothesis and also identifies the toxin gene in *Medea1*. Though promising, gene drives have many risks and are highly contentious globally. We used interviews to explore the early-career researcher perspective of responsible innovation (RI) and identified challenges and opportunities of applying the RI framework in genetic engineering research.

© Copyright 2021 by Nicole Gutzmann

All Rights Reserved

Making, Testing, and Debating Genetically Modified Insects

by
Nicole Elizabeth Gutzmann

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Entomology

Raleigh, North Carolina
2021

APPROVED BY:

Marcé Lorenzen Committee Chair	Jennifer Kuzma
Max Scott	Fred Gould

DEDICATION

To my parents who inspired my love of biology and gave me strength to pursue my dreams.

BIOGRAPHY

Nicole Gutzmann was born and raised in the Inland Empire of sunny Southern California. She completed her B.S. in Biology at University of California, Riverside and found her passion for molecular biology under the mentorship of Dr. Erin Wilson-Rankin. To prepare for a future career in biotechnology research, she sought training experiences that combined her research passion with her enthusiasm for working on teams and communicating across disciplines. These values led her to become an interdisciplinary student working on a Doctorate of Philosophy in Entomology with a minor in Genetic Engineering and Society. Her goal upon graduating is to gain entry at a biotechnology company concerned with sustainability and social responsibility, with the intention of working towards a position in scientific affairs.

ACKNOWLEDGMENTS

So many thanks to…

All my beetle babies that died… for science. I couldn't have done it without you.

Marcé for being my cheerleader, cultivating a positive lab culture, and pointing the way to a million (and one) interesting research opportunities.

Jennifer Kuzma for your guidance and patience. Chapter 5 would not have been possible without you.

Max Scott and Fred Gould for always having an open door and answering my many research design questions.

The Genetic Engineering and Society Center for all the invaluable research and professional development opportunities over the past six years. I am eternally grateful for being a part of the NSF IGERT fellowship cohort.

My labmates, especially Dr. Clay Chu for showing me the ropes and offering me valuable guidance and advise, and Dr. Nathaniel Grubbs for offering the exact opposite advice.

The fantastic undergrads Ashley Williams, Lyric Ellis, and Lauren Slayton, who helped me grow as a mentor and spent hours collecting data.

The office ladies of Thomas Hall. Your friendship made grad school a joyful and vibrant time of my life. They say human bonds are born of shared struggle, and struggle we did. Collecting that firewood, suffering through the acrid cheese hotel, and the never knowing when you might find a surprise rat -- in your water bottle, overhead cabinet, or yoga mat.

My husband Michael for sneaking to the grocery store at midnight to pick up surprise ice cream during the writing of this dissertation.

The funding sources that made traveling to conferences possible: the NCSU Genetic Engineering and Society Center, NSF, USDA, the National Corn Growers Association, and BASF Corporation.

This work was supported by the National Science Foundation [grant no. 1068676]. NSF IGERT 'Genetic Engineering and Society: The Case of Transgenic Pests.'

This work was supported by the Education and Workforce Development fellowship [grant no. 2019-67011-29617] from the USDA National Institute of Food and Agriculture.

This work was also supported in part by the National Science Foundation [grant no. MCB-1244772].

TABLE OF CONTENTS

LIST OF TABLES viii
LIST OF FIGURES x
LIST OF SUPPLEMENTARY MATERIALS xii

Chapter 1: Overview of Gene Editing and Gene Drive in Insects: Technical and Social Challenges 1
Dissertation Summary 2
Part I: Background and Key Terms 5
Part II: Technical Challenges of Genetic Pest Management 6
Part III: Social Challenges of Genetic Pest Management 12
References 17

Chapter 2: Improved Gene Editing Efficiency in *Tribolium* with Constitutive Expression of Cas9 Transgene 26
Introduction 27
Methods 29
  Tribolium strains 29
  Insect Rearing 30
  Plasmid DNAs 30
  Generation of transgenic Tribolium carrying Cas9 transgene 33
  Microinjection for insertion of Cas9 transgene 32
  Mating scheme for the development of Cas9 lines 33
  Fluorescence screening 35
  Inverse PCR 35
  High molecular weight gDNA extraction and Pac-Bio sequencing 36
  Fitness assay 37
  Microinjections for CRISPR/Cas9 knockout of EGFP 37
  Screening EGFP mosaic knockouts 38
  Screening EGFP gametic knockouts 38
  Statistical Analysis: One-way ANOVA 39
  Statistical Analysis: Correlation coefficient 39
Results 40
  Screening for Cas9 insertion in Tribolium genome 40
  Breeding to eliminate helper gene and multiple *Cas9* insertions 44
  Cas9 insertion locations 47
  Fitness of transgenic Cas9 Tribolium lines 49
  Ranking the Cas9 lines and other Cas9 tools in Tribolium 51
  Somatic knockout rates of muscle and eye EGFP using Cas9 beetle lines 49
  Comparing PubCas9 versus αTubCas9 lines 59
  Gametic knockout rates (Heredity of EGFP knock-out) 60
  Maternal preloading of Cas9 into embryos 63
  Comparing Cas9 tools: Cas9 plasmids, protein, and transgenic lines 65
Discussion 71
Conclusion 77
References 79
Supplementary Material 84

Chapter 3: Investigating the Role of Cadherin Receptors in Coleopteran Resistance to Bt Cry3Aa 121
Introduction 122
History and benefits of Bt pesticides 122
Mechanisms of resistance 124
Bt corn and field-evolved resistance in western corn rootworm 124
Mechanisms of resistance to Cry3 proteins in Coleoptera 125
Western corn rootworm (*Diabrotica virgifera virgifera*) 91
Mealworm beetle (*Tenebrio molitor*) 94
Lesser mealworm (*Alphitobius diaperinus*) 95
Red flour beetle (*Tribolium castaneum*) 95
Colorado potato beetle (*Leptinotarsa decemlineata*) 96
Leaf beetle pest of poplar trees (*Chrysomela tremula*) 96
Germline transformation to investigate resistance mechanisms 133
Methods 139
Insect strains 139
Homology-directed repair donor templates 139
CRISPR/Cas9 single guide RNAs (sgRNAs) 145
Microinjection 145
Florescent marker screening 109
Results 146
Discussion 154
Conclusion 155
References 157
Supplementary Material 166

Chapter 4: *Medea* Gene Drive as Transgene in Drosophila and Non-*Medea* Tribolium: Hurdles and a Path Forward 182
Introduction 183
The biology and prevalence of *Medea* 183
Medea gene candidates 184
Methods 187
Insect Strains 187
Generation of donor plasmids 149
Microinjections 193
Confirmation of *piggyBac* excision 193
Screening for *ATPSynthCDUF1703* insertion in beetles 194
Screening for *ATPSynthCDUF1703* insertion in flies 194
Results 195
Generation of Tribolium and Drosophila ATPsynthC-DUF1703 transgenes 195
Insertion of the *ATPsynthC-DUF1703* transgene in Tribolium 196
Insertion of the *ATPsynthC-DUF1703* transgene in Drosophila 203
Discussion 205
Using Wolbachia as a model 206
Possible bacterial endosymbiont origins of the M1 DUF1703 gene 208
Conclusion 213
References 214
Supplementary Material 222

Chapter 5: Responsible Innovation in Genetic Engineering Research: What Does it Mean to Early Career Researchers? 239
Introduction 240
Responsible Innovation 240
Responsible Innovation in biotechnology 242
Stakeholder and researcher perspectives 244
Early-career researcher perspectives 246
Methods 248
Research approach and sampling 248
Student backgrounds 251
Data collection 251
Data analysis 254
Results 257
Micro-level challenges and opportunities to RI 262
Student: Skills or Knowledge 264
Student: Duty 266
Student: Authority 268
Student: Time 270
Student: Independent Mentor 271
Professor: Authority 272
Professor: Mentorship 274
Professor: Access 275
Meso-level challenges and opportunities to RI 276
Collaboration 277
Lab focus and inheritance 280
Funding 281
Training opportunities 283
Public outreach 285
Public engagement 288
Extension programs 292
Macro-level challenges and opportunities to RI 293
Mechanics or practicality 294
Risks and values 297
Publication 298
Discussion 299
Conclusion 300
References 303
Supplementary Material 307

LIST OF TABLES

Table 1.1 Transgenic Cas9 Insect Strains 9

Table 2.1 Primer and guide sequences 33

Table 2.2 Inverse PCR primers 36

Table 2.3 Microinjection survival and transformation rates 42

Table 2.4 Names of all G_1 transformants 43

Table 2.5 Survival rate of injected G_0 eggs to larva (hatch rate) and from larva to pupa 54

Table 2.6 Somatic muscle EGFP knockout screening 55

Table 2.7 Somatic muscle EGFP knockout rates 56

Table 2.8 Somatic eye EGFP knockout screening 58

Table 2.9 Gametic muscle and eye EGFP knockout screening 61

Table 2.10 Somatic muscle EGFP knockout rates from Cas9 fathers versus Cas9 mothers 64

Table 2.11 Tribolium Cas9 tools ranked by knockout rate 66

Table 2.12 Tribolium Cas9 tools ranked by knockout rate at high and medium level 66

Table 2.13 Survival rate of injected G_0 eggs to larva (hatch rate) and from larva to pupa 67

Table 2.14 Somatic EGFP muscle knockout rates from Cas9 plasmids, purified protein, and transgenic lines 68

Table 2.15 Gametic EGFP muscle knockout rates from p(bhsp68Cas9), purified protein, and the top two performing transgenic lines 70

Table 2.16 Comparison between lines with the same Cas9 gene insertion location 76

Table 3.1 Sensitivity to Cry3 toxin following RNAi knockdown of binding proteins 131

Table 3.2 Resistance ratio following CRISPR/Cas9 knock-in of proposed resistance mutations 134

Table 3.3 Resistance ratio following CRISPR/Cas9 knockout of cadherin binding receptors 137

Table 3.4 Primers for building *p(TcCadHDRa-3xP3EGFP)* 141

Table 3.5 Primers for building *p(TcCadHDRb-PubEGFP)* 143

Table 3.6 Primers for sequencing and bacterial colony PCR 144

Table 3.7 CRISPR/Cas9 guide RNAs for targeting TcCad1 145

Table 3.8 CRISPR/Cas9-mediated homology directed repair results 150

Table 4.1 Primers for making a *piggyBac* donor plasmid *pB(TcPubEGFP-ATPsynthCDUF1703)* 188

Table 4.2 Primers for sequencing *piggyBac* donor plasmid *pB(TcPubEGFP-ATPsynthCDUF1703)* 191

Table 4.3 Primers for sequencing attB-containing donor plasmid *p(attB-DmPubEGFP-ATPsynthCDUF1703)* 192

Table 4.4 Primers for *piggyBac* excision assay 194

Table 4.5 Survival and somatic knock-in of Medea (ATPsynthCDUF1703) in injected beetles 200

Table 4.6 Efficiency of *piggyBac*-mediated germline transformation in *Tribolium castaneum* with two different donors 202

Table 4.7 Efficiency of PhiC31-mediated germline transformation in *Drosophila melanogaster* 204

Table 4.8 Top BlastX hits of the M1-associated DUF1703 in insects 210

Table 4.9 Top BlastX hits of the M1-associated DUF1703 in bacteria 211

Table 5.1 Semi-structured interview questions 253

Table 5.2 Number of observations at the macro, meso, and micro levels 259

Table 5.3 Summary table of themes at all levels (micro, meso, macro) 260

Table 5.4 Comparison of micro-level opportunities and challenges to RI 264

Table 5.5 Comparison of meso-level opportunities and challenges to RI 277

Table 5.6 Comparison of macro-level opportunities and challenges to RI 294

LIST OF FIGURES

Figure 2.1 G_1 Fluorescent Marker Screening in Pupae and Adults 41

Figure 2.2 Variation in eye fluorescence between G_1 individuals of different lines 45

Figure 2.3 EGFP expression variation between Cas9 lines 46

Figure 2.4 Cas9 insertion locations 48

Figure 2.5 Fitness of 10 Cas9 lines and 3 control lines 50

Figure 2.6 Tribolium lines used in EGFP knock-out experiment 52

Figure 2.7 Mosaic muscle EGFP knockout in pupae of EGFP Guide 3 injected individuals 53

Figure 2.8 Mosaic eye EGFP knockout in pupae of EGFP Guide 3 injected individual 57

Figure 2.9 Cas9 lines sorted by (A) muscle EGFP knockout rate, (B) High+Medium level muscle EGFP knockout rate, (C) eye EGFP knockout rates, and (D) fitness 59

Figure 3.1 Structure and sequence of insect cadherin Cry toxin receptors (Figure 2A Fabrick et al., 2009) 128

Figure 3.2 CRISPR/Cas9-mediated substitution of CR12 region of TcCad1 gene with TmCad1 147

Figure 3.3 CRISPR/Cas9-mediated substitution of CR8-10 region of TcCad1 gene with TmCad1 148

Figure 3.4 Fluorescent marker gene expression in G_1 offspring of HDRa-injected parents 149

Figure 3.5 Sanger sequencing to confirm seamless substitution 151

Figure 3.6 Transient EGFP from injected plasmids 152

Figure 3.7 Predicted promoter sequences within the promoter TcPub 154

Figure 4.1 Predicted domain architecture of ATPsynthC-DUF1703 196

Figure 4.2 Shortened piggyBac arms and mutation 197

Figure 4.3 Sequence alignment for *piggyBac* excision assay 198

Figure 4.4 Transient expression of Pub-EGFP from PiggyBac plasmid 198

Figure 4.5 Somatic transformation in G_0 beetles 199

Figure 4.6 Differences in sterility caused by Wolbachia versus Medea 207

Figure 5.1 Schematic for the organization of themes and observations 256

Figure 5.2 Proportion of opportunities versus challenges of RI at multiple levels, as perceived by early-career researchers 258

LIST OF SUPPLEMENTARY MATERIALS

Supplementary Material 2.1 Mating scheme to eliminate the X-linked piggyBac helper from Cas9 lines 85

Supplementary Material 2.2 Mating scheme to segregate multiple insertions of Cas9 transgene 86

Supplementary Material 2.3 Optimized protocol for High Molecular Weight (HMW) gDNA extractions for PacBio sequencing 87

Supplementary Material 2.4 Inverse PCR results which revealed concatemerization of *piggyBac* construct 90

Supplementary Material 2.5 Concatemerization of *piggyBac* construct 91

Supplementary Material 2.6 *Cas9* insertion location and *piggyBac* junctions 92

Supplementary Material 2.7 Correlation coefficient between fitness and knockout rates 93

Supplementary Material 2.8 Gametic EGFP muscle knockout 95

Supplementary Material 2.9 One-way ANOVA 105

Supplementary Material 2.10 Gametic EGFP muscle knockout rates from *p(bhsp68Cas9)*, purified protein, and the top two performing transgenic lines 106

Supplementary Material 2.11 SnapGene feature map of *pB(TcPubCas9-3xP3EGFP)* and *pB(TcαTubCas9-3xP3EGFP)* 108

Supplementary Material 2.12 >pB(TcPubCas9-3xP3EGFP) 109

Supplementary Material 2.13 >pB(TcαTubCas9-3xP3EGFP) 115

Supplementary Material 3.1 Chimeric primer design for building HDRa 167

Supplementary Material 3.2 CLUSTAL O(1.2.4) multiple sequence alignment between TcCad1 and the edited versions (HDRa in CR12 and HDRb in CR8-10) 168

Supplementary Material 3.3 Substitution of CR12 region of TcCad1 gene with TmCad1 and marker gene 171

Supplementary Material 3.4 Substitution of CR8-10 region of TcCad1 gene with TmCad1 and marker gene 172

Supplementary Material 3.5 SnapGene feature maps of *p(TcCadHDRa-3xP3EGFP)* and *p(TcCadHDRb-PubEGFP)* 173

Supplementary Material 3.6 >p(TcCadHDRa-3xP3EGFP) 174

Supplementary Material 3.7 >p(TcCadHDRb-PubEGFP) 177

Supplementary Material 3.8 Point mutations from cloning with *E. coli* 181

Supplementary Material 4.1 CLUSTAL O(1.2.4) multiple sequence alignment of M1 DUF1703 and bacteria DUF1703 from the PD-(D/E)XK nuclease review paper by Knizewski et al. (2007) 223

Supplementary Material 4.2 SnapGene feature maps 226

Supplementary Material 4.3 > *pB(TcPubEGFP- ATPsynthCDUF1703)* 227

Supplementary Material 4.4 > *pB(attB-PubEGFP- DmATPsynthCDUF1703)* 233

Supplementary Material 5.1 Codebook 308

CHAPTER 1

Overview of Gene Editing and Gene Drive in Insects:

Technical and Social Challenges

Dissertation Summary

Small but mighty, insects are essential to terrestrial life through their ecosystem roles as pollinators, soil aerators, decomposers, pest controllers, and providers of food for other taxa. Humans benefit greatly from these services, yet are also negatively affected by insects in their roles as herbivores of agricultural products and vectors of plant and human diseases. Each year over 700,000 people die from vector-borne diseases, and agricultural pests cause $70 billion USD in damages and food waste (WHO 2020, FAO 2019). Pest insect populations are managed through a variety of tactics including cultural, physical, biological, chemical, and genetic. Genetic methods include the use of pest-resistant plant varieties and genetic engineering of the pests themselves (called Genetic Pest Management or GPM). GPM uses modified versions of a pest to mate with wild pests in a target area.

The first GPM strategy Sterile Insect Technique (SIT) has been used globally since the 1950s (Klassen et al. 2005). Many more strategies exist today, advancing alongside improvements in gene editing techniques. Perhaps the most notable is gene drive which has been imagined and considered since the 1960s and in the last ten years has had exponential growth in interest and research. Genetic pest management has been discussed and developed as an approach to increase food security and to prevent the spread of some of the deadliest human diseases (Harris et al. 2012, Jin et al. 2013, Esvelt et al. 2014, Hammond et al. 2016). However, developers also recognize it is not a silver bullet and gene editing is historically and presently a highly controversial and polarizing topic. Thus, there has been an effort by the academic community to develop the technology in a way that is considered socially responsible. This dissertation explores a couple of facets of GPM from varying levels of analysis ranging from technical methods to perceptions of responsible development of the technology.

The following chapters are organized by level of analysis starting with specialized technical methods and moving up to broad level analysis. Chapter 2 introduces the major recent advancement in gene editing technology called CRISPR-Cas9. The aim of this chapter is to

develop genetic tools to mitigate some of the biggest limitations of CRISPR-Cas9 in the red flour beetle (*Tribolium castaneum*), a global pest of stored grain and model organism for development and pest biology. We assess two gene promoters for driving expression of *Cas9* in beetles, create multiple transgenic *Cas9* beetle lines, and compare efficiency between multiple Cas9 and guide RNA delivery methods. To our knowledge, transgenic *Cas9* lines have only been developed in the silkworm and multiple species of flies and mosquitos. The methods and tools we developed are easily transferable to other beetle species and also facilitate functional genomic studies in this model organism. Testing of promoters is additionally essential for future applications in gene drive, a concept that is covered in later chapters.

Chapter 3 uses methods and knowledge built in Chapter 2 to investigate regions of a pesticide receptor found in the midgut of beetles. In the US, over 80% of corn and cotton acres are planted with transgenic varieties that express one or more Cry toxin genes from the bacterium *Bacillus thuringiensis* (Bt) (USDA 2017). The aim of this chapter is to gain an understanding of Bt mode of action in beetles through two objectives: (1) Modify a Cry3Aa toxin binding protein, cadherin, by replacing small portions of the *Cad1* gene in Cry3Aa-tolerant *Tribolium castaneum* with the equivalent portion from a related Cry3Aa-susceptible species, *Tenebrio molitor* and (2) Describe the future direction of comparing mortality of modified and wild-type beetles when reared on Cry3Aa. A better understanding of Bt mode of action will inform the engineering of modified Bt toxins to counter insecticide resistance. These findings are therefore relevant to the plant engineering and integrated pest management communities. The genetic pest management community may also find this relevant as at least one genetic pest management strategy, Release of Insects carrying a Dominant Lethal (RIDL), has been suggested as a form of resistance management when used in conjunction with Bt crops (Harvey-Samuel et al. 2015).

Chapter 4 introduces gene drives, a specific type of genetic engineering where the goal is to spread engineered genes through a wild population by releasing a relatively small number of transgenic individuals. Towards developing a gene drive system, we investigated the function of

a naturally occurring gene drive found in the red flour beetle, known as *Medea* factors. Named for the Greek mythological woman who killed her children, *Medea* factors are hypothesized to selectively kill offspring via a toxin/antidote mechanism. Through our analysis, we build on this toxin/antidote hypothesis commonly held in the literature, identify the toxin gene, identify the molecular mechanism of toxin transfer from mother to egg, and refute an earlier hypothesis that the toxin and antidote both stemmed from a single gene. Identification of the toxin gene and mechanism of toxin transfer led us to the revelation that *Medea* is like a female version of cytoplasmic incompatibility (CI). In CI caused by the insect endosymbiont *Wolbachia*, sperm carry a toxin and are unable to form viable offspring unless an egg has the antidote. In *Medea*, we hypothesize eggs carry a toxin within their maternally inherited mitochondria and the antidote can come from either the sperm or the egg. *Medea* is also similar to cytonuclear incompatibility which results from lethal interactions between genes in the mitochondrial (located in the cytoplasm) and nuclear genomes, with the caveat that *Medea* toxin is nuclear encoded and later shuttled into the mitochondria. This discovery supports the candidacy of using *Medea* to build an engineered gene drive, though future discovery of the antidote gene and precise molecular mechanism are essential.

Chapter 5 looks at CRISPR-Cas9 technologies from a slightly higher level of analysis. We introduce a social science framework called Responsible (Research and) Innovation (RI/RRI) that can be used to address ethical and social concerns during the research and innovation process. RI topics include anticipation of impacts, examination of research beliefs and framing, inclusion of relevant publics, and incorporation of these findings back into the research process. Previous studies have analyzed the perspectives on RI of biotechnologies held by stakeholders and established researchers, yet early career researchers have been largely overlooked. Therefore, we sought to assess and describe what early career researchers think about RI and the feasibility of incorporating it into their research and work. We centered the conversation around gene editing, gene drive, and genetic pest management. From their input we identified potential

challenges and opportunities, and categorized them into micro (individual), meso (institutional), and macro (wider societal) levels. The top three most referenced challenges were skills and knowledge, public engagement, and student authority. The top three most referenced opportunities were duty, public outreach, and collaboration. We conclude graduate education may be an ideal environment to train genetic engineering researchers in RI.

Background and Key Terms

Starting in the 1920s, intentional but random genetic changes could be made to organisms through mutagenesis. This process uses radiation or mutagenic chemicals to create random changes with the hopes that some of them will be useful. This has many applications which include the creation of new crop varieties (i.e. seedless watermelons, hard shell peanuts, barley for beer and whiskey) and new strains of insects (i.e. medfly sexing strain Vienna-42). To this day, mutagenesis is still widely used for plant breeding, while genetic modification of animals almost exclusively relies on genetic engineering. Unlike mutagenesis, genetic engineering is a non-random process and is thus more precise and controlled. Genetic engineering has been possible since the 1970s, with systems varying based on type of protein used. There are pros and cons to each system, and many are still used routinely for genetic engineering. Today the most common types of proteins used for genetic modification of insects include transposases (e.g. *piggyBac*), recombinases (e.g. PhiC31), homing endonucleases (e.g. Sce-I) and CRISPR-associated proteins (e.g. Cas9). The major difference in these proteins are the mechanisms by which the protein interacts with DNA and the variety of genome locations that are possible to manipulate. This dissertation includes the use of Cas9, *piggyBac*, and PhiC31 but mostly focuses on Cas9.

In 2012, CRISPR/Cas9 marked a major breakthrough in gene editing (Jinek et al. 2012). CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats. In the context of gene editing, this name is a bit of a misnomer since it describes the natural version of this system found in prokaryotes. In prokaryotes (bacteria and archaea), CRISPR/Cas9 is an adaptive immune system that defends against foreign pathogenic DNA (Terns and Terns 2011). The

CRISPR part of the system stores information that leads to acquired immunity and antiviral defense. Cas9 is an endonuclease, or protein that can cut DNA, and is the portion of the system that was developed as a tool for gene editing. Importantly, Cas9 is guided by a short RNA sequence which can be designed to match a desired DNA sequence. The first gene editing systems depended on endonucleases that have fixed recognition sequences. It is possible to use protein engineering to change a zinc finger endonuclease to fit a chosen recognition sequence, but this is a long, expensive, and arduous process. Compared to ZFNs, TALENs are easier to design and engineer, but they have more off-target binding (Mahfouz, Piatek, and Stewart 2014). Target binding or specificity of both ZFNs and TALENs depend on protein-DNA interactions that may be limited by repeat context dependence and epigenetic status. In contrast, the CRISPR/Cas9 system depends on Watson–Crick base pairing, which is highly predictable. Guide sequences can be manufactured within a week and purchased for less than one hundred dollars; a far cry from the years and millions of dollars required for protein engineering. Thus, genetic modification capabilities have rapidly advanced in the past nine years, adding to the genetic toolbox developed since the 1920s. Though of course, there are many challenges that keep gene editing with Cas9 from being a silver bullet. In the case of GPM, challenges are both technical and social in nature (Gutzmann et al. 2017).

This dissertation addresses some challenges of CRISPR/Cas9 and GPM at both the technical and social levels. Chapters 2, 3, and 4 cover technical issues like optimizing Cas9 efficiency and investigating a naturally occurring gene drive. Chapter 5 covers the social science framework RI and investigates early career researcher perceptions of RI and feasibility in genetic engineering research.

Technical Challenges of Genetic Pest Management

Although recent advancements in gene editing have increased the possibilities of GPM, many technical challenges remain. The first challenge of GPM addressed in this dissertation is the development of gene editing tools in pest species, in this case the optimization of

CRISPR/Cas9 in Tribolium beetles. The amount and quality of genetic resources or tools varies greatly between species. To date, most research on transgenic expression of *Cas9* has been in mosquitoes, flies, and the silkworm (Table 1.1). Therefore, we sought to optimize this tool in beetles.

There are some inherent limitations to using CRISPR/Cas9 as a gene editing system. While cutting DNA with Cas9 is highly efficient and ‘programmable’ using CRISPR guides (gRNA), the CRISPR/Cas9 machinery does not actually perform the gene editing step. Gene editing relies upon endogenous DNA repair pathways in the target organism after being cut by Cas9. Thus, CRISPR/Cas9 mediates gene editing but does not perform it. Deletions or knockouts rely on non-homologous end joining (NHEJ) and are easier to create since NHEJ is a more common repair pathway, and many different events can lead to the same desired phenotypic outcome. Insertions or substitutions rely on homology directed repair (HDR) which is not the preferred pathway. In addition to being less common, only one event where HDR occurs between the designed construct and chromosome will lead to the desired outcome. Thus, gene knockouts are easier to achieve than gene insertions, even in cases where tools have not been built or optimized for a particular species. For example, a plasmid expressing Cas9 via a Drosophila promoter might work well enough in a beetle to cut DNA and achieve a gene knockout. On the other hand, the same plasmid might not drive expression of Cas9 enough to achieve a gene insertion, which relies on a less common repair pathway, requires more components, and only one event leads to the desired outcome.

Chapter 2 addresses some challenges of transgenesis through the development of transgenic Cas9 *Tribolium* which will expedite development of similar tools in other non-model beetle species. Specifically, we detail the desirability of *polyubiquitin* and *alpha-tubulin* promoters for driving transgenic *Cas9* expression. This is the first Cas9 beetle, as so far all transgenic Cas9 insects have been flies, mosquitoes, or the silk worm (Table 1.1). Furthermore, we compare the most common modes of Cas9 delivery (protein, plasmid, and transgene) and rank their efficiency.

A ranking of these Cas9 delivery methods will inform the methodology of other researchers, especially those working in beetle systems.

Table 1.1

Transgenic Cas9 Insect Strains

Insect	Cas9 Promoter	Special Features	Reference
Aedes aegypti	*Exuperentia, 4-nitrophenylphosphatase, trunk, Nup50, polyubiquitin, ubiquitin L40*		Li et al. (2017)
	sds3, bgcn, and nup50	Gene drive	Verkuijl et al. (2020)
Anopheles gambiae	*vasa2*	Gene drive	Hammond et al. (2016)
	zpg, exu, nanos	Gene drive	Hammond et al. (2018)
	zpg, vasa2	Gene drive	Kyrou et al. (2018)
	zpg, vasa	Gene drive	Simoni et al. (2020)
Anopheles stephensi	*nanos*	Gene drive	Gantz et al (2015)
	nanos	Gene drive	Adolfi et al. (2020)
Bombyx mori	*IE1*		Li et al. (2015)
	IE1		Zeng et al. (2016)
	39K		Dong et al. (2018)
	nanos	W-specific Cas9	Zhang et al. (2018)
	nanos, IE1, actin3		Xu et al. (2019)

Note. Listed by species name and gene promoter in order of publication date. Also includes any special attributes like if it was a part of a gene drive or gene activation system.

Table 1.1 *(Continued)*

Insect	Cas9 Promoter	Special Features	Reference
Drosophila melanogaster	*nanos*		Kondo and Ueda (2013)
	nanos		Ren et al. (2013)
	vasa		Gratz et al. (2014)
	nanos, nanos 3'UTR		Port et al. (2014)
	nanos, nanos 5'UTR, nanos 3'UTR		Ren et al. (2014)
	vasa		Sebo et at (2014)
	vasa		Xue et al. (2014)
	actin5C		Zhang et al. (2014)
	UAS-Cas9 with *bamPGAL4*	UAS-Gal4	Chen et al. (2015)
	vasa	Gene drive	Gantz and Bier (2015)
	UAS-dCas9 with *dpp-Gal4*	UAS-Gal4, dCas9	Lin et al. (2015)
	nanos, vasa, actin5C		Port et al. (2015)*
	actin5C		Port and Bullock (2016)

Note. Listed by species name and gene promoter in order of publication date. Also includes any special attributes like if it was a part of a gene drive or gene activation system. *Compared the *D. melanogaster* Cas9 lines thus far.

Table 1.1 *(Continued)*

Insect	Cas9 Promoter	Special Features	Reference
Drosophila melanogaster	*vasa*	Lig4 deficient	Zimmer et al. (2016)
	nanos, *vasa*	Gene drive	Champer et al. (2017)
	nanos	Gene drive	Oberhofer et al. (2018)
	nanos	Gene drive	Oberhofer et al. (2019)
	Ubiquitin 63E, nanos, vasa	pgSIT	Kandul et al. (2019)
	nanos	Gene drive (TARE)	Champer et al. (2020)
	nanos	Gene drive (ClvR/TARE)	Oberhofer et al. (2020)
	Ubiquitin 63E, BicC, nanos, vasa	Gene drive	Kandul et al. (2020)
	exuL, Rcd1r, βTub	Gene drive (HomeR)	Kandul et al. (2021)
Tribolium castaneum	*polyubiquitin, alpha tubulin*		Gutzmann (Chapter 2)

Note. Listed by species name and gene promoter in order of publication date. Also includes any special attributes like if it was a part of a gene drive or gene activation system.

The second challenge of GPM addressed in this dissertation is a mechanism for spreading engineered genes through wild populations. Gene drives are able to push, or drive, a gene into a natural population faster than would be expected with normal Mendelian inheritance. There are two main GPM strategies that require gene drives. The first is population suppression (e.g by the spread of a lethal gene), sometimes referred to as insect birth control in the media (Manu 2013). Population suppression by genetic modification has been around since the 1950's, though at that time used radiation instead of lethal genes (Klassen et al. 2005). Just like UV rays can break DNA and cause skin cancer in humans, gamma rays can break DNA in insects and cause sterilization. Sterile Insect Technique (SIT) uses radiation to sterilize male insects before releasing them into the wild. Wild females then mate with sterile males and their eggs fail to hatch which leads to a population crash. This technique is still used around the world today, but is limited to certain species since some insects, like mosquitoes, cannot withstand radiation. Thus, lethal genes are being developed as another mechanism of insect birth control. In order for these lethal genes to spread through a population, they need to be accompanied by a gene drive.

The second GPM strategy that requires a gene drive is population replacement, where the goal is to make wild insects less pestiferous. For example, making the Malaria mosquito unable to transmit Malaria. It would be impossible to capture all mosquitoes and bring them to the lab for genetic engineering, so a gene drive is used to spread an engineered gene more quickly from a small percentage of engineered individuals into and through a wild population. Chapter 4 investigates a naturally occurring gene drive system that is found in the red flour beetle (*Tribolium castaneum*) which has been proposed by multiple scholars as a possible tool for GPM. To date, most gene drive research has been done in insects.

Social Challenges of Genetic Pest Management

For as long as genetic modification has been possible, debate and uncertainty around the technology have existed. Scientists at the famous Asilomar conference in 1975 expressed biohazard concerns around recombinant DNA. During this time, scientists put their research on

pause and deliberation at this conference resulted in guidelines for recombinant DNA research in the laboratory, published by the US government. However, these were just guidelines and regulation came much later. Additionally, much of the discussion on potential hazards at Asilomar focused on the worst-case scenario which is still the focus in many debates today. Though some general concerns remain the same, the topics most covered in the media have tracked popular areas of development. In the late 1980s through the 1990s, most applications were in agriculture, with genetically modified (GM) crops sparking a public controversy often called “the GMO debate.” However, since 2014, 90% of industry research using CRISPR/Cas9 gene editing has focused on human therapeutics. CRISPR/Cas9 human therapeutics companies constitute $10 billion USD of the NASDAQ, and the first clinical trials have been completed for a few diseases including sickle cell anemia (Demirci et al. 2018). Many current public conversations about gene editing are concerned over the possible slippery slope from disease treatment to eugenics (Walton 2017). In an event highly frowned upon by the global community, the first CRISPR babies were born in 2018. The scientist who performed the gene editing, He Jiankui, was heavily fined and sent to prison for this illegal medical practice, otherwise criticized as a “reckless ethical disaster” (Greely 2019). In short, genetic engineering has evolved alongside technology advancements, yet has always been a controversial topic and remains polarizing to this day.

There are three major differences between present day and the 1970s: 1) scientists generally regard this now commonplace technology as safe, 2) genetic engineering has become a commercial interest in addition to academic, and 3) governance and regulation of the technology evolved into formal systems of oversight for products of biotechnology. While the technology and applications have changed over the last 45 years, some concerns like bioweapon production, eugenics, and “playing God” have remained the same. With the advent of newer genetic modification methods, like human gene editing, CRISPR-based gene drives, and application of gene editing to a wider array of plants and animals, ethical issues are gaining more attention. Today those ethical questions are coming back to the forefront, perhaps in response to continued

public distrust and protest over the technology. Generally, scientists regard genetic engineering as safe, but some have come to question their ability to self-govern the technology and feel the public should be involved in risk assessments (Barinaga 2000).

Questions have also been raised about how to incorporate ideas of risk, ethics, and responsibility into research practice. For example, world authorities on gene drive have written collaborative papers on how to safeguard use of the technology in terms of research ethics and ecological risk (Akbari et al. 2015, Champer et al. 2019). Another paper on gene drive risk management proposed multidisciplinary methods to explore societal and ethical questions (Lunshof and Birnbaum 2017). For example, the authors highlight the importance of including human well-being in risk management and propose genetic engineering teams work with "ethics consultants" to address this need. Alternatively, a paper on global governance of gene drives cited public inclusion and deliberation as essential (Kofler et al. 2018).

As proposed in this dissertation, the question of if and how to include relevant publics and stakeholders can perhaps be addressed by the RI framework. The RI framework was developed in 2003 and began gaining traction in policy discourse in the EU in 2010 (Hellström 2003, Owen, Macnaghten & Stilgoe 2012). Multiple versions of the RI framework have been described in the literature, with the most common having four components: anticipation, reflexivity, inclusion and responsiveness. Common threads exist between RI and GPM discourse; for example, elements of RI, though not explicitly referenced, were included in several ethical and social discussions of GPM (Resnik 2014, Lavery 2008). The 2016 report titled 'Gene Drives on the Horizon' by the US National Academy of Sciences, Engineering, and Medicine (NASEM) covered topics including human values, public and stakeholder engagement, risk, and governance. These topics have considerable overlap with the four pillars of RI, but RI takes them further in giving various publics voice in the research and innovation process, responding to their concerns, including them in the anticipation of risks, and asking researchers to reflect on the underlying purposes, motivations, and goals of innovation (Stilgoe et al. 2013).

In a few cases, the RI framework has been investigated with respect to GPM. For example, one paper applied the RI framework to the release of GM mosquitoes in Brazil (De Campos et al. 2017). It has also been applied to spotted wing drosophila, an invasive pest of berries, for which pest management efforts have focused on GM sterile males, gene drive, and RNAi yeast (Kokotovich et al. 2020). Principles of RI were also incorporated into a framework for more “procedurally robust risk assessments” for GM mosquitos and gene drive organisms (Kuzma 2019). Chapter 5 considers if and how the RI framework could help early-career genetic engineers to meet the goals of transparency and aligning development with public needs.

In addition to the academic value of answering these questions, the practical value can be seen through a couple of case examples of GM insect research and deployment. In 2016, GM mosquitoes were approved for release by the Florida Keys Mosquito Control District (FKMCD) and the US FDA. These GM mosquitoes are engineered to prevent the spread of dengue fever and Zika virus and had successful field trials in Brazil and Grand Cayman (Carvalho et al. 2015, Harris et al. 2012). The FKMCD began considering the use of GM mosquitoes in 2009 but at the time made no strategic attempts at public involvement due to regulatory uncertainty (Doyle 2016). In 2012 the FKMCD made the decision to pursue GM mosquitos and began informing the public through radio shows, local newspapers, and public events. A 2013 survey to gauge public opinion reported 60% of local residents supported the use of GM mosquitos, 17% were neutral, and 23% were opposed (Ernst et al 2015). However, despite legal approval and support by a majority of residents, the field trail in Key West was canceled due to protests. Opposition was mainly driven by three influential residents and national environmental organizations (Doyle 2016). A petition from Change.org had many signatures of which 83% were from non-Floridian residents and only 0.35% from residents of Key West. Former executive director of FKMCD Michael Doyle shared three takeaways from this experience: 1) they shouldn’t have assumed that support from a numerical majority equaled the public perception of a majority, 2) the public should have been provided data and answers earlier in the process, and 3) they should have met with individual

residents before the opposition group did. The RI framework could have mitigated these challenges through upstream public engagement and reflection of assumptions by experts. Following a surge of dengue fever in 2020, GM mosquitoes have been reapproved and releases are slated to begin in 2021.

RI could have also been useful in the case of GM diamond back moth, a global agricultural pest that causes $4-5 billion USD in damages annually (Zalucki et al. 2012). In 2015, Cornell University conducted caged field trials of GM diamond back moth with Bt broccoli, a GM crop. The authors concluded GM moths could delay resistance to Bt, an insecticide commonly used in GM crops (Harvey-Samuel et al. 2015). Bt resistance in beetles is investigated in Chapter 3. In 2017, Cornell obtained a permit for open field trials from the USDA-APHIS-BRS and an Environmental Assessment yielded a Finding of No Significant Impact (FONSI). Despite legal approval, open field trials were postponed following vandalism and political activism against the open field trial which was later completed in 2017 (Shelton et al. 2020). While the RI framework should not be used as a tool to gain public buy-in, it can be used to establish a relationship between the community and researchers to deliberate alternative perspectives, values, and preferences in governance of the technology.

In summary, insects contribute vital services to ecosystems and also play an antagonistic role as herbivores and vectors of human disease. Some emerging pest management strategies make use of genetic engineering to decrease pest populations or make them less harmful. In the US, genetic engineers are aware of the precedence previous GM products like GM crops have set for GM insects regarding public perception and acceptance. Developers are making an effort to avoid mistakes made with the 1st generation of GMOs by focusing on transparency and science communication. Goals are to ensure GM technologies are developed in ways that benefit society while mitigating risks to consumers and the environment. One way to do this is to have more inclusive conversations between developers and the public. Towards this goal, the RI framework may be useful.

REFERENCES

Adolfi, A., Gantz, V., Jasinskiene, N., Lee, H., Hwang, K., Terradas, G., Bulger, E., Ramaiah, A., Bennett, J., Emerson, J., Marshall, J., Bier, E. & James, A. (2020). Efficient population modification gene-drive rescue system in the malaria mosquito Anopheles stephensi. *Nature Communications*, 11(1).

Akbari, O. S., Bellen, H. J., Bier, E., Bullock, S. L., Burt, A., Church, G. M., G.M., Cook, K.R., Duchek, P., Edwards, O.R., Esvelt, K.M., Gantz, V.M., Golic, K. G., Gratz, S. J., Harrison, M. M., Hayes, K. R., James, A. A., Kaufman, T. C., Knoblich, J., Malik, H. S., Matthews, K. A., O'Connor-Giles, K. M., Parks, A. L. Perrimon, N., Port, P., Russell, S., Ueda, R., & Wildonger, J. (2015). Safeguarding gene drive experiments in the laboratory. *Science*, *349*(6251), 927-929.

Barinaga, M. (2000). Asilomar revisited: lessons for today?. *Science*, *287*(5458), 1584-1585.

Brown, S., Shippy, T., Miller, S., Bolognesi, R., Beeman, R., & Lorenzen, M. et al. (2009). The Red Flour Beetle, Tribolium castaneum (Coleoptera): A Model for Studies of Development and Pest Biology. *Cold Spring Harbor Protocols*, *2009*(8), pdb.emo126-pdb.emo126. doi: 10.1101/pdb.emo126

Carvalho, D., McKemey, A., Garziera, L., Lacroix, R., Donnelly, C., & Alphey, L. et al. (2015). Suppression of a Field Population of Aedes aegypti in Brazil by Sustained Release of Transgenic Male Mosquitoes. *PLOS Neglected Tropical Diseases*, *9*(7), e0003864. doi: 10.1371/journal.pntd.0003864

Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G., & Messer, P. W. (2017). Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS genetics*, *13*(7), e1006796.

Champer, J., Chung, J., Lee, Y., Liu, C., Yang, E., Wen, Z., Clark, A. and Messer, P. (2019). Molecular safeguarding of CRISPR gene drive experiments. *eLife*, 8.

Champer, J., Lee, E., Yang, E., Liu, C., Clark, A. G., & Messer, P. W. (2020). A toxin-antidote CRISPR gene drive system for regional population modification. *Nature communications*, *11*(1), 1-10.

Chen, H. M., Huang, Y., Pfeiffer, B. D., Yao, X., & Lee, T. (2015). An enhanced gene targeting toolkit for Drosophila: Golic+. *Genetics*, *199*(3), 683-694.

Demirci, S., Uchida, N., & Tisdale, J. F. (2018). Gene therapy for sickle cell disease: An update. *Cytotherapy*, *20*(7), 899-910.

De Campos, A. S., Hartley, S., de Koning, C., Lezaun, J., & Velho, L. (2017). Responsible Innovation and political accountability: genetically modified mosquitoes in Brazil. *Journal of Responsible Innovation*, *4*(1), 5-23.

Dong, Z., Huang, L., Dong, F., Hu, Z., Qin, Q., Long, J., Cao, M., Chen, P., Lu, C. and Pan, M. (2018). Establishment of a baculovirus-inducible CRISPR/Cas9 system for antiviral research in transgenic silkworms. *Applied Microbiology and Biotechnology*, 102(21), pp.9255-9265.

Doyle, M. (2016, November). Public Engagement about Potential Release of GE Mosquitoes: Lessons Learned. Oral presentation at the Genetic Engineering and Society Center, NCSU, Raleigh, NC. Available online at https://vimeo.com/191714601.

Ernst, K. C., Haenchen, S., Dickinson, K., Doyle, M. S., Walker, K., Monaghan, A. J., & Hayden, M. H. (2015). Awareness and support of release of genetically modified "sterile" mosquitoes, Key West, Florida, USA. *Emerging infectious diseases*, *21*(2), 320.

Esvelt, K. M., Smidler, A. L., Catteruccia, F., & Church, G. M. (2014). Emerging technology: concerning RNA-guided gene drives for the alteration of wild populations. *elife*, *3*, e03401.

FAO. "New Standards to Curb the Global Spread of Plant Pests and Diseases." *Food and Agriculture Organization of the United Nations*, 3 Apr. 2019, www.fao.org/news/story/en/item/1187738/icode/#:~:text=FAO%20estimates%20that%20annually%20between,insects%20around%20US%2470%20billion.

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences*, *112*(49), E6736-E6743.

Gantz, V. M., & Bier, E. (2015). The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*, *348*(6233), 442-444.

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., & O'Connor-Giles, K. M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics*, *196*(4), 961-971.

Greely, H. T. (2019). CRISPR'd babies: human germline genome editing in the 'He Jiankui affair'. *Journal of Law and the Biosciences*, *6*(1), 111-183.

Gutzmann, N., Elsensohn, J. E., Barnes, J. C., Baltzegar, J., Jones, M. S., & Sudweeks, J. (2017). CRISPR-based gene drive in agriculture will face technical and governance challenges. *EMBO reports*, *18*(9), 1479-1480.

Hammond, A., Galizi, R., Kyrou, K., Simoni, A., Siniscalchi, C., Katsanos, D., Gribble, M., Baker, D., Marois, E., Russell, S., Burt, A., Windbichler, N., Crisanti, A. and Nolan, T. (2016). A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nature Biotechnology*, 34(1), pp.78-83.

Hammond, A., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Kranjc, N., Carpi, F., D'Aurizio, R., Crisanti, A. and Nolan, T. (2017). The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLOS Genetics*, 13(10), p.e1007039.

Hammond, A., Kyrou, K., Gribble, M., Karlsson, X., Morianou, I., Galizi, R., Beaghton, A., Crisanti, A., & Nolan, T. (2018). Improved CRISPR-based suppression gene drives mitigate resistance and impose a large reproductive load on laboratory-contained mosquito populations. *BioRxiv*.

Harris, A., McKemey, A., Nimmo, D., Curtis, Z., Black, I., Morgan, S., Oviedo, M., Lacroix, R., Naish, N., Morrison, N., Collado, A., Stevenson, J., Scaife, S., Dafa'alla, T., Fu, G., Phillips, C., Miles, A., Raduan, N., Kelly, N., Beech, C., Donnelly, C., Petrie, W. and Alphey, L. (2012). Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes. *Nature Biotechnology*, 30(9), pp.828-830.

Harvey-Samuel, T., Morrison, N. I., Walker, A. S., Marubbi, T., Yao, J., Collins, H. L., Davies T. E., Alphey N., Warner S., & Shelton, A. M. (2015). Pest control and resistance management through release of insects carrying a male-selecting transgene. *BMC biology*, *13*(1), 49.

Hellström, T. (2003). Systemic innovation and risk: technology assessment and the challenge of responsible innovation. *Technology in Society*, *25*(3), 369-384.

Jin, L., Walker, A., Fu, G., Harvey-Samuel, T., Dafa'alla, T., Miles, A., Marubbi, T., Granville, D., Humphrey-Jones, N., O'Connell, S., Morrison, N. and Alphey, L. (2013). Engineered Female-Specific Lethality for Control of Pest Lepidoptera. *ACS Synthetic Biology*, 2(3), pp.160-166.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816-821.

Joseph, Manu. "New Delhi's Dengue Fever Problem." *The New York Times*, The New York Times, 25 Sept. 2013, www.nytimes.com/2013/09/26/world/asia/new-delhis-dengue-fever-problem.html.

Kandul, N. P., Liu, J., Wu, S. L., Marshall, J. M., & Akbari, O. S. (2019). Transforming insect population control with precision guided sterile males with demonstration in flies. *Nature communications*, *10*(1), 1-12.

Kandul, N. P., Liu, J., Buchman, A., Gantz, V. M., Bier, E., & Akbari, O. S. (2020). Assessment of a split homing based gene drive for efficient knockout of multiple genes. *G3: Genes, Genomes, Genetics*, *10*(2), 827-837.

Kandul, N. P., Liu, J., Bennett, J. B., Marshall, J. M., & Akbari, O. S. (2021). A confinable home-and-rescue gene drive for population modification. *Elife*, *10*, e65939.

Klassen, W., Curtis, C. F., Dyck, V. A., Hendrichs, J., & Robinson, A. S. (2005). History of the sterile insect technique. *Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management. Dordrecht, The Netherlandfs: Springer*, 1-34.

Kofler, N., Collins, J., Kuzma, J., Marris, E., Esvelt, K., Nelson, M., Newhouse, A., Rothschild, L., Vigliotti, V., Semenov, M., Jacobsen, R., Dahlman, J., Prince, S., Caccone, A., Brown, T. and Schmitz, O. (2018). Editing nature: Local roots of global governance. *Science*, 362(6414), pp.527-529.

Kondo, S., & Ueda, R. (2013). Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. *Genetics*, *195*(3), 715-721.

Kyrou, K., Hammond, A., Galizi, R., Kranjc, N., Burt, A., Beaghton, A., Nolan, T. and Crisanti, A. (2018). A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes. *Nature Biotechnology*, 36(11), pp.1062-1066.

Li, Z., You, L., Zeng, B., Ling, L., Xu, J., Chen, X., Zhang, Z., Palli, S., Huang, Y. and Tan, A. (2015). Ectopic expression of ecdysone oxidase impairs tissue degeneration in Bombyx mori. *Proceedings of the Royal Society B: Biological Sciences*, 282(1809), p.20150513.

Li, M., Bui, M., Yang, T., Bowman, C. S., White, B. J., & Akbari, O. S. (2017). Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti. *Proceedings of the National Academy of Sciences*, *114*(49), E10540-E10549.

Lin, S., Ewen-Campen, B., Ni, X., Housden, B. E., & Perrimon, N. (2015). In vivo transcriptional activation using CRISPR/Cas9 in Drosophila. *Genetics*, *201*(2), 433-442.

Lunshof, J. E., & Birnbaum, A. (2017). Adaptive risk management of gene drive experiments: biosafety, biosecurity, and ethics. *Applied Biosafety*, *22*(3), 97-103.

Mahfouz, M. M., Piatek, A., & Stewart Jr, C. N. (2014). Genome engineering via TALENs and CRISPR/Cas9 systems: challenges and perspectives. *Plant biotechnology journal*, *12*(8), 1006-1014.

Oberhofer, G., Ivy, T., & Hay, B. A. (2020). Gene drive and resilience through renewal with next generation Cleave and Rescue selfish genetic elements. *Proceedings of the National Academy of Sciences*, *117*(16), 9013-9021.

Oberhofer, G., Ivy, T., & Hay, B. A. (2018). Behavior of homing endonuclease gene drives targeting genes required for viability or female fertility with multiplexed guide RNAs. *Proceedings of the National Academy of Sciences*, *115*(40), E9343-E9352.

Oberhofer, G., Ivy, T., & Hay, B. A. (2019). Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive. *Proceedings of the National Academy of Sciences*, *116*(13), 6250-6259.

Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and public policy*, *39*(6), 751-760.

Port, F., Chen, H. M., Lee, T., & Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proceedings of the National Academy of Sciences*, *111*(29), E2967-E2976.

Port, F., Muschalik, N., & Bullock, S. L. (2015). Systematic evaluation of Drosophila CRISPR tools reveals safe and robust alternatives to autonomous gene drives in basic research. *G3: Genes, Genomes, Genetics*, *5*(7), 1493-1502.

Port, F., & Bullock, S. L. (2016). Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nature methods*, *13*(10), 852.

Ren, Xingjie, Jin Sun, Benjamin E. Housden, Yanhui Hu, Charles Roesel, Shuailiang Lin, Lu-Ping Liu et al. "Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9." *Proceedings of the National Academy of Sciences* 110, no. 47 (2013): 19012-19017.

Ren, X., Yang, Z., Mao, D., Chang, Z., Qiao, H., Wang, X., Sun, J., Hu, Q., Cui, Y., Liu, L., Ji, J., Xu, J. and Ni, J. (2014). Performance of the Cas9 Nickase System in Drosophila melanogaster. *G3: Genes|Genomes|Genetics*, 4(10), pp.1955-1962.

Sebo, Z. L., Lee, H. B., Peng, Y., & Guo, Y. (2014). A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. *Fly*, *8*(1), 52-57.

Shelton, A. M., Long, S. J., Walker, A. S., Bolton, M., Collins, H. L., Revuelta, L., Johnson, L. M. & Morrison, N. I. (2020). First Field Release of a Genetically Engineered, Self-Limiting Agricultural Pest Insect: Evaluating Its Potential for Future Crop Protection. *Frontiers in Bioengineering and Biotechnology*, *7*, 482.

Simoni, A., Hammond, A., Beaghton, A., Galizi, R., Taxiarchi, C., Kyrou, K., Meacci, D., Gribble, M., Morselli, G., Burt, A., Nolan, T., & Crisanti, A. (2020). Author Correction: A male-biased sex-distorter gene drive for the human malaria vector Anopheles gambiae. *Nature Biotechnology*, 38(9), pp.1097-1097.

Terns, M. P., & Terns, R. M. (2011). CRISPR-based adaptive immune systems. *Current opinion in microbiology*, *14*(3), 321-327.

USDA. "Recent Trends in GE Adoption." USDA ERS - Recent Trends in GE Adoption. July 12, 2017. Accessed November 17, 2017. http://www.ers.usda.gov/data-products/adoption-of-genetically-engineered-crops-in-the-us/recent-trends-in-ge-adoption/.

Verkuijl, S. A., Gonzalez, E., Li, M., Ang, J., Kandul, N. P., Anderson, M. A., Akbari, O. S., Bonsall, M. B., & Alphey, L. (2020). A CRISPR endonuclease gene drive reveals two distinct mechanisms of inheritance bias. *bioRxiv*.

Walton, D. (2017). The slippery slope argument in the ethical debate on genetic engineering of humans. *Science and engineering ethics*, *23*(6), 1507-1528.

WHO. "Vector-Borne Diseases." *World Health Organization*, World Health Organization, 2 Mar. 2020, www.who.int/news-room/fact-sheets/detail/vector-borne-diseases.

Xu, J., Chen, R., Chen, S., Chen, K., Tang, L., Yang, D., Yang, X., Zhang, Y., Song, H. and Huang, Y. (2019). Identification of a germline-expression promoter for genome editing in Bombyx mori. *Insect Science*, 26(6), pp.991-999.

Xue, Z., Ren, M., Wu, M., Dai, J., Rong, Y. S., & Gao, G. (2014). Efficient gene knock-out and knock-in with transgenic Cas9 in Drosophila. *G3: Genes, Genomes, Genetics*, *4*(5), 925-929.

Zalucki, M. P., Shabbir, A., Silva, R., Adamson, D., Shu-Sheng, L., & Furlong, M. J. (2012). Estimating the economic cost of one of the world's major insect pests, Plutella xylostella (Lepidoptera: Plutellidae): just how long is a piece of string?. *Journal of economic entomology*, *105*(4), 1115-1129.

Zeng, B., Zhan, S., Wang, Y., Huang, Y., Xu, J., Liu, Q., Li, Z., Huang, Y. and Tan, A. (2016). Expansion of CRISPR targeting sites in Bombyx mori. *Insect Biochemistry and Molecular Biology*, 72, pp.31-40.

Zhang, X., Koolhaas, W. H., & Schnorrer, F. (2014). A versatile two-step CRISPR-and RMCE-based strategy for efficient genome engineering in Drosophila. *G3: Genes, Genomes, Genetics*, *4*(12), 2409-2418.

Zhang, Z., Niu, B., Ji, D., Li, M., Li, K., James, A., Tan, A. and Huang, Y. (2018). Silkworm genetic sexing through W chromosome-linked, targeted gene integration. *Proceedings of the National Academy of Sciences*, 115(35), pp.8752-8756.

Zimmer, C. T., Garrood, W. T., Puinean, A. M., Eckel-Zimmer, M., Williamson, M. S., Davies, T. E., & Bass, C. (2016). A CRISPR/Cas9 mediated point mutation in the alpha 6 subunit of the nicotinic acetylcholine receptor confers resistance to spinosad in Drosophila melanogaster. *Insect biochemistry and molecular biology*, *73*, 62-69.

CHAPTER 2

Improved Gene Editing Efficiency in *Tribolium castaneum* With Constitutive Expression of Cas9 Transgene

Introduction

In the short time since the adaptation of CRISPR/Cas9 as a gene editing technology, it has revolutionized many agricultural fields. The new accessibility of pest insect gene editing has increased the feasibility of studies investigating development, evolution, behavior, molecular biology, and gene function (Tanning et al. 2017). The comparative advantages of gene editing with Cas9 have led to rapid adoption, though it is not without limitations. The most significant limitations for applications in insects are the delivery of molecular components into eggs, lab rearing, and availability of genomic data. Second to this is the need for optimization of expression from constructs or transgenes in different species, including the identification of promoters to drive expression of *Cas9*. To overcome this limitation, we sought to test two promoters for driving *Cas9* expression in an insect model, the red flour beetle (*Tribolium castaneum)*. We used transgenic *Cas9* to help alleviate the constraint of a small injection volume for homology directed repair constructs. In other words, using transgenic *Cas9* means more injection mix volume can be dedicated to guides and knock-in constructs, which will make gene alterations or insertions easier. Additionally, transgenic *Cas9* is crucial for Cas9-mediated gene drive, with potential use in genetic pest management.

Cas9 mediated gene editing is dependent on the efficiency of Cas9 cutting, effectiveness of the guide, and the DNA repair pathways of the insect. In this study we focus on improving the efficiency of Cas9 cutting. This is affected by the delivery method of Cas9, and the Cas9 promoter when applicable. Cas9 nuclease can be provided in many forms (DNA construct, mRNA, or purified protein), as can the guides (DNA construct, gRNA). Gene editing reagents are usually delivered by microinjecting into insect embryos prior to cellularization. The timing of injections and reagent concentration and form (DNA, RNA, or protein) all affect the success rate of transgenesis. Concentrations should be high enough to result in heritable gene editing but not so high as to be toxic. For the best chance at making a heritable gene edit, the insect eggs should be injected at an early developmental stage, pre-blastoderm, in the first minutes to hours of being laid. Timing

of the pre-blastoderm stage varies by species. When Cas9 is provided via a plasmid, cutting does not happen immediately upon injection. There is a lag time while the embryo uses its own cell machinery to transcribe off of the injected plasmid DNA and translate that message into Cas9 protein. In the past five years, researchers have improved issues with timing, with one of the biggest improvements thus far being the use of a primed ribonucleoprotein complex (RNP; Cas9 protein pre-complexed with single-guide RNAs) (Woo et al. 2015). Another approach is to use transgenic Cas9 to ensure the presence of active Cas9 at the pre-blastoderm stage.

The ideal gene promoter for driving expression of *Cas9* would address both timing and concentration issues. Germline specific promoters have been used for this purpose in multiple insect species and a summary of transgenic Cas9 insects is described in Table 1.1 (page 9-10). However, germline specific promoters have not been used for driving transgenes in Tribolium. Our next best option would be a ubiquitous promoter known to be active in early embryos, and it would be an added bonus if transcripts are maternally preloaded. In early egg development, insect mothers load the egg with important transcripts. In *Anopheles stephensi* and *Drosophila melanogaster*, transgenic *Cas9* mothers have been found to load *Cas9* into eggs when using the promoter *nanos* or *vasa* (Gantz et al. 2015, Kondo and Ueda 2013, Sebo et at. 2014, Champer et al. 2017). In Tribolium, *polyubiquitin* (*TcPub*) and *alpha-Tubulin* (*TcαTub*) transcripts are preloaded into embryos and the respective promoters have already been validated for driving transgenes (Lorenzen et al. 2002, Siebert et al. 2008). This is our reasoning for creating *TcPubCas9* and *TcαTubCas9*.

Methods of doing CRISPR gene editing in insects have changed drastically in the past six years. Optimizing the gene editing system involved finding appropriate promoters for driving expression of Cas9 and guide RNAs. Initially, plasmids containing a *Cas9* gene were used because Cas9 protein was not yet available on the market. When Cas9 first became available, mostly through small start-ups, it was of questionable quality. Now reliable Cas9 protein is available, though the high price may still make it unavailable to some labs. Guide plasmids were

used instead of synthetic sgRNA for the same reason. However, while sgRNAs were not yet available for purchase, they could be made in-house. Thus, our aims were two-fold: create transgenic Cas9 *Tribolium* to improve the efficiency of gene editing and compare their efficiency with multiple methods of Cas9 delivery that have changed over the years.

To achieve these goals we developed the first transgenic Tribolium lines that express *Cas9*, testing Cas9 expression under the control of two different promoters, *polyubiquitin* (*TcPub*) and *alpha-Tubulin* (*TcαTub*). Moreover, we test the hypotheses that transgenic expression via constitutively active promoters would ensure the presence of Cas9 in younger embryos than is possible via microinjection and therefore increase Cas9 cutting efficiency and gene editing. Herein we also compare gene editing rates between five *PubCas9* and five *αTubCas9* lines, each with different genomic locations of the transgene. Gene editing rates were additionally compared to microinjected Cas9 protein, *PubCas9* plasmid, *αTubCas9* plasmid, and the previously published *bhsp68Cas9* plasmid (Gilles et al. 2015).

Tribolium is a worldwide pest of stored grains and considered a more representative model of pest insects than the highly specialized model insect, *Drosophila melanogaster* (Brown et al. 2009). Moreover, Tribolium is a model for studying molecular biology and gene function in insects, is easy to rear, and has a high-quality assembled genome and a wide range of transgenic tools. Here we report the first development of a transgenic Cas9 line in a beetle. To date, Cas9 lines have been developed for a number of fly, mosquito, and moth species and tools and data from these studies are not directly transferable to beetles. Therefore, results from Tribolium can inform development of CRISPR/Cas9 tools in many other economically significant beetle species.

Methods

Tribolium strains

The strains used in this work are as follows: *vermillion*^{white} (v^{W}), a white-eyed mutant strain described in Lorenzen et al. (2002); M26, a transgenic helper strain that expresses *piggyBac transposase*, described in Lorenzen et al. (2007); and A19, an EGFP-marked muscle-enhancer

trap strain described in Lorenzen et al. (2003). It should be noted that the transgenic strains are in different genetic backgrounds. Specifically, the M26 strain used was homozygous for the recessive v^w eye-color mutation, while the A19 strain was homozygous for a different eye-color mutation, *pearl* (*p*) (Park, 1937). Importantly, while both mutant strains exhibit white-eyed phenotypes, the mutations are in different genes so crossing them results in heterozygous individuals that display wild type (near black) eye pigmentation.

Insect rearing

Beetle rearing protocols developed by the USDA ARS were followed (Brown et al. 2009). Diet is created by mixing 9.5 cups Bay State Milling Co. (Quincy, MA, USA) organic whole wheat flour with 0.5 cups MP Biomedicals brewer's yeast (Cat#903312) then sifted with a #25 brass sieve to remove large particles. Flour is frozen for 24hrs upon delivery, and both flour and brewer's yeast are stored at 4ºC. To make special diet for doing egg collections, normal diet is sifted with a #30 mesh sieve. Both normal and egg-lay flour diets are stored in sealed containers at room temperature for up to 2 months. Beetle colonies are maintained in pint or half-pint glass mason jars and sub-cultured by paper transfer every 3-4 weeks. Colonies are stored in a dark incubation chamber set to 30ºC and ~70% humidity.

Plasmid DNAs

Five plasmids were used in this work: two *piggyBac* donor plasmids *pB(TcPubCas9-3xP3EGFP)* and *pB(TcαTubCas9-3xP3EGFP)*, two sgRNA guide plasmids *p(U6a-EGFPG2)* and *p(U6a-EGFPG3),* and a Cas9 plasmid created by Gilles et al. (2015) *p(bhsp68-Cas9)* (Addgene #65959). The donor plasmids contain a *Streptococcus pyogenes*-derived and human-codon optimized Cas9 (Cong et al. 2013), a monopartite nuclear localization signal (NLS) from the SV40 Large T-antigen at the N-terminus, and a nucleoplasmin NLS at the C-terminus.

The two guide plasmids were created in a guide plasmid backbone *p(U6a-BsaI-gRNA)* created by Gilles et al. (2015) (Addgene #65955). The vector *p(U6a-BsaI-gRNA)* was cloned using homemade TOP10 competent cells and purified using QIAprep spin mini prep kit (Qiagen,

Valencia, CA, USA). The vector was then restriction digested with BsaI (New England Biolabs, Ipswich, MA, USA) and purified using the Monarch PCR & DNA Clean-up Kit (New England Biolabs). EGFP guide sequences were designed and validated during a previous project: EGFP Guide 2 and EGFP Guide 3 (Table 2.1). For each guide, two oligonucleotides (Integrated DNA Technologies, Coralville, Iowa, USA) were designed with BsaI overhangs created by an additional AGTG on the 5' end of the forward primer and an additional AAAC on the 5' end of the reverse primer (Table 2.1). To anneal the forward and reverse oligos, 10μl of 2x annealing buffer (20mM Tris, 2mM EDTA, 100 mM NaCl, pH 8.0) was mixed with 5μl each 100μM oligo in a 20 μl reaction volume and the following PCR conditions: 98°C for 1min, 98-88°C for 5 s, decrease 0.1°C/cycle x99 cycles, 88-38°C for 10 s, decrease 0.1°C/cycle x499 cycles, 38-18°C for 1 s, decrease 0.1°C/cycle x99 cycles, and 18°C indefinitely. Annealed guide oligos were then phosphorylated using T4 Polynucleotide Kinase (New England Biolabs) with 1μl annealed oligos in a 20μl reaction volume. Phosphorylated oligos were diluted with 30 μl ddH_20. Guide oligos (1 μl of 5X diluted) were ligated into the vector (1 μl BsaI digested, purified) using T4 DNA ligase (New England Biolabs) in a total reaction volume of 10 μl for 3.5 hours at room temperature. The guide plasmids *p(U6a-EGFPG2)* and *p(U6a-EGFPG3)* were cloned with TOP10 competent cells and purified using a Plasmid DNA Midi Kit (Qiagen) and eluted with microinjection grade water.

To create the two *piggyBac* donor plasmids *pB(TcPubCas9-3xP3EGFP)* and *pB(TcαTubCas9-3xP3EGFP)*, the promoters were cloned into a vector backbone *pB(AscI-AflII-Cas9-3xP3EGFP)* built by Dr. Fu-Chyun Chu (unpublished) using *pBac(3xP3-EGFPaf)* from Horn and Wimmer (2000). Primers with restriction sites were used to amplify the *TcαTub* and *TcPub* promoter sequences such that the promoters had an AscI site on the 5' end and an AflII site on the 3' end (Table 2.1). Promoters were amplified in 50 μl PCR reactions with 5 μl 10X PCR buffer, 4 μl dNTPs, 2 μl of each 25 μM primer, 0.25 μl ExTaq DNA polymerase (Takara Bio, Kusatsu, Shiga, Japan) and 1 ng plasmid template. Previously built plasmids containing the promoters were used as template. Conditions followed manufacturer suggestions with an initial denaturation

at 98°C for 2 min and final elongation at 72°C for 5 min. Products were purified using Monarch PCR & DNA Clean-up Kit (New England Biolabs, Ipswich, MA, USA) and TA cloned into the pGEM-T Easy vector (Promega, Durham, NC, USA) with a 1 hr ligation at room temperature. TOP10 competent cells were transformed with either *pGem(TcPub)* or *pGem(TcαTub)* and the plasmid DNAs purified using QIAprep spin mini prep kit (Qiagen, Valencia, CA, USA).

The vector backbone *pB(AscI-AflII-Cas9-3xP3EGFP)*, *pGem(TcPub)* and *pGem(TcαTub)* were each digested with AscI and AflII (New England Biolabs). The vector backbone was purified using the Monarch PCR & DNA Clean-up Kit (New England Biolabs) and the fragments containing promoter inserts were purified using the Monarch Gel Extraction Kit (New England Biolabs). The promoter sequence inserts (8 ng) were then ligated into the digested vector (25ng) using T4 DNA ligase (New England Biolabs) in a 10 µl reaction volume for 1 hour at room temperature. TOP10 competent cells were transformed with the ligation products and plasmid DNA purified using a Plasmid DNA Midi Kit (Qiagen) and eluted with microinjection grade water. For SnapGene feature map and full plasmid sequences see Supplementary Material 2.10, 2.11, and 2.12.

Table 2.1

Primer and guide sequences

Name	Sequence	Anneal (°C)
Tc aTub-AscI F	GGCGCGCCCACACTGCAGTGAACGGT	55
Tc aTub-AflII R	CTTAAGACAAAAGAGCAGAAAAACGG	55
TcPub AscI- F	TAGGCGCGCCTATTGTCGTCCGTATTTACA	58
Tc Pub- AflII R	CTTAAGTACGACCTGTCTCACCACTTTTG	58
Cas9 Check F1	CGAAACTAATGGGGAAACTGG	-
Cas9 Check F2	TTATTCAGTCCTAGTGGTTGC	-
Cas9 Check R1	TGCTTATGCTGCTCCACAAAC	-
Cas9 Check R2	GTGATGGATTGATGGATAAGA	-
Cas 9 Check F6	GCAGAAGGGAAACGAACTGGC	-
EGFP G2 F	AGTGACCAGGATGGGCACCACCC	touchdown
EGFP G2 R	AAACGGGTGGTGCCCATCCTGGT	touchdown
EGFP G3 F	AGTGCACCGGGGTGGTGCCCATCC	touchdown
EGFP G3 R	AAACGGATGGGCACCACCCCGGTG	touchdown
EGFP G3 sgRNA	CACCGGGGUGGUGCCCAUCC	-

Note. Annealing temperature shown for PCR but not sequencing primers.

Generation of transgenic Tribolium carrying Cas9 transgene

To create transgenic beetle lines carrying a *Cas9* transgene, we used the *piggyBac transposase* transformation system (Cary et al. 1989, Tamura et al. 2000). Our *piggyBac* donor plasmids contained the *Cas9* gene and the eye-specific fluorescence marker, *3xP3-EGFP*, flanked by *piggyBac* recognition arms (R1 and L1). Donor plasmid was co-injected with *piggyBac*

transposase helper plasmid (phspBac; Handler and Harrell 1999) into a Tribolium *piggyBac transposase* helper line, M26 (Lorenzen et l. 2007). Helper plasmid was used in addition to M26 to increase likelihood of successful transformation. Transformation efficiency is inversely correlated with the size of genetic cargo, and *Cas9* is considered relatively large at ~5Kb. The total length of the insert with marker and *piggyBac* recognition arms is 8.9Kb for *PubCas9* and 8.7Kb for *αTubCas9*. During this process, *piggyBac transposase* finds recognition arms in the donor plasmid, cuts the DNA at the ends of these arms, then inserts the arms and genetic cargo (*Cas9* and green-eye marker) at a TTAA-target site at a random location in the genome.

Microinjection for Insertion of Cas9 transgene

For egg collection, 200-300 adults were placed in a jar of egg-lay flour (2x sifted flour) for two hours. Following the two hour egg-lay, adults and eggs were separated by sifting with a stacked #25 and #30 brass sieve. Adults were transferred back to normal diet, and the eggs were placed into a Falcon cell strainer (100µm nylon) inside a Petri dish. To clean eggs for injection, they were rinsed in the Falcon cell strainer with 2.5% bleach until they look clean and free of flour then rinsed with water. Eggs were transferred to glass cover slips with a non-toxic paintbrush and injected by standard protocols (Berghammer et al. 2009). Injection mix components and final concentrations were: 500ng/µl of *pB(TcPubCas9-3xP3EGFP)* or *pB(TcαTubCas9-3xP3EGFP)*, 250ng/µl *p(hspBac)*, and 20% phenol red. Glass cover slips with injected eggs were placed in a plastic Petri dish within a Billups-Rothenberg modular incubator chamber with a damp paper towel, stored in a 30ºC incubator. Injected eggs typically take ~4 days to hatch and larvae were transferred to Petri dishes with flour diet. Pupae were then sorted by sex and transferred to plastic vials with ~5g of diet and topper as described in the USDA ARS protocol (Beeman and Friesen 2019). See Brown et al. 2009 for a review on Tribolium biology, husbandry, genetics, genomics, and technical approaches. Plating eggs for injection and separating pupae by sex was done on an Olympus SZX7 stereomicroscope.

Mating scheme for the development of Cas9 lines

For creation of the Cas9 lines (*PubCas9* and *αTubCas9*), injected M26 (G_0) eggs were raised to adulthood and outcrossed to v^w. G_1 offspring were screened for green-eye fluorescence which indicates *Cas9* gene insertion, and red-eye fluorescence which indicates *piggyBac transposase* on the X chromosome. Before making each line homozygous for *Cas9*, we first segregated out the *piggyBac transposase* helper gene by outcrossing Cas9 individuals to v^w for two generations, each time selecting offspring without red eyes (Supplementary Material 2.1). We continued outcrossing until there was only a single *Cas9* insertion before breeding to homozygosity (Supplementary Material 2.2)

Fluorescence screening

Screening for EGFP and DsRed was done on a Leica MZFLIII microscope with a GFP2 (excitation filter: 480/40 nm, barrier filter: 510 nm) or ET DsRed filter (excitation filter: 540/25 nm, barrier filter: 620/60 nm). Photos were taken with a Jenoptik Gryphax camera.

Inverse PCR

Inverse PCR was used to identify location of the transgene within the genome. Genomic DNA was extracted using homemade cell lysis and protein precipitation solutions. 500ng of gDNA was digested with BglII (New England Biolabs) in a 20μl reaction volume. Fragments were then ligated with T4 DNA ligase (Promega) using 5μl of the digested reaction mixture in a 50μl reaction volume and an incubation at 18°C for one hour. One μl of unpurified ligation reaction mixture was used as template for inverse PCR. PCR conditions followed the standard MyTaq protocol, with an annealing temperature of 55°C and 30 second elongation time. Nested PCR was used to increase success. The primary round primers pBi-L1 and pBo-L1 and secondary round primers pBi-L2 and pBo-L2 (Table 2.2) have previously been validated (Chu et al. 2018). Primary PCR product was diluted (1:10) before using it as a template in the nested PCR. PCR product was visualized with GelGreen on a 1% agarose gel; bands above 200bp were extracted using

QIAquick Gel Extraction Kit (Qiagen) and sent to Eton BioScience (Research Triangle Park, NC) for sequencing. Sequencing results were aligned to our known construct sequence.

Table 2.2

Inverse PCR primers

Name	Sequence	Anneal (°C)
pBo-L1	ATCAGTGACACTTACCGCATTGACA	55
pBo-L2	CTCCAAGCGGCGACTGAG	55
pBi-L1	GAATCTTGACCTTGCCACAG	55
pBi-L2	CAAAGTCCACGAGCGGTAGC	55
pBi-L seq	GCACAGCGACGGATTC	-

Note. Annealing temperatures are shown for PCR primers and not the sequencing primer pBi-L seq.

High molecular weight gDNA extraction and PacBio sequencing

PacBio sequencing was used to determine *Cas9* insertion location after inverse PCR was inconclusive. For Cas9 lines #1-9, pupae from 3 lines were pooled together before extracting, for a total of 3 extractions. The three HMW gDNA preps were as follows: A) Cas9 lines 1, 2, 3, B) Cas9 lines 4, 5, 6, and C) Cas9 lines 7, 8, 9. Individual isolations and library preps weren't feasible due to time constraints and funding. For obtaining HMW gDNA we used Omega Bio-tek Mag-Bind Blood and Tissue DNA HDQ 96 Kit (Cat# M6399-01) to extract DNA from 138-140mg of pupal tissue. In order to obtain long fragments (preferably >30,000bp) we optimized the manufacturer protocol by making big changes to incubation and mixing methods. See Supplementary Material 2.3 for our optimized protocol. The three pooled HMW gDNA were sent to the Genomics and

Bioinformatics Research Unit at the USDA, ARS, Mid South Area for library prep and sequencing. Size selection was used to pull out 18-20kb fragments. Each of the three samples were barcoded during library prep so sequencing data from the three samples could be separated after sequencing. All three samples were run on one SMRT cell on the Pac-Bio Sequel II system. PCR based on the PacBio reads and *Tribolium castaneum* genome was used to confirm which strain contained each of the junction sequences.

Fitness assay

F_1 adult progeny production was measured to compare fitness between ten Cas9 lines and three control lines (v^W, A19, and GA2). v^W is the white-eyed line that Cas9 was created in, A19 is the transgenic EGFP line used in the gene knock-out assay, and GA2 is a non-transgenic but highly inbred lab line used in sequencing the Tribolium genome, available on NCBI (Tribolium Genome Sequencing Consortium 2008, Herndon et al. 2020). F_1 adult progeny production is the number of offspring produced that survive to adulthood (Pai et al. 2007, Jbilou et al. 2008). This assay measures more than fecundity because it also includes survival from egg to adult. For each Cas9 line, one virgin female and two virgin males (same line, both 1-2 weeks old) were placed in a vial with 5.000(+/-.001)g diet. There were four replicates per line. Vials were placed in a cardboard tray at a randomized position to account for microenvironmental differences. The tray was stored in the previously described 30ºC incubator. Adults were left for seven days, then removed by paper transfer, leaving the flour undisturbed. Eggs were allowed to develop for four weeks, at which point adults were removed by sieve and counted. Adults continued to be sieved out once every three days until all individuals were collected.

Microinjections for CRISPR/Cas9 knockout of EGFP

First we used Cas9 Lines #1-9 to create EGFP mosaic knockout individuals. Embryos heterozygous for EGFP were created by crossing each Cas9 line #1-9 with A19. A non-Cas9 control was created by crossing v^w and A19. Injection mix final concentrations were: 400ng/µl *p(U6aEGFPG2)* or *p(U6aEGFPG3)* and 20% phenol red (1 to 5 by volume) (See Table 2.1 for

guide sequences). Initial crosses were A19 males mated with *Cas9* or v^w females. For Cas9 lines 4 and 6 we also did the reciprocal cross with A19 females and Cas9 males.

Next we used plasmid or protein sources of Cas9 to create EGFP mosaic knockout individuals. Three injection mixes with Cas9 plasmid had final concentrations: 400ng/µl of one *Cas9* plasmid (either *p(bhsp68-Cas9, pB(TcPubCas9-3xP3EGFP), or pB(TcαTubCas9-3xP3EGFP)*), 400ng/µl *p(U6aEGFPG3)*, and 20% phenol red. Injection mix with Cas9 protein had final concentrations: 650ng/µl Cas9 protein (Fisher TrueCut Cat#A36498), 400 ng/µl sgRNAs (Synthego), and 20% phenol red. RNP complex was created by incubating Cas9 protein injection mix on ice for 30 minutes before injections.

After the *EGFP* gene is cut, it is repaired through error prone DNA repair pathways which can lead to indels. In turn, this can lead to a frame shift mutation and premature stop codons, resulting in a knockout of the green fluorescence phenotype. Injectees have one copy of *Cas9*, one copy of *3xP3-EGFP* with muscle enhancer trap expression, and an additional copy of *3xP3-EGFP* marking the *Cas9* insertion. By using *EGFP* muscle heterozygous individuals, any resulting knockouts will be mosaic individuals where some patches of muscles have no functional copies of *EGFP*. Knockouts of the *EGFP* eye phenotype can be used as an estimate of bi-allelic knockout rates. Though not identical to a homozygote, this should give a conservative estimate because Cas9 is having to target two different genomic locations.

Screening EGFP mosaic knockouts

Pupae were screened for mosaic knockout of EGFP in their muscles and eyes. The number of knockout and non-knockout individuals was recorded. For the EGFP muscle phenotype, the level of mosaicism was also recorded. Level of muscle EGFP knockout was categorized as: Non (0%), Low (1 - 49%), Medium (50 - 95%), and High (96 -100%). High and medium somatic knockout levels are desirable as this may be indicative of gametic knockout levels (heritable gene editing).

Screening EGFP gametic knockouts

F_1 offspring of mosaic knockout parents were screened for EGFP. There were 44 treatment categories: four somatic knockout levels (non, low, medium, high) and eleven Cas9 sources (Cas9 Lines #1-9, *p(bhsp68Cas9)*, and Cas9 protein). For each treatment a single EGFP mosaic adult was given two v^W mates and F_1 pupae screened for EGFP. There were five replicates of each treatment for a total of 220 crosses.

Statistical Analysis: One-way ANOVA

One-way Analysis of Variance (ANOVA) was used to assess the influence of somatic knockout level (non, low, medium, high) on the gametic knockout rate (percentage of offspring with an inactive copy of EGFP). Our null hypothesis assumed that the non, low, medium, and high somatic knockout groups are identical and the alternative hypothesis assumed that at least one of the groups is different from the others. This test can identify a significant difference but not the strength of association. We performed the test in JMP 15 software by SAS Institute Inc. To meet assumptions of heteroscedasticity (equal variances) and normal distribution, the number of outliers and missing rows (some beetle crosses were sterile and therefore had no observations) was reduced by using three of five replicates in the statistical analysis. For each treatment, two replicates with the lowest number of observed offspring were removed from the analysis. One-way ANOVA estimates the variance explained in the continuous variable using the discrete variable. Our continuous dependent variable was percentage of knockout offspring and our categorical independent variable was parental somatic body tissue knockout.

Statistical Analysis: Correlation coefficient

The correlation coefficient (r) was calculated to assess if there was a significant negative correlation between fitness and EGFP knockout rates (proxy for Cas9 cutting efficiency). The correlation coefficient is a unit-free value between -1 (strong negative relationship) and 1 (strong positive relationship). Values at or close to zero imply a weak or no linear relationship. This analsysis was used to compare between fitness and three measures of Cas9 cutting efficiency:

muscle EGFP knockout rates (high to low level), muscle EGFP knockout rates (high and medium level), and eye EGFP knockout rates.

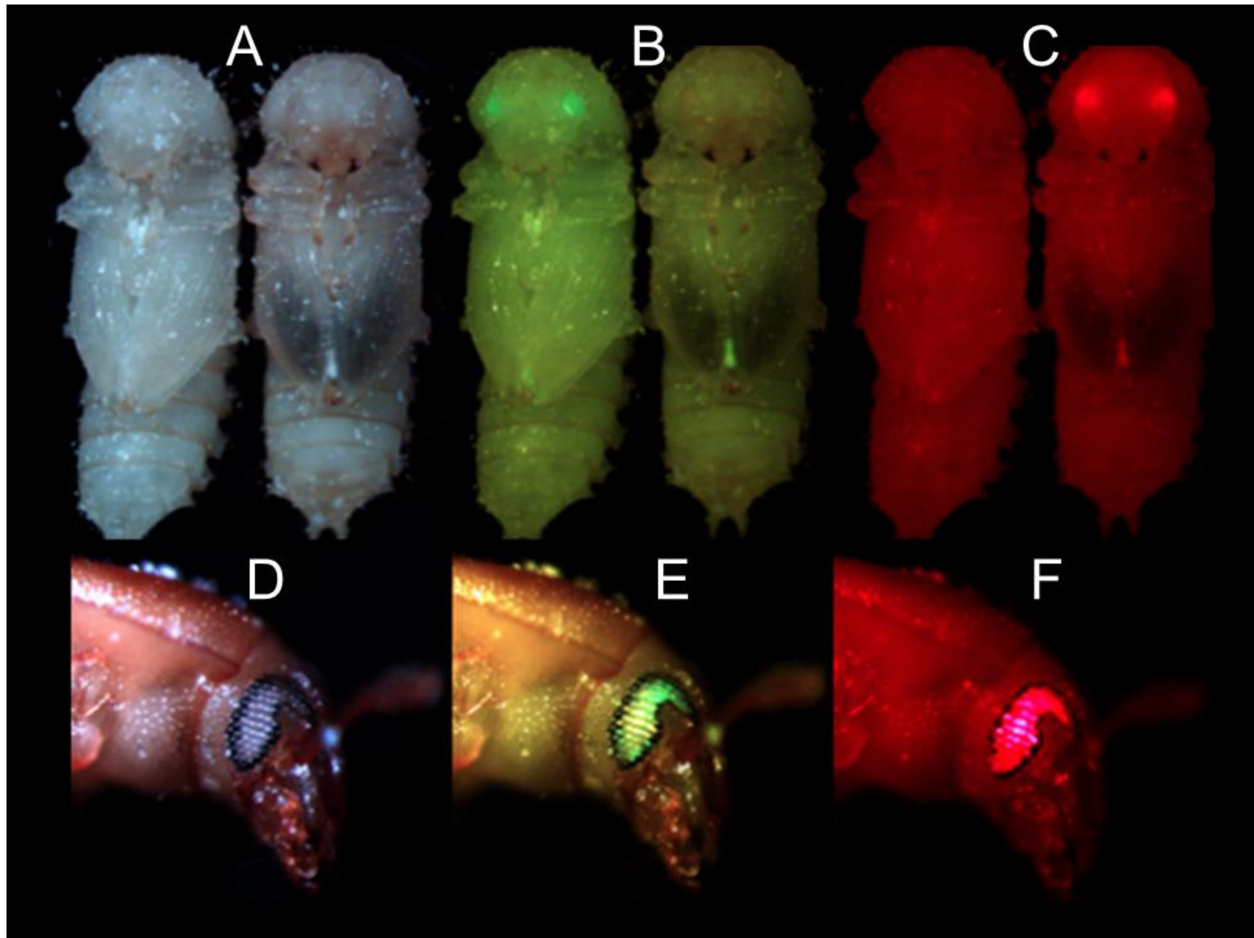
Results

Screening for Cas9 insertion into Tribolium genome

There were a total of 20 G_1 individuals with green eye fluorescence, with 12 possible events from *pB(3xP3-EGFP-TcPub-Cas9)* injections, and 8 possible events from *pB(3xP3-EGFP-TcαTub-Cas9)* (Table 2.3). See Figure 2.1 for EGFP and DsRed eye phenotypes. For *Pub-Cas9* injections, 191 fertile G_0 crosses produced five transformed progeny giving a transformation rate of 2.6%. For *αTub-Cas9* injections, 100 fertile G_0 crosses produced two transformed progeny giving a transformation rate of 2.0% (Table 2.3).

Figure 2.1

G_1 Fluorescent Marker Screening in Pupae and Adults



Note. Top row: Two pupae under A) white light, B) EGFP filter, or C) DsRed filter. Left individual has a *Cas9* insertion (EGFP) and is male from male G_0, so does not have *piggyBac* helper (DsRed) on the X chromosome. Right individual does not have a *Cas9* insertion (EGFP) and is male from female G_0, so has *piggyBac* helper. Bottom row: Single G_1 adult under D) white light, E) EGFP filter, or F) *DsRed* filter. The adult has both *Cas9* (EGFP) and helper (DsRed).

Table 2.3

Microinjection survival and transformation rates

Injection construct	Eggs	Larvae	Hatch Rate	Adults (G_0)	Fertile Crosses	Offspring (G_1)	Successful Crosses	Transformation Rate
pB(3XP3-EGFP-TcPub-Cas9)	750	252	33.6%	229	191	10,833	5	2.6%
pB(3XP3-EGFP-TcaTub-Cas9)	358	163	45.5%	127	100	4,993	2	2.0%

Note. Constructs were injected into *Tribolium* line M26 (v^w background). Some of the successful crosses had multiple transformed offspring. In total, there were 12 transformed G_1 from *Pub-Cas9* and 8 from *aTub-Cas9* (see Table 2.2).

Some G_1 transformants came from the same G_0 parent, so it was possible for the transgene insertion site to be identical between these siblings. We had at least five unique *Pub-Cas9* insertion locations with seven more needing to be confirmed with sequencing, and at least two unique *TcαTub-Cas9* insertion locations with six more needing to be confirmed (Table 2.4). Each G_1 transformant was temporarily assumed to have a unique insertion location and used to establish its own line.

Table 2.4

Names of all G_1 transformants

Cas9 Promoter	Name of Each G_1 Individual				
Pub	1F8-1	1H6-1[1]	2B8-1[D]	3A6-1	3C6-1
		1H6-2[2]	2B8-2[4]		3C6-2
		1H6-3[3]	2B8-3		3C6-3
			2B8-4		
αTub	3H1-1[5]	3H3-1[6]			
	3H1-2	3H3-2[7]			
	3H1-3	3H3-3[8]			
	3H1-4	3H3-4[9]			

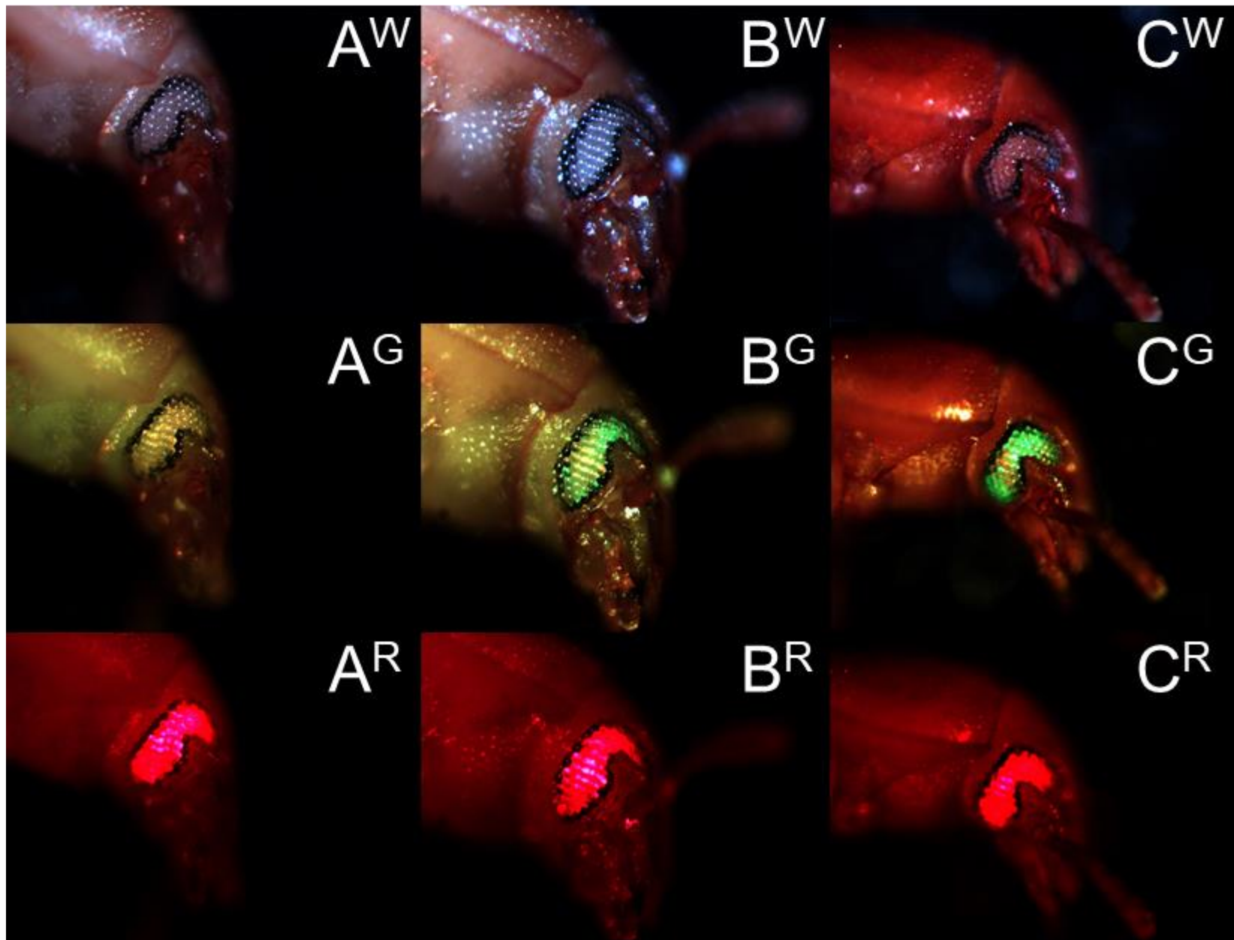
Note. The naming scheme we used for the 20 lines created from the 20 G_1 transformants was the mating box number, row letter, column number, and a ranked number if they came from the same G_0 parent. The number in superscript indicates the ten lines that survived and were tested in other assays. Line D (for dead) was eventually lost.

Breeding to eliminate helper gene and multiple Cas9 insertions

Segregating out the transposase gene and multiple Cas9 insertions took up to five generations for all the lines (Supplementary Material 2.2). During this time, 11 of 20 lines died out and for the remainder of the chapter the lines are named numerically (Cas9 Line #1-9). Die off was presumably due to the disruption of important genes following transgene insertion or *Cas9* expression being too high. Transgenic organisms commonly have lower fitness than non-transgenic lab colonies. During the screening process for all these crosses, we noticed some lines have a higher level of eye fluorescence than others. While screening was done during the pupal stage, these differences in eye fluorescence are especially noticeable in adult beetles. As shown in Figure 2.2, the individual from Cas9 Line 8 has a higher proportion of ommatidia with fluorescence than Cas9 Line 7.

Figure 2.2

Variation in eye fluorescence between G_1 individuals of different lines

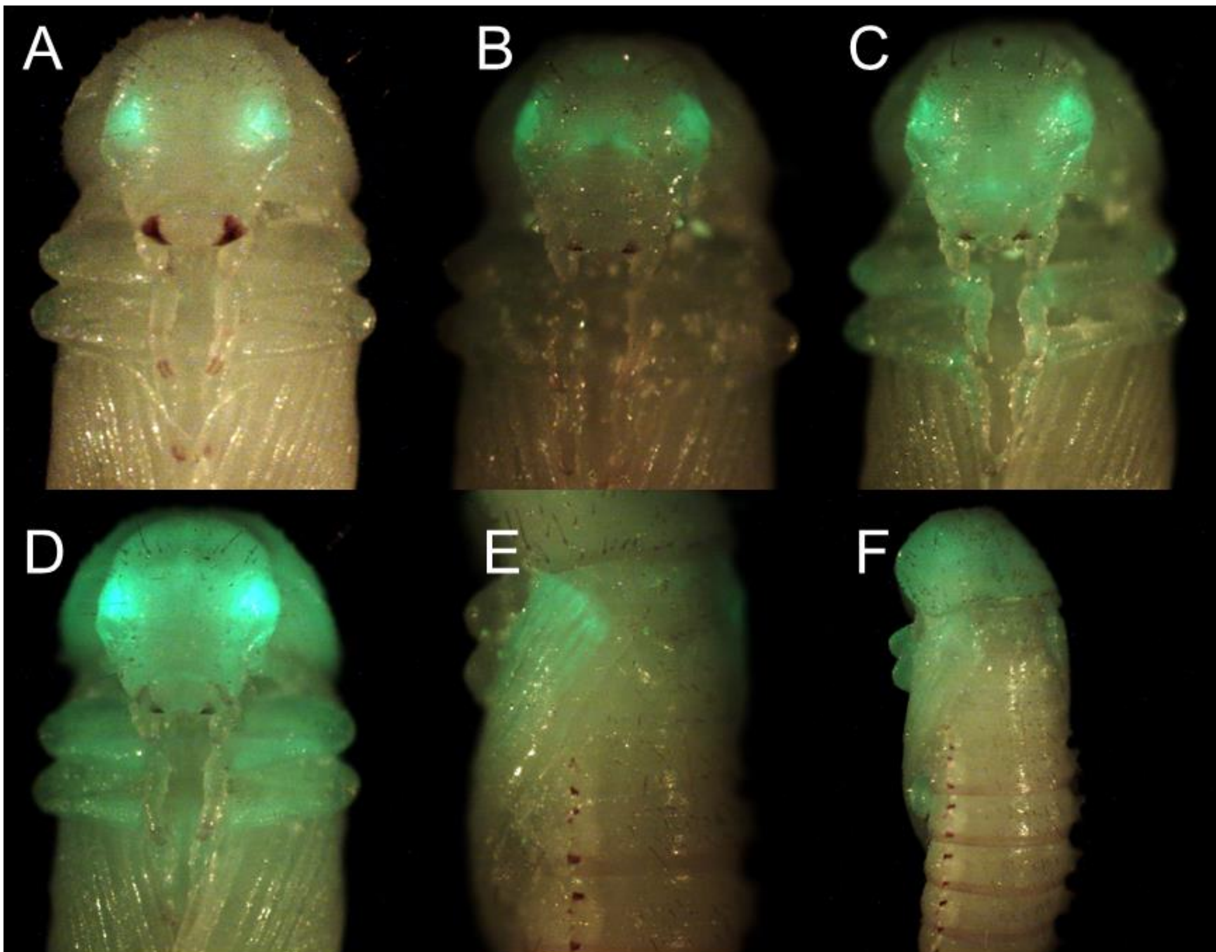


Note. Three individuals (A, B, C) under different filters: W: white light and no filter, G: EGFP2 filter, R: DsRed filter. A: M26 (control), B: G_1 adult from Cas9 Line 7, C: G_1 adult from Cas9 Line 8. Green fluorescence indicates *Cas9* and red fluorescence indicates *piggyBac* helper.

Furthermore, in addition to having green fluorescent eyes, multiple lines had EGFP expression in additional body parts (Figure 2.3). This is likely due to the transgene landing near an enhancer and should not affect the efficiency of Cas9 at the pre-blastoderm embryonic stage.

Figure 2.3

EGFP expression variation between Cas9 lines



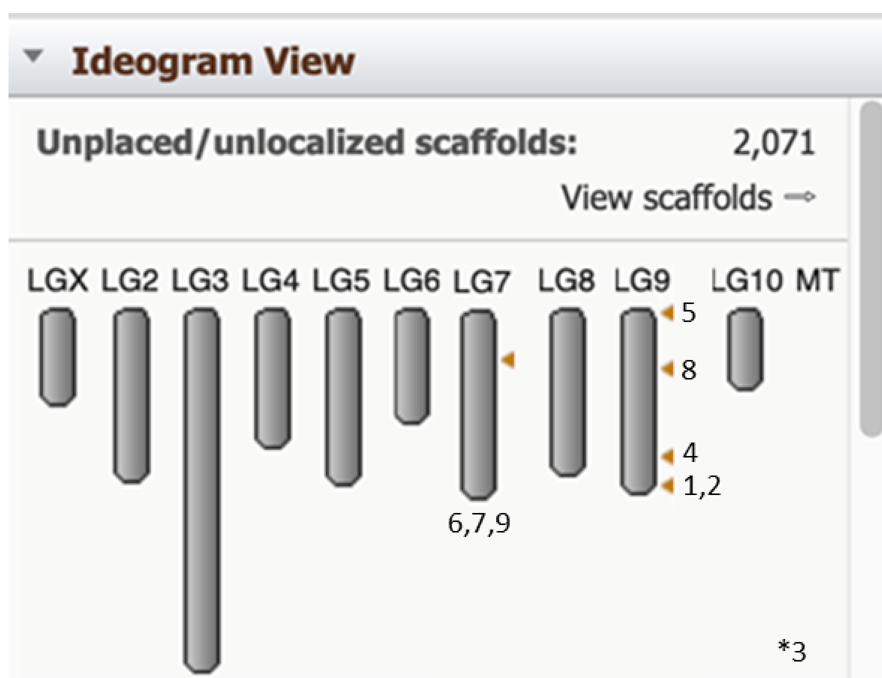
Note. Each photo is an individual from a different line. Variations are possibly due to enhancer traps and have expression in the following body parts: A: No enhancer traps, B: frons, C: clypeus and leg joints, D: head, pronotum, femur, and tibia, E: elytron, F: elytron, head, pronotum, femur, and tibia.

Cas9 insertion locations

Inverse PCR revealed that our *piggyBac* donors *TcPubCas9-3xP3EGFP* and *TcαTubCas9-3xP3EGFP* had formed concatemers before inserting into the Tribolium genome. The 5’ end of the inverse PCR sequence aligned to the right *piggyBac* arm (R1) and ended with a TTAA-target site as expected. However, the downstream sequence was not Tribolium genome but random segments of vector backbone (Supplementary Material 2.4). It seems likely that during injections our construct first inserted into itself before inserting into the *Tribolium* genome (Supplementary Material 2.5). We think it is unlikely to be due to a cloning error, since Sanger sequencing of our constructs came back as expected. Furthermore, the different Cas9 lines had different segments of backbone adjacent to the transgene. An alternative method of identifying insertion site junctions is universal PCR, but this would produce the same data as inverse PCR and therefore be inconclusive. Thus, we used PacBio genome sequencing to map the *Cas9* insertions (Figure 2.4, Supplementary Material 2.6)

Figure 2.4

Cas9 insertion locations



Note. Lines 6, 7, and 9 are identical and located on LG7. Five lines have insertions on LG9 (Lines 1, 2, 4, 5, 8) and Lines 1 and 2 are identical. Line 3 matches a scaffold that has not been mapped and therefore it is unclear what linkage group/chromosome Line 3 is on.

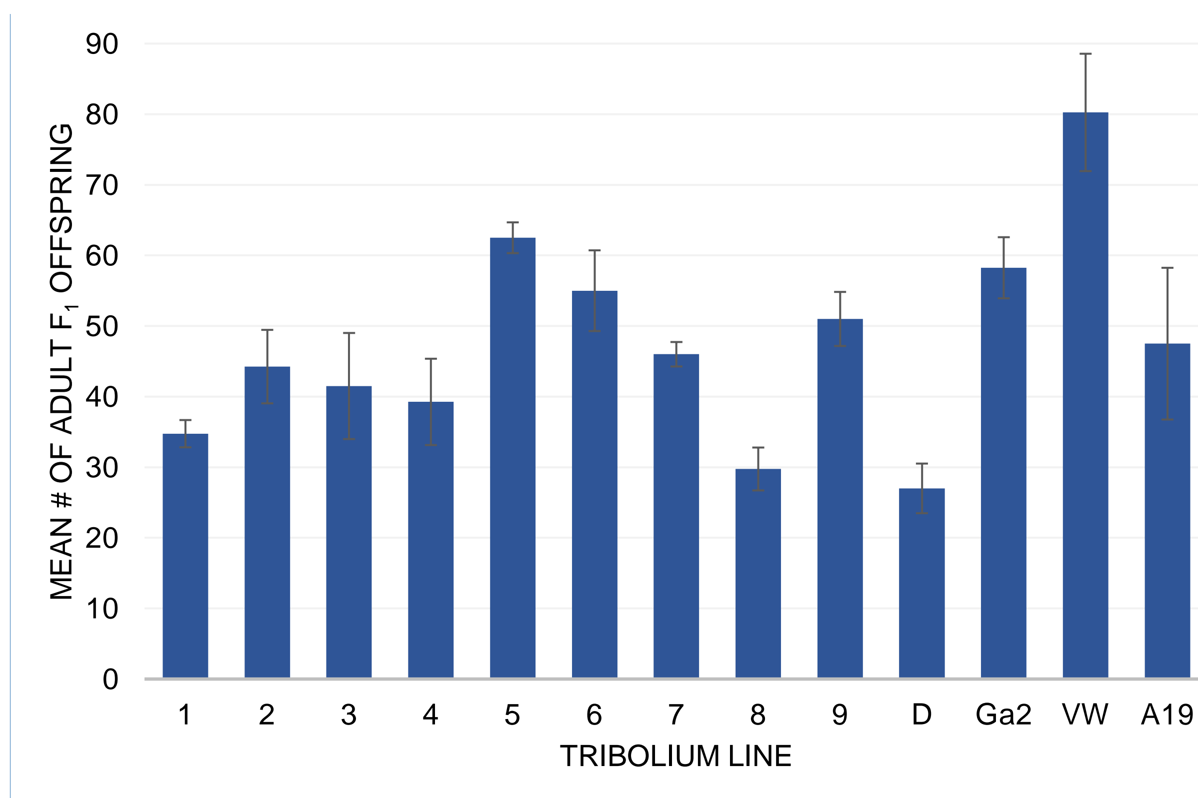
Fitness of transgenic Cas9 Tribolium lines

Our 20 potential lines quickly reduced to nine in a couple of generations. PacBio data was received and analyzed after completing all other analyses. The names Cas9 Lines #1-9 will be used for the remainder of the chapter, and the consistency of data for Lines 1 and 2 (single insertion event) and Lines 6, 7, and 9 (single insertion event) will be addressed in the discussion. Fitness effects were also apparent from observed differences in female egg production. We used F_1 adult progeny production to compare fitness between ten Cas9 lines and three controls. Competitive fitness is more robust but the egg laying capacity of some lines was so low that collecting the number of eggs needed for a competition assay was impossible.

Line 5 had the highest fitness of the *αTubCas9* lines (Figure 2.5). All *PubCas9* lines had similar fitness, with Line 2 having significantly higher fitness than Line 1 and the rest being indistinguishable. The unnumbered Cas9 line 2B8_1 that died out after this fitness assay (indicated in Figure 2.5 as line D), unsurprisingly had the lowest fitness along with Line 8. There was no significant difference between the rest of the Cas9 lines.

Figure 2.5

Fitness of 10 Cas9 lines and 3 control lines



Note. 10 Cas9 lines (#1-9 and D) and 3 control lines (v^W, A19, and Ga2). D (for dead) indicates Cas9 line 2B8_1 which died out immidiatly after this assay. V^W is the white-eye line that the Cas9 lines were created in, A19 is the transgenic EGFP line used in the gene knock-out assay, and Ga2 is a non-transgenic but highly inbred lab line used in sequencing the *Tribolium* genome. Error bars indicate standard error. Lines 1-4 (and D) are *PubCas9* and Lines 5-9 are *αTubCas9*.

Ranking the lines and other Cas9 tools in Tribolium

Once we established nine lines of transgenic *Cas9* Tribolium (which were later identified to be six independent insertion events) we aimed to determine the most efficient line and promoter, then compare to other available CRISPR/Cas9 tools.

1) Compare somatic knockout rates (muscle and eye EGFP)
2) Compare somatic knockout level (muscle mosaicism)
3) Determine if heredity of knockout differs by promoter or somatic knockout level, and
4) Compare the efficiency of current Cas9 tools (protein, plasmids, and transgenic lines)

Somatic knockout rates of muscle and eye EGFP using Cas9 beetle lines

In organisms with transgenic *Cas9*, protein levels may vary between lines due to position effects and the promoter, in this case *polyubiquitin* or *alpha-tubulin*. Therefore, differences in gene editing efficiency are expected. To compare lines 1-9 and between *PubCas9* versus *αTubCas9,* we conducted a knockout experiment using the strain A19 which has *3xP3-EGFP* in a muscle enhancer trap (Figure 2.6). We measured somatic knockout rates of muscle EGFP (one copy) and eye EGFP (two copies). For the muscle EGFP we also recorded the level of knockout ranked from high to low (Figure 2.7). It is important to note that individuals had one copy of *Cas9*. Results from our pilot study using EGFP Guide 2 were inconclusive due to an extremely high efficiency of ~90-100% for all lines. We switched to the less efficient EGFP Guide 3 (previously validated, Grubbs et al. *in preparation*) for the remainder of the study. Injection hatch rate and survival from larva to pupa is reported in Table 2.5. For knockout of muscle EGFP see Table 2.6 and Table 2.7. For knockout of eye EGFP see Figure 2.8 and Table 2.8. Following the completion of all assays, PacBio sequencing revealed there were six independent insertion events. Consistency of data for Lines 1 and 2 (single insertion event) and Lines 6, 7, and 9 (single insertion event) will be addressed in the discussion.

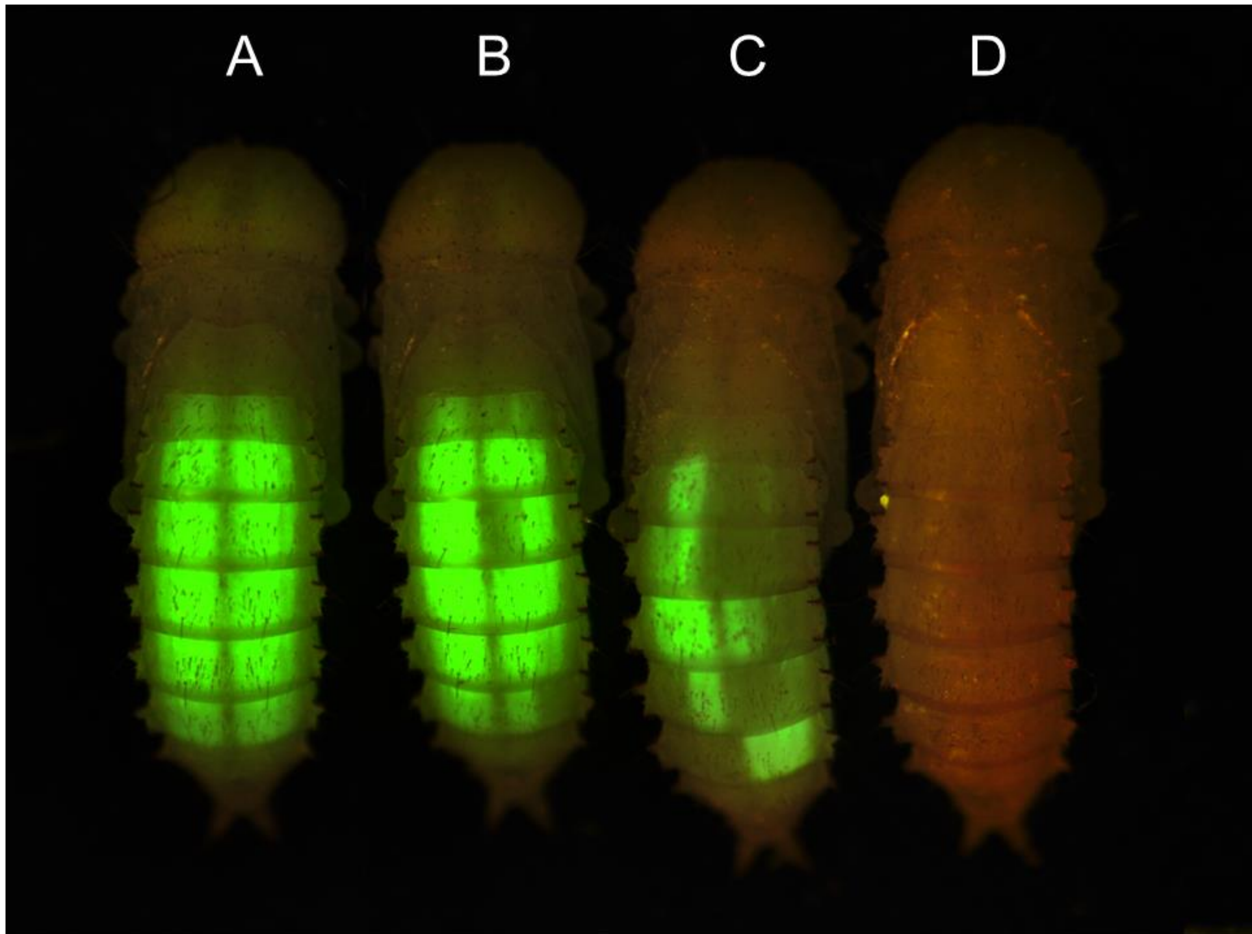
Figure 2.6

Tribolium lines used in EGFP knock-out experiment

Note. Pupae from three distinct *Tribolium* strains under white light (left) and EGFP filter (right). A: *vermilion*^{white} (v^w), white eyes. B: Cas9, EGFP eyes. C: A19, EGFP eyes and muscle. Experimental: Cas9 females mated with A19 males (B and C), Control: v^w females mated with A19 males (A and C). Embryos injected with CRISPR guides targeting EGFP. After EGFP is cut, insertions or deletions (indels) can to a premature stop codon and knockout of the green phenotype. EGFP knockout rates used as a proxy for cutting efficiency.

Figure 2.7

Mosaic muscle EGFP knockout in pupae of EGFP Guide 3 injected individuals



Note. Individuals ranked based on knockout level. A: Non-knockout (N), B: Low-knockout (L), C: Medium-knockout (M), D: High-knockout (H).

Table 2.5

Survival rate of injected G_0 eggs to larva (hatch rate) and from larva to pupa

Promoter	Cas9 Line	Injected Eggs	Hatched	Survival	Pupa	Survival
Pub	1	165	113	68%	94	83%
Pub	2	146	110	75%	102	93%
Pub	3	186	104	56%	55	53%
Pub	4	215	163	76%	151	93%
αTub	5	176	116	66%	70	60%
αTub	6	193	105	54%	73	70%
αTub	7	190	126	66%	104	83%
αTub	8	156	125	80%	114	91%
αTub	9	177	136	77%	128	94%

Note. Rates are conditionally formatted to show highest rates in yellow and lowest rates in blue.

Table 2.6

Somatic muscle EGFP knockout screening

Promoter	Cas9 Line	Knockout					
		High	Med	Low	Non	Total	High+Med
Pub	1	6	18	28	42	52	24
Pub	2	3	26	45	28	74	29
Pub	3	2	9	17	27	28	11
Pub	4	32	71	28	20	131	103
αTub	5	0	6	13	51	19	6
αTub	6	14	15	19	25	48	29
αTub	7	8	13	27	56	48	21
αTub	8	0	18	44	52	62	18
αTub	9	7	13	43	65	63	20

Note. Somatic muscle EGFP knockout rates at a given level of body knockout (high, med, low, non), total number of knockouts (K/O), and the sum of high and medium knockouts (High+Med). The number of knockout individuals out of the total number of pupae at each level (high-low), for each Cas9 line (1-9). We defined knockout rates as the proportion of injectees from each treatment with inactivated *EGFP*, observable by patches of eye or muscle without EGFP.

Table 2.7

Somatic muscle EGFP knockout rates

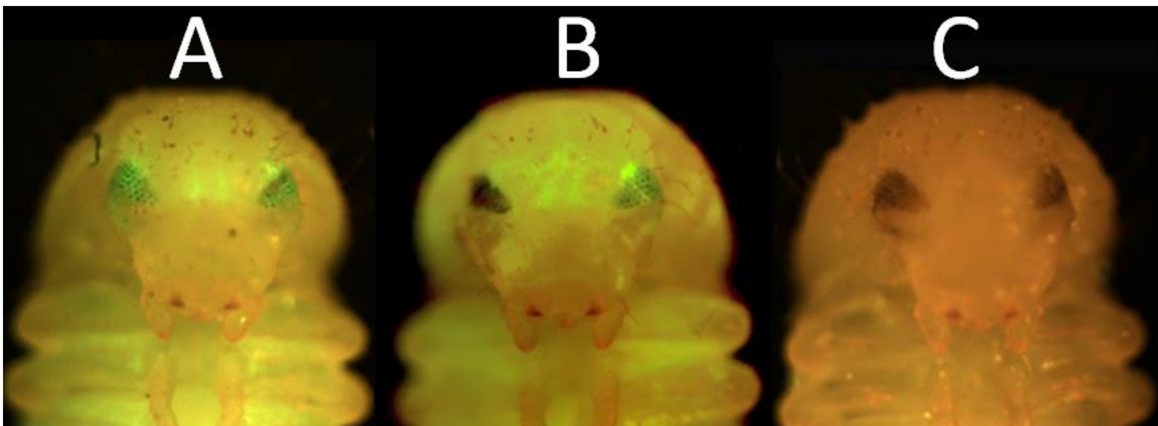
Promoter	Cas9 Line	High	Med	Low	Non	K/O	High+Med
Pub	1	6%	19%	30%	45%	55%	26%
Pub	2	3%	25%	44%	27%	73%	28%
Pub	3	4%	16%	31%	49%	51%	20%
Pub	4	21%	47%	19%	13%	87%	68%
αTub	5	0%	9%	19%	73%	27%	9%
αTub	6	19%	21%	26%	34%	66%	40%
αTub	7	8%	13%	26%	54%	46%	20%
αTub	8	0%	16%	39%	46%	54%	16%
αTub	9	5%	10%	34%	51%	49%	16%

Note. Somatic muscle EGFP knockout rates at a given level of body knockout (high, med, low, non), total number of knockouts (K/O), and the sum of high and medium knockouts (High+Med). Rates were calculated as the number of knockout individuals out of the total number of pupae at each level (high-low), for each Cas9 line (1-9). Knockout rates are conditionally formatted to show highest rates in yellow and lowest rates in blue. See Table 2.6 for raw numbers.

Eye EGFP knockout rates were used as an estimate of bi-allelic knockout rates. The eye EGFP knockout assay was more stringent than the muscle knockout assay and therefore gave a conservative estimate of bi-allelic knockout rates. Despite embryos being injected at the posterior end, this is still a strong assay because unlike Drosophila which undergoes long germband embryogenesis, Tribolium undergoes an intermediate germband mode of embryogenesis. The blastodermal fate map of Tribolium shows most of the anterior end is devoted to extraembryonic cell fates, while segments of the head are located more centrally. In contrast, cells fated to become head segments in Drosophila are located in the anterior end of the embryo and extraembryonic cell fates are restricted to the dorsal side of the embryo. As expected, we found eye knockout rates (2 copies *EGFP*) to be lower than muscle knockout rates (1 copy *EGFP*). Although there is the caveat that mosaic eye knockouts are harder to observe under a microscope than muscle knockouts. Unlike muscle knockouts which are large and patchy, eye knockouts are often small and inconspicuous. The most obvious eye knockouts had an entire eye missing a functional copy of EGFP, but this was rare (Figure 2.8).

Figure 2.8

Mosaic eye EGFP knockout in pupae of EGFP Guide 3 injected individual



Note. A: Non-knockout, B: Knockout, one eye, C: Knockout, both eyes.

Table 2.8

Somatic eye EGFP knockout screening

Promoter	Cas9 Line	Non K/O	K/O	K/O%
Pub	1	57	37	39%
Pub	2	49	53	52%
Pub	3	30	25	45%
Pub	4	35	116	77%
αTub	5	62	8	11%
αTub	6	47	26	36%
αTub	7	78	26	25%
αTub	8	94	20	18%
αTub	9	89	39	30%

Note. We defined knockout rates as the proportion of injectees from each treatment with inactivated *EGFP*, observable by patches of eye or muscle without EGFP. Rates were calculated as the number of knockout individuals out of the total number of pupae for each Cas9 line (1-9). Knockout rates are conditionally formatted to show highest rates in yellow and lowest rates in blue.

Comparing PubCas9 versus αTubCas9 lines

PubCas9 lines 2 and 4 had higher muscle EGFP knockout rates than *αTubCas9* lines (Figure 2.9, A). *PubCas9* line 4 and *αTubCas9* line 6 had the highest knockout rates at the medium or full body level of mosaicism (Figure 2.9, B). All *PubCas9* lines had higher eye EGFP knockout rates than *αTubCas9* lines (Figure 2.9, C). All *PubCas9* lines had higher eye EGFP knockout rates than any of the *αTubCas9* lines.

Figure 2.9

Cas9 lines sorted by (A) muscle EGFP knockout rate, (B) High+Medium level muscle EGFP knockout rate, (C) eye EGFP knockout rates, and (D) fitness

A)

Promoter	Cas9 Line	Knockout
Pub	4	87%
Pub	2	73%
αTub	6	66%
Pub	1	55%
αTub	8	54%
Pub	3	51%
αTub	9	49%
αTub	7	46%
αTub	5	27%

B)

Promoter	Cas9 Line	High+Med
Pub	4	68%
αTub	6	40%
Pub	2	28%
Pub	1	26%
αTub	7	20%
Pub	3	20%
αTub	8	16%
αTub	9	16%
αTub	5	9%

C)

Promoter	Cas9 Line	Knockout
Pub	4	77%
Pub	2	52%
Pub	3	45%
Pub	1	39%
αTub	6	36%
αTub	9	30%
αTub	7	25%
αTub	8	18%
αTub	5	11%

D)

Promoter	Cas9 Line	Fitness
αTub	5	63
αTub	6	55
αTub	9	51
αTub	7	46
Pub	2	44
Pub	3	42
Pub	4	39
Pub	1	35
αTub	8	30

Note. Lines sorted from high (yellow) to low (blue) knockout rate. *Pub Cas9* lines are labeled gray while *αTubCas9* lines are labeled white.

There appeared to be a weak negative correlation between fitness and knockout rates, so the correlation coefficient (r) was calculated between fitness (Figure 2.9 D) and three measures of Cas9 cutting efficiency (Figure 2.9 A, B, C). There was not a significant negative correlation between fitness and any of the three measures: muscle EGFP knockout rates ($r = -0.4431$, $P = 0.2323$), muscle EGFP knockout rates (high and medium level) ($r = -0.2225$, $P = 0.5650$), and eye EGFP knockout rates ($r = -0.3336$, $P = 0.3804$). However, correlations best describe linear relationships and the 95% density ellipse of each scatterplot was more circular than ovular (Supplementary Material 2.7), which indicates the relationship may be more complex and not well captured by correlation.

Gametic knockout rates (Heredity of EGFP knock-out)

After assessing the level of somatic knockout, we were curious as to the level of gametic knockout. We had two aims: assess any correlation between a) level of parental knockout (non-high) and percent knockout offspring, and b) Cas9 line and percent knockout offspring. In other words, we want to know if there is a correlation between a categorical independent variable (low-high parental K/O) and a continuous dependent variable (% offspring non-GFP). Parental somatic body knockout level ($F = 53.64$, d.f. = 3, $P < 0.001$) and parental somatic eye knockout level ($F = 17.85$, d.f. = 3, $P < 0.001$) were found to be significant predictors of offspring knockout rate, whereas Cas9 line was not ($P = 0.55$) (Table 2.9). This was likely due to the structure of the assay (non, low, medium, and full knockouts were considered for each line) and the wide range of efficiency between lines (some of the best αTubCas9 lines outperformed some of the worst PubCas9 lines). Additionally, we expected high levels of somatic knockout to be indicative of high levels of gametic knockout, since it was more likely that future germ cells were targeted. Indeed, injectees with a high level of muscle EGFP knockout produced the most offspring with the knockout (Table 2.9). See Supplementary Material 2.8 for a raw data table with replicates and Supplementary Material 2.9 for a one-way ANOVA graph with percent of knockout offspring plotted against parental somatic body knockout level.

Table 2.9

Gametic muscle and eye EGFP knockout screening

Line	Promoter	Parental Knockout	Offspring				
			Pupae	Non-EGFP Muscles		Non-EGFP Eyes	
6	αTub	High	505	505	100.0%	307	60.8%
9	αTub	High	414	414	100.0%	186	44.9%
1	Pub	High	377	374	99.2%	284	75.9%
7	αTub	High	241	227	94.2%	188	82.8%
3	Pub	High	384	360	93.8%	268	74.4%
2	Pub	High	179	159	88.8%	157	98.7%
8	αTub	Med	588	489	83.2%	378	77.3%
4	Pub	High	267	218	81.6%	207	95.0%
3	Pub	Med	415	325	78.3%	275	84.6%
7	αTub	Med	440	335	76.1%	272	81.2%
2	Pub	Low	497	358	72.0%	279	77.9%
6	αTub	Med	497	355	71.4%	255	71.8%
2	Pub	Med	545	370	67.9%	266	71.9%
7	αTub	Low	556	359	64.6%	237	66.0%
4	Pub	Med	518	332	64.1%	194	58.4%
9	αTub	Med	400	252	63.0%	170	67.5%
4	Pub	Low	189	115	60.8%	83	72.2%
1	Pub	Med	288	175	60.8%	112	64.0%

Note. Table sorted by percent offspring with non-EGFP muscles. Data shown is the sum of five replicates for each treatment. See Supplementary Material 2.8 for replicates.

Table 2.9 *(Continued)*

Line	Promoter	Parental Knockout	Offspring				
			Pupae	Non-EGFP Muscles		Non-EGFP Eyes	
3	Pub	Low	529	316	59.7%	210	66.5%
9	αTub	Low	470	275	58.5%	159	57.8%
3	Pub	Non	500	283	56.6%	220	77.7%
2	Pub	Non	392	219	55.9%	160	73.1%
6	αTub	Low	360	200	55.6%	172	86.0%
8	αTub	Low	499	276	55.3%	168	60.9%
4	Pub	Non	426	233	54.7%	149	63.9%
6	αTub	Non	433	234	54.0%	148	63.2%
7	αTub	Non	350	188	53.7%	127	67.6%
5	αTub	Non	313	162	51.8%	115	71.0%
1	Pub	Low	250	129	51.6%	104	80.6%
8	αTub	High	2	1	50.0%	1	100.0%
5	αTub	Low	125	62	49.6%	48	77.4%
8	αTub	Non	578	283	49.0%	163	57.6%
1	Pub	Non	242	112	46.3%	108	96.4%
9	αTub	Non	584	268	45.9%	145	54.1%
5	αTub	High	0	0	n/a	0	n/a
5	αTub	Med	0	0	n/a	0	n/a

Note. Table sorted by percent offspring with non-EGFP muscles. Data shown is the sum of five replicates for each treatment. See Supplementary Material 2.8 for replicates.

Maternal preloading of Cas9 into embryos

In *Anopheles stephensi*, *Anopheles gambiae*, and *Drosophila melanogaster*, transgenic *Cas9* mothers have been found to load *Cas9* into eggs when using the promoter *nanos*, *vasa,* or *Ubiquitin 63E* (Gantz et al. 2015, Hammond et al. 2017, Kondo and Ueda 2013, Sebo et at. 2014, Champer et al. 2017, Kandul et al. 2019). To explore possible maternal preloading effects when using *Pub* and *αTub*, we repeated the EGFP knockout experiments with the reciprocal cross (Cas9 males X A19 females) of two best lines: *PubCas9* Line 4 and *αTubCas9* Line 6.

When Cas9 and EGFP heterozygous eggs were injected with EGFP Guide 3, knockout rates were much higher in offspring of Cas9 females than Cas9 males (Table 2.10). This is evidence that *PubCas9* and *αTubCas9* are both maternally preloaded and that preloading is crucial for high Cas9 cutting efficiency. Knockout rates were 11.5% versus 86.8% in *PubCas9* and 1.6% versus 65.8% in *αTubCas9*. Knockout rates at the high and medium levels of mosaicism were also a lot lower, with male *PubCas9* and *αTubCas9* both producing <1% knockouts at this level. No eye knockouts were observed from *αTubCas9* fathers and <3.4% from *PubCas9* fathers.

Table 2.10

Somatic muscle EGFP knockout rates from Cas9 fathers versus Cas9 mothers

Promoter	Cas9 Line	Sex	High	Med	Low	Non	Total K/O	High+Med
Pub	4	♂	0 (0.0%)	1 (0.7%)	16 (10.8%)	131 (88.5%)	17 (11.5%)	1 (0.7%)
Pub	4	♀	32 (21.2%)	71 (47.0%)	28 (18.5%)	20 (13.2%)	131 (86.8%)	103 (68.2%)
αTub	6	♂	0 (0.0%)	1 (0.8%)	1 (0.8%)	124 (98.4%)	2 (1.6%)	1 (0.8%)
αTub	6	♀	14 (19.2%)	15 (20.5%)	19 (26.0%)	25 (34.2%)	48 (65.8%)	29 (39.7%)

Note. Somatic muscle EGFP knockout rates at a given level of body knockout (high, med, low, non), total number of knockouts (K/O), and the sum of high and medium knockouts (High+Med). Rates were calculated as the number of knockout individuals out of the total number of pupae at each level (high-low), for each Cas9 line (1-9).

Comparing Cas9 tools: Cas9 plasmids, protein, and transgenic lines

Methods for gene editing insects with CRISPR/Cas9 have changed over the last six years. Of highest interest to this chapter is the method of Cas9 delivery into embryos. Tools continue to be developed and optimized, including *Cas9* plasmids with native promoters for different species and transgenic *Cas9* lines. Another recent change is the use of RNP complexes (protein complexed in vitro with guide RNAs). The somatic EGFP knockout experiment was repeated to compare efficiency between multiple Cas9 delivery methods (plasmid, protein, transgene). Gametic knockout efficiency (heritable gene editing) was also compared.

RNP complex ranked highest both in terms of somatic knockout rate (Table 2.11) and somatic knockout rate at a medium to high level of mosaicism (Table 2.12). Transgenic lines *PubCas9* Line 4 and *αTubCas9* Line 6 were second best, followed by plasmid-based Cas9 expression. Of three plasmids with different promoters, *Pub* performed the best, followed by *bhsp68,* then *αTub*. See Tables 2.13 and 2.14 for raw data.

Table 2.11

Tribolium Cas9 tools ranked by somatic knockout rate

Cas9 Source	Knockout
Protein	96%
PubCas9 Line 4	87%
pB(TcPub-Cas9)	73%
αTubCas9 Line 6	66%
p(bhsp68-Cas9)	57%
pB(TcαTub-Cas9)	50%

Table 2.12

Tribolium Cas9 tools ranked by somatic knockout rate at the high and medium level

Cas9 Source	High+Med
Protein	96%
PubCas9 Line 4	68%
αTubCas9 Line 6	40%
pB(TcPub-Cas9)	23%
p(bhsp68-Cas9)	10%
pB(TcαTub-Cas9)	10%

Table 2.13

Survival rate of injected G_0 eggs to larva (hatch rate) and from larva to pupa

Cas9 Source	Cas9 (ng/µl)	Guide Source	Guide (ng/µl)	Injected Eggs	Hatched	Survival	Pupa	Survival
PubCas9 Line 4	-	*p(U6a-EGFP-G3)*	400	215	163	76%	151	93%
Protein	625	EGFP G3 sgRNA	400	342	235	69%	206	88%
p(bhsp68-Cas9)	400	*p(U6a-EGFP-G3)*	400	351	260	68%	247	95%
Buffer control	-	-	-	278	186	67%	144	77%
Protein	625	EGFP G3 sgRNA	1000	288	171	59%	144	84%
αTubCas9 Line 6	-	*p(U6a-EGFP-G3)*	400	193	105	54%	73	70%
pB(TcPub-Cas9)	400	*p(U6a-EGFP-G3)*	400	330	170	52%	78	46%
pB(TcαTub-Cas9)	400	*p(U6a-EGFP-G3)*	400	332	172	37%	105	61%

Note. Sorted by hatch rate.

Table 2.14

Somatic EGFP muscle knockout rates from Cas9 plasmids, purified protein, and transgenic lines

Cas9 Source	Cas9 (ng/µl)	Guide (ng/µl)	High	Med	Low	Non	Total K/O	High+Med
Protein	625	1000	141 (97.9%)	1 (0.7%)	0 (0.0%)	2 (1.4%)	142 (98.6%)	142 (98.6%)
Protein	625	400	190 (92.2%)	7 (3.4%)	1 (0.5%)	8 (3.9%)	198 (96.1%)	197 (95.6%)
PubCas9 Line 4	-	400	32 (21.2%)	71 (47.0%)	28 (18.5%)	20 (13.2%)	131 (86.8%)	103 (68.2%)
αTubCas9 Line 6	-	400	14 (19.2%)	15 (20.5%)	19 (26.0%)	25 (34.2%)	48 (65.8%)	29 (39.7%)
pB(TcPub-Cas9)	400	400	3 (3.8%)	15 (19.2%)	39 (50.0%)	21 (26.9%)	57 (73.1%)	18 (23.1%)
p(bhsp68-Cas9)	400	400	5 (2.0%)	19 (7.7%)	116 (47.0%)	107 (43.3%)	140 (56.7%)	24 (9.7%)
pB(TcαTub-Cas9)	400	400	2 (1.9%)	8 (7.6%)	43 (41.0%)	52 (49.5%)	53 (50.5%)	10 (9.5%)
Buffer control	-	-	0 (0.0%)	0 (0.0%)	0 (0.0%)	144 (100.0%)	0 (0.0%)	0 (0.0%)

Note. Sorted by High+Medium knockout. Somatic EGFP muscle knockout rates at a given level of body knockout (high, med, low, non), total number of knockouts (K/O), and the sum of high and medium knockouts (High+Med). Rates were calculated as the number of knockout individuals out of the total number of pupae at each level (high-low).

When Cas9 delivery methods were ranked by gametic knockout, a similar order was observed, with RNP complex followed by transgenic *Cas9*, then plasmid (Table 2.15). For this assay only plasmid *p(bhsp68Cas9)* was assessed. For simplicity, only the top two performing Cas9 lines were considered. It was not possible to use a one-way ANOVA since there were not enough mosaic individuals at each level (non, low, medium, high) for each Cas9 delivery method (protein, plasmid, transgene) to have an adequate number of replicates. RNP complex did not have enough non, low, and medium level knockouts while *p(bhsp68Cas9)* did not have enough high level knockouts. Each replicate was a single injectee outcrossed to two v^{W} mates. The best comparison given the available replicates (Supplementary Material 2.10) is at the high level of parental (injectee) somatic muscle knockout, summarized in Table 2.15. This table shows the sum of all available replicates. Given the results of previous assays, the results of this assay fell within our expectations except for *αTubCas9* Line 6 outperforming *PubCas9* Line 4. However, two replicates of *PubCas9* Line 4 with high parental knockout only produced two offspring each and may have skewed the data. Further testing may decrease the difference between the lines, although it is compelling that there was not a single non-knockout offspring from five crosses with an αTubCas9 Line 6 parent.

Table 2.15

Gametic EGFP muscle knockout rates from p(bhsp68Cas9), purified protein, and the top two performing transgenic lines

Source	Parental	Offspring		
	Knockout	Pupae	Muscle EGFP	Non-EGFP
Protein	High	241	0	100.0%
αTubCas9 Line 6	High	505	0	100.0%
PubCas9 Line 4	High	267	47	82.4%
p(bhsp68Cas9)	High	149	33	77.9%

Note. Sorted by percentage of non-EGFP offspring.

Discussion

This chapter described the development of a new CRISPR/Cas9 gene editing tool in Tribolium with potential applications for functional genomics and gene drive for pest management. For functional genomics, transgenic *Cas9* is more efficient than plasmids and may offer some convenience over Cas9 protein. Tribolium are small and low maintenance while purified Cas9 protein, as a consumable, needs to be reordered regularly, is moderately expensive, and can loose enzymatic activity over time. Some labs may elect to keep both Cas9 protein and Cas9 transgenic lines on hand to have more options and greater flexibility. We will be maintaining *PubCas9* Line 4 (efficiency of 68.2% of injectees having a med-high level of somatic knockout, of which produced 70.1% knockout offspring) and *αTubCas9* Line 6 (efficiency of 39.7% med-high level somatic knockout, of which produced 85.8% knockout offspring). In comparison, protein had an efficiency of 95.6% of injectees with a medium to high level of somatic knockout, of which produced 100% knockout offspring.

All knockout assays were performed in double heterozygotes, so efficiency may be higher when using a *Cas9* homozygote or lower for gene knockout in a target gene homozygote. Efficiency of the *Cas9* strains for gene replacement or gene knock-in have not been tested and strategies that rely on HDR are generally less efficient. Additionally, the gRNA used will affect overall efficiency and this chapter highlights the importance and wide range of guide efficiency *in vivo*, given that using an inefficient gRNA was crucial for the analysis. In the pilot study we used a guide we knew to have high efficiency (~78% when EGFP homozygotes were microinjected with Cas9 and guide plasmids), which resulted in all nine Cas9 lines to be indistinguishable, each with a knockout rate of 98% or greater. To adequately compare the lines, we switched to a guide previously tested to have a lower efficiency (~34%). The efficiency of gRNA is affected by gRNA sequence features (composition, nucleotide position, GC content), energetics properties (RNA secondary structure, melting temperature, free energy), and genetic and epigenetic features (Liu

et al. 2020). See Liu et al. (2020) for a summary, evaluation, and comparison of computational tools for designing highly efficient gRNAs.

Some examples of potential functional genomics studies enabled by these lines are the investigation of female embryonic lethal genes for genetic sexing strains, developmental genes that could be targeted with RNAi or gene drive for pest management, pesticide receptors and resistance genes, and genes involved in metabolic or immunity pathways to increase our general understanding of pest biology. Two limitations of using transgenic Cas9 strains versus Cas9 protein are the slightly lower fitness of the Cas9 strains compared to non-transgenic strains (although they had comparable fitness to transgenic strain A19) and the potentially undesired presence of the *Cas9* transgene which would need to be eliminated by outcrossing. Tribolium has a generation time of around 27 days at 30°C, which may seem long to Drosophila researchers though is still much faster than some beetle pests which can have generation times of a year or more. Regardless, the process of outcrossing to eliminate the *Cas9* transgene before homozygosing the line (if that is part of the goal) can take multiple months. It is for this reason that I did not use one of my Cas9 strains in Chapter 3 where I use functional genomics to investigate *TcCad*, a potential Cry3Aa resistance gene in Tribolium. If not for the time constraints of the dissertation process, I would have used the strains developed in Chapter 2 towards meeting the goals in Chapter 3 and therefore been able to report on their efficiency in gene editing with homology-directed repair.

Functional genomics can be achieved through gene editing and mutagenesis. The CRISPR/Cas9 system has been used for targeted mutagenesis in several insect orders and species such as *Drosophila melanogaster* (Bassett et al. 2013), *Spodoptera litura* (Bi et al. 2016), *Acyrthosiphon pisum* (Le Trionnaire et al. 2019) and *Leptinotarsa decemlineata* (Gui et al. 2020) to name a few. Just like gene editing, mutagenesis does not require transgenic *Cas9*, but it simplifies the microinjection process by reducing the number of components, provides convenience, and reliably produces heritable mutations. Both *PubCas9* Line 4 and αTubCas9

Line 6 reliably produced heritable gene knockouts, which supports their use in mutagenesis to transmit mutations through the germline to make stable mutant lines.

Testing different promoters for driving expression of Cas9 is also crucial for CRISPR gene drive systems. While constitutive promoters are not desirable for most homing drive systems, they are suitable for the split-drive system Home-and-Rescue (HomeR), and would also likely work in toxin-antidote drive systems such as Cleave and Rescue (ClvR) and Toxin-Antidote Recessive Embryo (TARE). In homing drive systems, having a tightly-regulated germline-specific promoter is essential as leaky expression and maternal deposition of Cas9 can lead to NHEJ-induced repair events and cleavage-resistant alleles (Hammond et al. 2017, Champer et al. 2017). In contrast, TARE, ClvR, and HomeR systems are more flexible and should function with a variety of promoters.

TARE (Champer et al. 2020) and ClvR (Oberhofer et al. 2019) systems function by two components, one that disrupts an essential gene and a second that provides a recoded version of the essential gene resistant to cleavage. The difference is that in ClvR, the drive allele and target gene reside at two different genomic loci, while TARE can be either same-site or distant-site. Both systems are far less prone to the formation of resistance alleles than homing-type drives, though both have only been tested with the *nanos* promoter thus far (Oberhofer et al. 2019, Champer et al. 2020). HomeR is a homing split-drive system that has stringent design criteria to avoid the accumulation of resistance alleles, and has successfully used two constitutive promoters, *Ubiquitin 63E* and *Actin 5C*, for driving Cas9 expression (Kandul et al. 2021). One HomeR design criteria is the use of lethal biallelic mosaicism, where maternal carryover of Cas9/gRNA complexes results in biallelic disruption of an essential target gene throughout development, and thereby ensures resistant alleles result in dominant lethal mutations.

Constitutive promoters are also suitable for precision guided sterile insect technique (pgSIT), which entails the release of eggs that exclusively give rise to sterile males (Kandul et al. 2019). Unlike gene drives which rely on HDR, this CRISPR-based pest management strategy

relies on NHEJ-based repair to simultaneously disrupt genes essential for female viability and male fertility. Like HomeR, PgSIT has tested *Ubiquitin 63E* for driving expression of Cas9 and benefits from lethal biallelic mosaicism (Kandul et al. 2019).

Although there was not a statistically significant negative correlation between Cas9 activity and fitness of the Cas9 lines, there was an observable trend. Differences in fitness were observed even in the absence of gRNA, leading us to believe fitness effects were caused by high expression levels in some of the lines. Many scholars have reported on the toxicity of Cas9, though usually in the context of the Cas9/gRNA complex which can have off-target cutting that leads to cytotoxicity. Toxicity has been observed in studies that used microinjected Cas9 (Sebo et al. 2014), transgenic *Cas9* lines (Port et al. 2014), and CRISPR gene drive insects (Champer et al. 2020, Kandul et al. 2021). However, there have also been transgenic *Cas9* mosquito (Li et al. 2017) and fly lines (Ren et al. 2013) reported to not have significant toxicity or fitness effects.

In addition to off-target cutting, some of the fitness cost is likely attributed to expression of the large Cas9 protein itself (Champer et al. 2017). Expression of high amounts of Cas9 protein is toxic in various organisms, including bacteria, algae, mammals, and insects. In Drosophila, crossing UAS-Cas9 transgenic lines to strong ubiquitous Gal4 drivers resulted in substantial lethality, even with a catalytically dead mutant protein (Port et al. 2014). Comparatively, enhancer-driven Cas9 was found to be less cytotoxic (Poe et al. 2019). Expression of a *Cas9* transgene was also predicted to be a major contributor to fitness costs in TARE and HomeR gene drive systems (Champer et al. 2020, Kandul et al. 2021). Therefore, the potential for *polyubiquitin* and *α-tubulin* promoters for driving Cas9 expression in a gene drive system requires empirical testing. Meanwhile, their use in gene editing is well balanced between Cas9 activity and toxicity, as the lines had statistically equivalent fitness to transgenic strain A19 and non-transgenic strain Ga2.

As mentioned previously, Cas9 lines #1-9 were tested before knowing their insertion sites. Some G_1 individuals used to establish the nine lines came from the same G_0 parent, so it was possible for the *Cas9* transgene insertion site to be identical between these siblings. Following

the completion of all assays, PacBio sequencing revealed there were six independent insertion events, with a single insertion event shared between Lines 1 and 2, and another between Lines 6, 7, and 9. Lines 1 and 2 were established from two G_1 individuals from the same G_0 parent along with Line 3, but Line 3 was determined to have a different insertion site. Similarly, Lines 6, 7, and 9 were established from G_1 individuals from the same G_0 parent along with Line 8, but Line 8 had a different insertion site. There are multiple possibilities for the reason why Lines 3 and 8 are different from Lines 1 and 2, and Lines, 6, 7, and 9, respectively. The initial insertion event could have been different, the initial insertion could have been the same and subsequently remobilized due to the availability of *piggyBac* transposase from the helper strain M26, or there could have been multiple insertion events which were then separated by outcrossing.

After determining which lines had the same insertion site, we assessed the similarity of the data between the lines for each assay. Overall, the results of the fitness and knockout assays align with Lines 1 and 2 being the same, as well as Lines, 6, 7, and 9. However, there were some differences which are possibly due to background genetics and variations in microinjection. Depending on the sex of the G_0 and G_1 used to establish the line, I had to outcross to v^W a different number of times to eliminate the *piggyBac transposase* gene on the X chromosome. Additionally, some of the first inbreeding crosses revealed there were multiple insertions in some of the lines, so I completed additional outcrosses. Lines 1, 2, and 7 were outcrossed for four generations before breeding to homozygosity, and Lines 6 and 9 were outcrossed for five generations.

The fitness of Lines 6, 7, and 9 were not statistically different, but Lines 1 and 2 were (Table 2.16). Fitness was measured by production of F_1 adult offspring. Competitive fitness is more robust but the egg laying capacity of some on the nine lines was so low that collecting the number of eggs needed for a competition assay was impossible. However, Lines 1, 2, 6, 7, and 9 all had egg production high enough to conduct a competitive fitness assay, which could offer more insight to the similarities or differences between these lines. It is also possible that Lines 1

and 2 have differences in fitness due to their genetic backgrounds, which may be different enough from the many generations of outcrosses to cause differences in organismal phenotypes.

Table 2.16

Comparison between lines with the same Cas9 gene insertion location

Cas9 Line #:	1	2	6	7	9
Production of F_1 adult offspring (Fitness)	34.75 ±1.93	44.25 ±5.20	55.00 ±5.72	46.00 ±1.73	51.00 ±3.83
Somatic muscle EGFP knockout (Low-High level mosaicism)	55%	73%	66%	46%	49%
Somatic muscle EGFP knockout (Med-High level mosaicism)	26%	28%	40%	20%	16%
Somatic eye EGFP knockout	39%	52%	36%	25%	30%
Gametic muscle EGFP knockout (% K/O offspring from a high somatic K/O parent)	99%	89%	100%	94%	100%

One-way ANOVA showed there was no significant difference in gametic knockout levels between the lines. Thus, the lines can be considered to have equivalent gametic knockout efficiency. No statistical analysis was run on the somatic knockout data sets because the criteria were to have at least 100 hatchlings for each line, but did not include replicates. For each measure of somatic EGFP knockout shown in Table 2.16, there was a range of 18%, 2%, and 13% between Lines 1 and 2, and a range of 20%, 24%, and 5% between Lines 6, 7, and 9. When all nine Cas9 lines were ranked by efficiency (Figure 2.9), Lines 1 and 2, and 6, 7, and 9 tended to group together, though not always. Like fitness, differences could be due to differences in genetic background, though they are more likely attributed to variation in microinjections. Although eggs

were microinjected following the same protocol, there can be unavoidable variations due to slight differences in the needle tip and egg pressure differences from room humidity and pressure.

Conclusions

The red flour beetle is a global pest of stored grain and a model organism for development and pest biology. Optimization of CRISPR/Cas9 tools in *Tribolium* will facilitate studies in this model organism and expedite development of similar tools in other non-model beetle species. To increase Cas9 gene editing efficiency we created multiple transgenic Cas9 lines. We found the ubiquitous promoters *polyubiquitin* (*TcPub*) and *alpha-tubulin* (*TcαTub*) sufficient for driving expression of *Cas9.* Transgenic *TcPubCas9* lines ranked higher in cutting efficiency but lower in fitness than *TcαTubCas9* lines. Therefore, usefulness of these Cas9 lines is balanced between fitness and cutting efficiency.

Assessing the level of eye-specific EGFP knockout revealed that biallelic knockout was higher when *Cas9* gene expression was driven by the *Pub* than *αTub* promoter. We think this is likely due to a combination of promoter timing and maternal preloading. We saw a 64-76% higher knock out rate in injected eggs from *Cas9* females versus *Cas9* males and we know from previous work (Lorenzen et al. unpublished) that naturally occurring *Pub* transcripts are maternally preloaded into embryos at higher levels than *αTub*. Promoter timing also seems to play a role as evidenced by injections with *PubCas9* plasmids which outperformed *αTubCas9* plasmids. The *TcPub* promoter may be earlier acting than *TcαTub* and the higher rate of bi-allelic knockout with *TcPubCas9* may be attributed to Cas9 being present in higher quantities. In summary, differences are likely due to greater maternal preloading and earlier zygotic transcription.

Of the Cas9 delivery methods available for Tribolium, we found RNP complexed protein to be the most efficient, followed by *Cas9* transgene then plasmid. Cas9 protein may be prohibitively expensive to some labs, in which case transgenic Cas9 Tribolium may be preferable. Cas9 protein and transgene had higher rates of knockout at levels >50% somatic tissue (96% for protein and 40-68% for transgene) than plasmids (10-23%). For *Cas9* transgene, maternal

preloading was critical to achieve high levels of mosaicism. This is important because we found level of somatic gene editing to be a predictor of heritable gene editing.

REFERENCES

Bassett, A. R., Tibbit, C., Ponting, C. P., & Liu, J. L. (2013). Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. *Cell reports*, *4*(1), 220-228.

Beeman, R. W., and K. Friesen. “Beetle Wrangling Tips: An Introduction to the Care and Handling of Tribolium Castaneum.” *Stored Product Insect and Engineering Research: Manhattan, KS*, USDA ARS, 31 Dec. 2019, www.ars.usda.gov/plains-area/mhk/cgahr/spieru/docs/tribolium-stock-maintenance/.

Berghammer, A. J., Weber, M., Trauner, J., & Klingler, M. (2009). Red flour beetle (Tribolium) germline transformation and insertional mutagenesis. *Cold Spring Harbor Protocols*, *2009*(8), pdb-prot5259.

Bi, H. L., Xu, J., Tan, A. J., & Huang, Y. P. (2016). CRISPR/Cas9-mediated targeted gene mutagenesis in Spodoptera litura. *Insect science*, *23*(3), 469-477.

Brown, S., Shippy, T., Miller, S., Bolognesi, R., Beeman, R., & Lorenzen, M. et al. (2009). The Red Flour Beetle, Tribolium castaneum (Coleoptera): A Model for Studies of Development and Pest Biology. *Cold Spring Harbor Protocols*, *2009*(8), pdb.emo126-pdb.emo126. doi: 10.1101/pdb.emo126

Cary, L. C., Goebel, M., Corsaro, B. G., Wang, H. G., Rosen, E., & Fraser, M. J. (1989). Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*, *172*(1), 156-169.

Champer, J., Reeves, R., Oh, S. Y., Liu, C., Liu, J., Clark, A. G., & Messer, P. W. (2017). Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS genetics*, *13*(7), e1006796.

Champer, J., Lee, E., Yang, E., Liu, C., Clark, A. G., & Messer, P. W. (2020). A toxin-antidote CRISPR gene drive system for regional population modification. *Nature communications*, *11*(1), 1-10.

Chu, F. C., Klobasa, W., Grubbs, N., & Lorenzen, M. D. (2018). Development and use of a piggyBac-based jumpstarter system in Drosophila suzukii. *Archives of insect biochemistry and physiology*, *97*(3), e21439.

Cong, L., Ran, F., Cox, D., Lin, S., Barretto, R., & Habib, N. et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819-823. doi: 10.1126/science.1231143

Gantz, V. M., Jasinskiene, N., Tatarenkova, O., Fazekas, A., Macias, V. M., Bier, E., & James, A. A. (2015). Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. *Proceedings of the National Academy of Sciences*, *112*(49), E6736-E6743.

Gilles, A. F., Schinko, J. B., & Averof, M. (2015). Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle Tribolium castaneum. *Development*, *142*(16), 2832-2839.

Gui, S., Taning, C. N. T., Wei, D., & Smagghe, G. (2020). First report on CRISPR/Cas9-targeted mutagenesis in the Colorado potato beetle, Leptinotarsa decemlineata. *Journal of insect physiology*, *121*, 104013.

Hammond, A., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., Kranjc, N., Carpi, F., D'Aurizio, R., Crisanti, A., & Nolan, T. (2017). The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLOS Genetics*, 13(10), p.e1007039.

Handler, A. M., & Harrell R. A. (1999). Germline transformation of Drosophila melanogaster with the piggyBac transposon vector. *Insect molecular biology*, *8*(4), 449-457.

Herndon, N., Shelton, J., Gerischer, L., Ioannidis, P., Ninova, M., Dönitz, J., Waterhouse, R.M., Liang, C., Damm, C., Siemanowski, J. & Kitzmann, P. (2020). Enhanced genome assembly and a new official gene set for Tribolium castaneum. *BMC genomics*, *21*(1), 1-13.

Horn, C., Jaunich, B., & Wimmer, E. A. (2000). Highly sensitive, fluorescent transformation marker for Drosophila transgenesis. *Development genes and evolution*, *210*(12), 623-629.

Jbilou, R., Amri, H., Bouayad, N., Ghailani, N., Ennabili, A., & Sayah, F. (2008). Insecticidal effects of extracts of seven plant species on larval development, α-amylase activity and offspring production of Tribolium castaneum (Herbst)(Insecta: Coleoptera: Tenebrionidae). *Bioresource technology*, *99*(5), 959-964.

Kandul, N. P., Liu, J., Wu, S. L., Marshall, J. M., & Akbari, O. S. (2019). Transforming insect population control with precision guided sterile males with demonstration in flies. *Nature communications*, *10*(1), 1-12.

Kondo, S., & Ueda, R. (2013). Highly improved gene targeting by germline-specific Cas9 expression in Drosophila. *Genetics*, *195*(3), 715-721.

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3 Bioinformatics 23(10):1289-91

Le Trionnaire, G., Tanguy, S., Hudaverdian, S., Gleonnec, F., Richard, G., Cayrol, B., Monsion, B., Pichon, E., Deshoux, M., Webster, C., Uzest, M., Herpin, A., & Tagu, D. (2019). An integrated protocol for targeted mutagenesis with CRISPR-Cas9 system in the pea aphid. *Insect Biochemistry and Molecular Biology*, 110, pp.34-44.

Li, M., Bui, M., Yang, T., Bowman, C. S., White, B. J., & Akbari, O. S. (2017). Germline Cas9 expression yields highly efficient genome engineering in a major worldwide disease vector, Aedes aegypti. *Proceedings of the National Academy of Sciences*, *114*(49), E10540-E10549.

Liu, G., Zhang, Y., & Zhang, T. (2020). Computational approaches for effective CRISPR guide RNA design and evaluation. *Computational and structural biotechnology journal*, *18*, 35-44.

Lorenzen, M. D., S. J. Brown, R. E. Denell, and R. W. Beeman. "Transgene expression from the Tribolium castaneum Polyubiquitin promoter." *Insect molecular biology*11, no. 5 (2002): 399-407.

Lorenzen, M. D., Berghammer, A. J., Brown, S. J., Denell, R. E., Klingler, M., & Beeman, R. W. (2003). piggyBac-mediated germline transformation in the beetle Tribolium castaneum. *Insect molecular biology*, *12*(5), 433-440.

Lorenzen, M. D., Kimzey, T., Shippy, T. D., Brown, S. J., Denell, R. E., & Beeman, R. W. (2007). piggyBac-based insertional mutagenesis in Tribolium castaneum using donor/helper hybrids. *Insect molecular biology*, *16*(3), 265-275.

Oberhofer, G., Ivy, T., & Hay, B. A. (2019). Cleave and Rescue, a novel selfish genetic element and general strategy for gene drive. *Proceedings of the National Academy of Sciences*, *116*(13), 6250-6259.

Pai, A., Feil, S., & Yan, G. (2007). Variation in polyandry and its fitness consequences among populations of the red flour beetle, Tribolium castaneum. *Evolutionary Ecology*, *21*(5), 687-702.

Park, T. (1937). The inheritance of the mutation" pearl" in the flour beetle, Tribolium castaneum Herbst. *The American Naturalist*, *71*(733), 143-157.

Poe, A., Wang, B., Sapar, M., Ji, H., Li, K., Onabajo, T., Fazliyeva, R., Gibbs, M., Qiu, Y., Hu, Y. and Han, C. (2019). Robust CRISPR/Cas9-Mediated Tissue-Specific Mutagenesis Reveals Gene Redundancy and Perdurance in Drosophila. *Genetics*, 211(2), pp.459-472.

Port, F., Chen, H. M., Lee, T., & Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proceedings of the National Academy of Sciences*, *111*(29), E2967-E2976.

Ren, X., Sun, J., Housden, B. E., Hu, Y., Roesel, C., Lin, S., ... & Ni, J. Q. (2013). Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. *Proceedings of the National Academy of Sciences*, *110*(47), 19012-19017.

Sebo, Z. L., Lee, H. B., Peng, Y., & Guo, Y. (2014). A simplified and efficient germline-specific CRISPR/Cas9 system for Drosophila genomic engineering. *Fly*, *8*(1), 52-57.

Siebert, Kendra S., Marcé D. Lorenzen, Susan J. Brown, Yoonseong Park, and Richard W. Beeman. "Tubulin superfamily genes in Tribolium castaneum and the use of a Tubulin promoter to drive transgene expression." *Insect biochemistry and molecular biology* 38, no. 8 (2008): 749-755.

Tamura, T., Thibert, C., Royer, C., Kanda, T., Eappen, A., Kamba, M., Kômoto, N., Thomas, J.L., Mauchamp, B., Chavancy, G. and Shirk, P.,Fraser, M., Prudhomme, C., & Couble, P. (2000). Germline transformation of the silkworm Bombyx mori L. using a piggyBac transposon-derived vector. *Nature biotechnology*, *18*(1), 81-84.

Taning, C. N. T., Van Eynde, B., Yu, N., Ma, S., & Smagghe, G. (2017). CRISPR/Cas9 in insects: Applications, best practices and biosafety concerns. *Journal of insect physiology*, *98*, 245-257.

Tribolium Genome Sequencing Consortium. (2008). The genome of the model beetle and pest Tribolium castaneum. *Nature*, *452*(7190), 949.

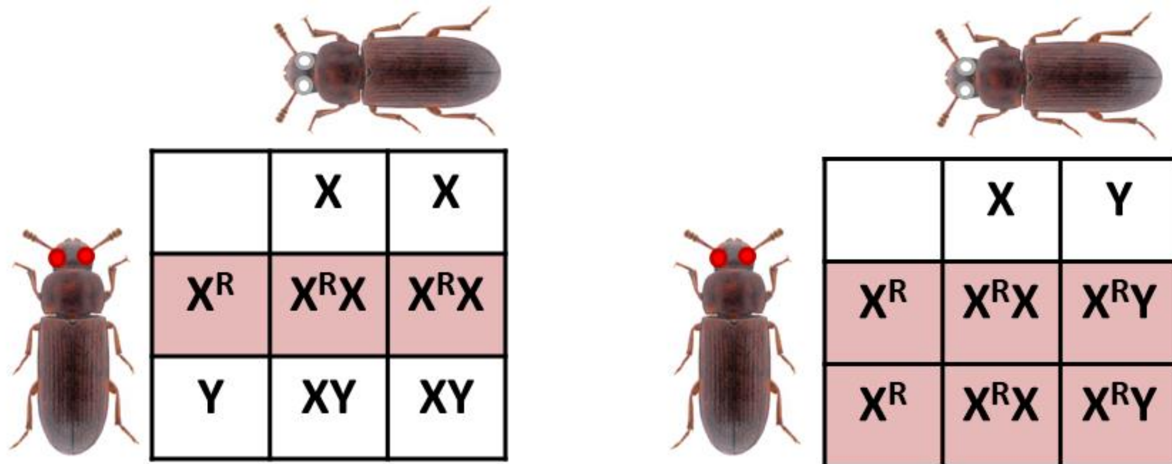
Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3 - new capabilities and interfaces. Nucleic Acids Research 40(15):e115

Woo, J. W., Kim, J., Kwon, S. I., Corvalán, C., Cho, S. W., Kim, H., ... & Kim, J. S. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature biotechnology*, *33*(11), 1162-1164.

SUPPLEMENTARY MATERIAL

Supplementary Material 2.1

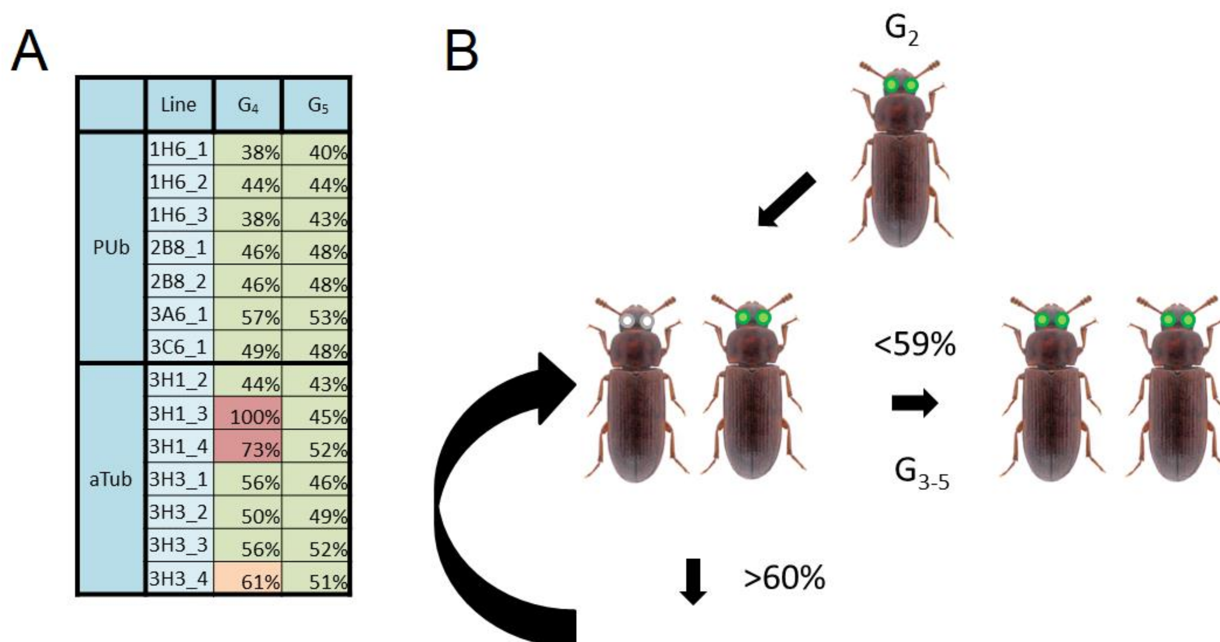
Mating scheme to eliminate the X-linked piggyBac helper from Cas9 lines



Note. Punnett squares have X-linked *piggyBac helper* indicated by X^R and red shading. Injected G_0 adults (red eye) were outcrossed to v^W (white eye). G_0 males had one copy helper and females had two copies helper. Outcrosses were continued until the helper gene was purified from each Cas9 line.

Supplementary Material 2.2

Mating scheme to segregate multiple insertions of Cas9 transgene



	Line	G_4	G_5
PUb	1H6_1	38%	40%
	1H6_2	44%	44%
	1H6_3	38%	43%
	2B8_1	46%	48%
	2B8_2	46%	48%
	3A6_1	57%	53%
	3C6_1	49%	48%
aTub	3H1_2	44%	43%
	3H1_3	100%	45%
	3H1_4	73%	52%
	3H3_1	56%	46%
	3H3_2	50%	49%
	3H3_3	56%	52%
	3H3_4	61%	51%

Note. A: G_2 individuals (heterozygous for *Cas9* and *EGFP*) were outcrossed to v^W and their offspring screened for EGFP. Using the Hardy–Weinberg principle we would expect 50% of offspring to be EGFP (used for approximation since the assumption of infinite population size is not met). Lines that had 59% or fewer EGFP offspring were then self-crossed beginning at G_3. Lines that had 60% or more EGFP offspring were assumed to have multiple insertions and outcrossed until multiple insertions could not be detected. All lines had a single insertion by G_5, with three lines needing to be outcrossed for 4 generations before starting self-crosses as seen in B. B: Screening of EGFP eyes in G_4 or G_5, reported as percentage of offspring. By G_4 only two lines (marked with red) likely had multiple insertions and one line (orange) was also suspected. By G_5 all lines had begun self-crosses.

Supplementary Material 2.3

Optimized protocol for High Molecular Weight (HMW) gDNA extractions for PacBio sequencing

Purpose
We would like to find a happy balance between mixing our samples roughly enough to get high yields from our genomic extractions, and treating them gently enough to meet high molecular weight qualifications for PacBio sequencing. We are aiming to have at least 10μg of 50,000bp for each sample prep, and a DIN of at least 9. DNA length, quality, and concentration should be assessed.

Materials:
Omega Bio-tek Mag-Bind Blood and Tissue DNA HDQ 96 Kit (Cat# M6399-01)
Omega Bio-tek Magnetic Separation Device (MSD-02)
BioTix 1.5mL microcentrifuge tube (Cat# MT-0150-R)
Olympus wide-bore tips (Cat#22424)
Blue polypropylene pellet pestle (In drawer under shaking incubator)
Invitrogen 0.1M DTT (Bottom drawer in -20)
Thermo Fisher Scientific RNase A 10mg/mL (Cat# EN0531) (Bottom drawer in -20)

Equipment:
Vortex Mixer (Fisher Scientific)
Test Tube Rocker (Unico)
C24KC Incubator Shaker (New Brunswick Scientific)
Multi-Blok Heater (American Scientific)
5415D Centrifuge (Eppendorf)
NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific)
2200 TapeStation (Agilent)

Methods:
1. Add samples to 1.5mL tube:

Tribolium castaneum	~130-180mg	~40-60 pupae

   *If possible, use early stage pupae. Supposed to be better for HMW DNA since they haven't formed complex body tissues like wings, and exoskeleton has minimal tanning.

2. Add 400μL 0.1M DTT to 1000μL TL Buffer for final concentration 28.6mM DTT

   *Typo in manufacturer's protocol: Step 2 "40μM" should be "40mM"

   *Because manufacturer suggests using 1M DTT and we only have 0.1M, I decreased the volume (and final concentration) of DTT since I was concerned about the final concentration of the TL buffer getting too low. Concentrations aside, this amount of liquid (350μL) per sample seemed appropriate for the amount of tissue. The manufacturer protocol calls for up to 10mg of tissue and 250μL TL (or 260μL if add 1.0M DTT).
3. Squish pupae with blue pestle with vertical motion ten times. Use just enough pressure to squish and stop once you think most of the internal tissues are released. Don't worry about shearing the exoskeleton.

4. Add 20μL Proteinase K and mix by vortexing at speed 3 for 5 sec.

5. Incubate at 55°C for 3 hours: In shaking incubator at 100rpm with tube laying horizontally
6. Place in tube rack at room temp and add 5μL RNase A. Vortex at speed 3 for 5 sec then incubate 5 min at room temp.
7. Centrifuge at max speed (16,000xg) for 5 min.
8. Transfer 300μL supernatant with wide-bore tips to new 1.5mL tube.

   *Because starting volume was 350 instead of 250, increased supernatant from 200μL to 300μL
9. Add 345μL AL Buffer. Vortex at speed 3 for 5 sec, put on rocker for 5 min.

   *Because have 300μL supernatant, increased AL from 230μL to 345μL to maintain 1.15:1 ratio of Buffer AL:Supernatant
10. Add 480μL HDQ and 20μL MagBind. Vortex at speed 3 for 5 sec, put on rocker for 5 min.

    *Vortex beads before pipetting from bottle

    *Increased HDQ from 320μL to 480μL to maintain 1.6:1 ratio of HDQ:Supernatant.
11. Place on magnetic rack for 2 min
12. Pipette off the cleared supernatant
13. Add 600μL VHB. Vortex at speed 3 for 5 sec, put on rocker for 5 min.

    *Beads don't resuspend at vortex speed 3.

    *Beads don't completely resuspend on rocker, mostly just swish around.
14. Place on magnetic rack for 2 min
15. Pipette off the cleared supernatant
16. Add 600μL VHB (Repeated). Vortex at speed 3 for 5 sec, put on rocker for 5 min.
17. Place on magnetic rack for 2 min
18. Pipette off the cleared supernatant
19. Add 600μL SPM. Vortex at speed 8 for 5 sec, put on rocker for 1 min.
20. Place on magnetic rack for 2 min
21. Pipette off the cleared supernatant
22. Leave on magnetic rack additional 1 min
23. Pipette off waste and dry beads with tube lid open for 10 min

24. Add 50μL 70°C Elution Buffer for 1° Elution. Vortex at speed 8 for 5 sec, put on rocker for 5 min.

    *Expect 1° Elution to have lower concentration than 2° Elution.

25. Place on magnetic rack for 2 min. Transfer 1° Elution with wide bore pipette tip to new tube.
26. Add 50μL 70°C Elution Buffer for 2° Elution. Vortex at speed 8 for 5 sec, rocker for 5 min, vortex at speed 8 for 5 sec.

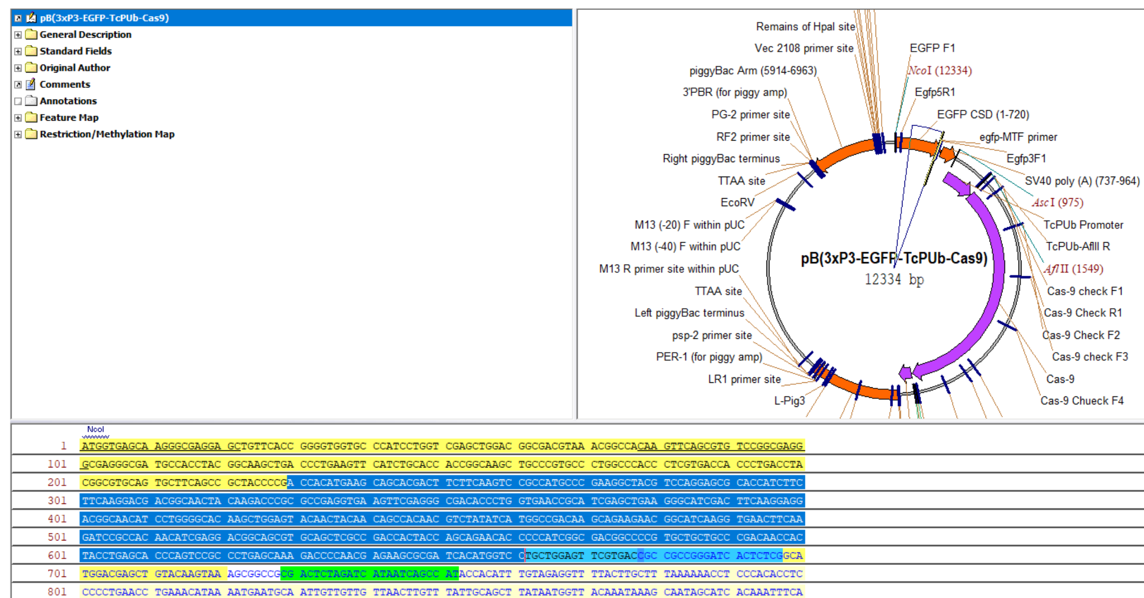
    *Expect 2° Elution to have higher concentration than 1° Elution.

    *3° Elution is not worth it because has very low yield.

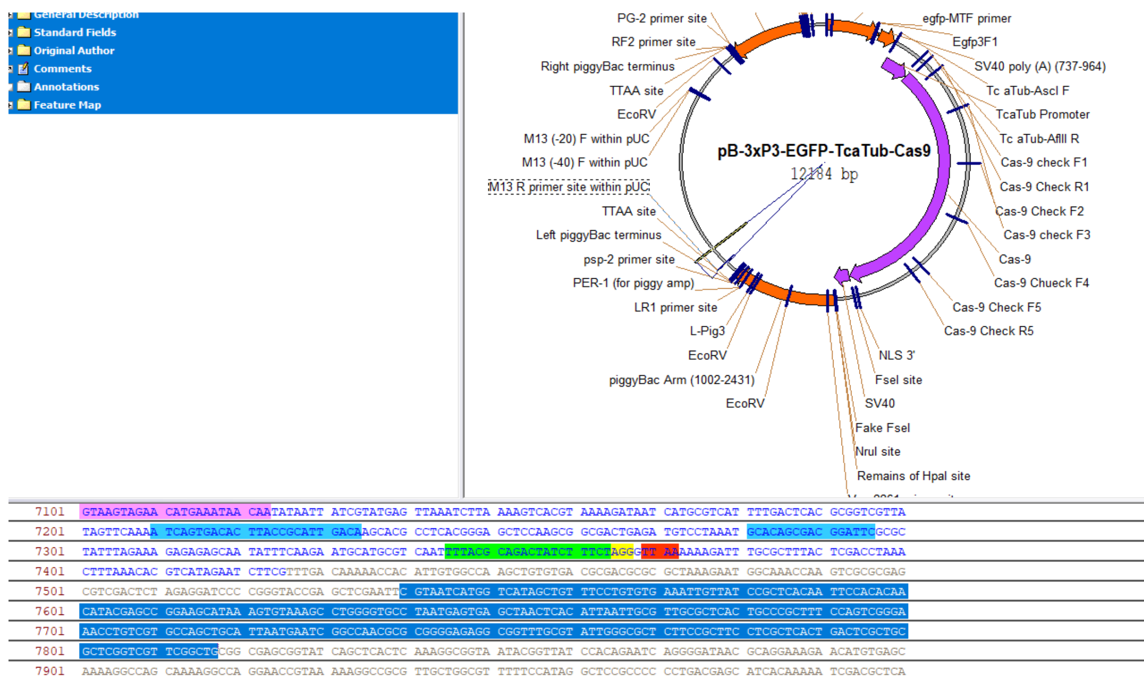
27. Place on magnetic rack for 2 min. Transfer 2° Elution with wide bore pipette tip to new tube.
28. Check DNA quality and concentration with the Nanodrop. Store at -80°C
29. Send samples to NCSU Genomics Sciences Laboratory to run TapeStation

Supplementary Material 2.4

Inverse PCR results which revealed concatemerization of piggyBac construct



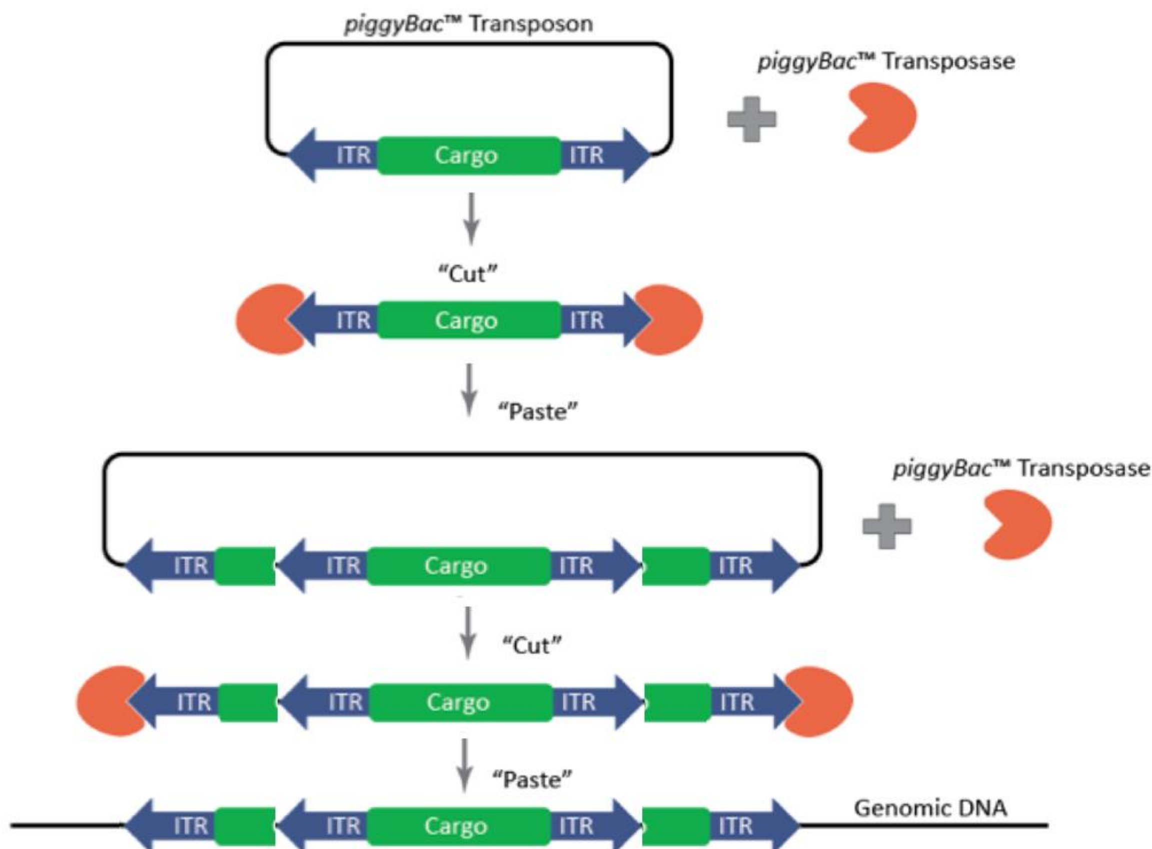
Note. Cas9 Line 9 has EGFP sequence (as highlighted) immediately downstream piggyBac arm R1 and TTAA site. If there wasn't concatemerization, TTAA site would be next to genomic DNA.



Note. Cas9 Line 3 has backbone sequence (as highlighted) immediately downstream piggyBac arm R1 and TTAA site.

Supplementary Material 2.5

Concatemerization of piggyBac construct



Note. Normally, a *piggyBac* donor construct would undergo a single "cut and paste" process and be inserted straight into the genome. We believe our construct underwent at least two "cut and paste" processes, the first insertion being in another plasmid construct and the second into the genome. This resulted in genomic insertion of an intact *Cas9*, marker gene, and *piggyBac* arms, and additional incomplete fragments of the construct. Each line had different fragments, and the figure shown is an example of one possibility. Original figure found at https://blog.addgene.org/piggybac-ing-through-the-genome-editing-field.

Supplementary Material 2.6

Cas9 insertion location and piggyBac junctions

Line

1 *cagactatctttctaggg***ttaa**cgaggcgatggaataaaaccgcattttcctccgttctgcgattctgaaag

2 *cagactatctttctaggg***ttaa**cgaggcgatggaataaaaccgcattttcctccgttctgcgattctgaaag

3 maps to highly repetitive region

4 *cagactatctttctaggg***ttaa**tggtcagtttactaaattctttcggaataaataaaaaagataaaactagc

5 *cagactatctttctaggg***ttaa**attatcttttaaattgttcttcagataattttttttcttcactgaaagct

6 *cagactatctttctaggg***ttaa**ccgtggcgtccggaattgtgcacgtttggacttgaataaatttattattg

7 *cagactatctttctaggg***ttaa**ccgtggcgtccggaattgtgcacgtttggacttgaataaatttattattg

8 *cagactatctttctaggg***ttaa**ttgcaataatttttactaaatgcgtaatatgtgacaaaaataattaacaa

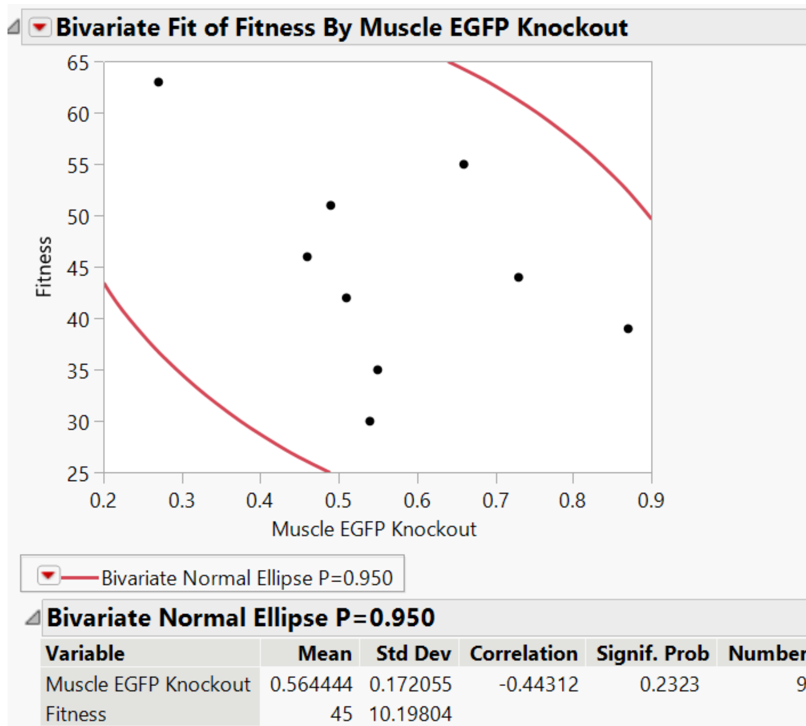
9 *cagactatctttctaggg***ttaa**ccgtggcgtccggaattgtgcacgtttggacttgaataaatttattattg

Note. Left *piggyBac* arm is shown in italics, the TTAA target site in bold, followed by Tribolium genomic sequence. Lines 1 and 2 are identical as well as Lines 6, 7, and 9.

Supplementary Material 2.7

Correlation coefficient between fitness and knockout rates

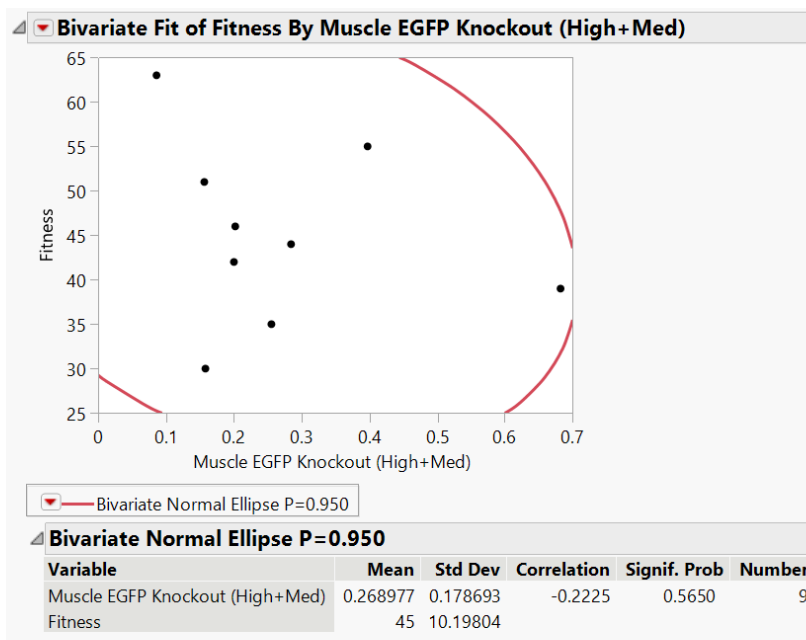
A)



Bivariate Normal Ellipse P=0.950

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
Muscle EGFP Knockout	0.564444	0.172055	-0.44312	0.2323	9
Fitness	45	10.19804			

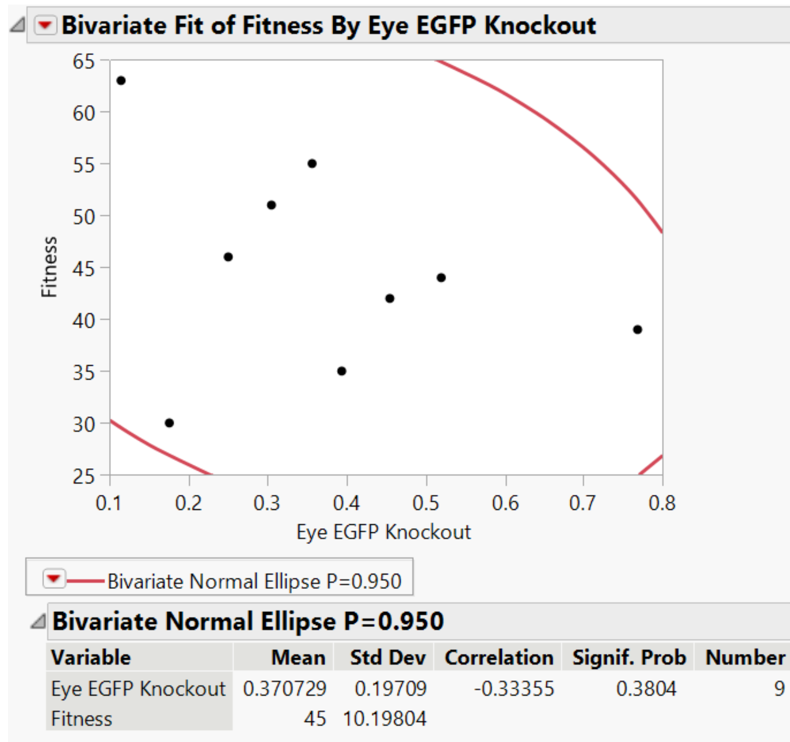
B)



Bivariate Normal Ellipse P=0.950

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
Muscle EGFP Knockout (High+Med)	0.268977	0.178693	-0.2225	0.5650	9
Fitness	45	10.19804			

C)



Bivariate Normal Ellipse P=0.950

Variable	Mean	Std Dev	Correlation	Signif. Prob	Number
Eye EGFP Knockout	0.370729	0.19709	-0.33355	0.3804	9
Fitness	45	10.19804			

Supplementary Material 2.8

Gametic EGFP muscle knockout

Line	Promoter	K/O	Replicate	Pupae	Body EGFP	Eye EGFP
1	Pub	High	1	119	3	34
1	Pub	High	2	8	0	0
1	Pub	High	3	84	0	31
1	Pub	High	4	78	0	1
1	Pub	High	5	88	0	27
1	Pub	Med	1	39	5	9
1	Pub	Med	2	46	22	33
1	Pub	Med	3	103	34	66
1	Pub	Med	4	74	40	51
1	Pub	Med	5	26	12	17
1	Pub	Low	1	50	23	16
1	Pub	Low	2	0	0	0
1	Pub	Low	3	130	61	95
1	Pub	Low	4	3	0	0
1	Pub	Low	5	67	37	35
1	Pub	Non	1	99	47	51
1	Pub	Non	2	45	29	29
1	Pub	Non	3	0	0	0
1	Pub	Non	4	98	54	54
1	Pub	Non	5	0	0	0

Line	Promoter	K/O	Replicate	Pupae	Body EGFP	Eye EGFP
2	Pub	High	1	0	0	0
2	Pub	High	2	75	10	7
2	Pub	High	3	104	10	15
2	Pub	High	4	0	0	0
2	Pub	High	5	0	0	0
2	Pub	Med	1	91	33	55
2	Pub	Med	2	128	40	57
2	Pub	Med	3	116	31	45
2	Pub	Med	4	110	42	68
2	Pub	Med	5	100	29	54
2	Pub	Low	1	104	52	69
2	Pub	Low	2	112	40	81
2	Pub	Low	3	81	26	36
2	Pub	Low	4	140	8	10
2	Pub	Low	5	60	13	22
2	Pub	Non	1	89	26	45
2	Pub	Non	2	81	31	43
2	Pub	Non	3	0	0	0
2	Pub	Non	4	102	38	60
2	Pub	Non	5	120	78	84

Line	Promoter	K/O	Replicate	Pupae	Body EGFP	Eye EGFP
3	Pub	High	1	90	1	1
3	Pub	High	2	75	1	32
3	Pub	High	3	124	3	43
3	Pub	High	4	0		
3	Pub	High	5	95	19	40
3	Pub	Med	1	0	0	0
3	Pub	Med	2	91	12	21
3	Pub	Med	3	108	20	18
3	Pub	Med	4	93	32	44
3	Pub	Med	5	123	26	57
3	Pub	Low	1	116	46	69
3	Pub	Low	2	104	46	74
3	Pub	Low	3	103	61	89
3	Pub	Low	4	112	12	20
3	Pub	Low	5	94	48	67
3	Pub	Non	1	100	46	69
3	Pub	Non	2	105	49	79
3	Pub	Non	3	69	23	34
3	Pub	Non	4	119	51	24
3	Pub	Non	5	107	48	74

Line	Promoter	K/O	Replicate	Pupae	Body EGFP	Eye EGFP
4	Pub	High	1	2	0	0
4	Pub	High	2	46	8	9
4	Pub	High	3	111	36	46
4	Pub	High	4	106	3	3
4	Pub	High	5	2	0	0
4	Pub	Med	1	101	11	32
4	Pub	Med	2	118	50	83
4	Pub	Med	3	109	45	80
4	Pub	Med	4	79	36	53
4	Pub	Med	5	111	44	76
4	Pub	Low	1	0	0	0
4	Pub	Low	2	63	29	31
4	Pub	Low	3	126	45	75
4	Pub	Low	4	0	0	0
4	Pub	Low	5	0	0	0
4	Pub	Non	1	107	56	75
4	Pub	Non	2	76	40	48
4	Pub	Non	3	119	42	68
4	Pub	Non	4	124	55	86
4	Pub	Non	5	0	0	0

Line	Promoter	K/O	Replicate	Pupae	Body EGFP	Eye EGFP
5	αTub	High	1	0	0	0
5	αTub	High	2	0	0	0
5	αTub	High	3	0	0	0
5	αTub	High	4	0	0	0
5	αTub	High	5	0	0	0
5	αTub	Med	1	0	0	0
5	αTub	Med	2	0	0	0
5	αTub	Med	3	0	0	0
5	αTub	Med	4	0	0	0
5	αTub	Med	5	0	0	0
5	αTub	Low	1	125	63	77
5	αTub	Low	2	0	0	0
5	αTub	Low	3	0	0	0
5	αTub	Low	4	0	0	0
5	αTub	Low	5	0	0	0
5	αTub	Non	1	111	65	65
5	αTub	Non	2	99	34	59
5	αTub	Non	3	102	51	73
5	αTub	Non	4	0	0	0
5	αTub	Non	5	1	1	1

Line	Promoter	K/O	Replicate	Pupae	Body EGFP	Eye EGFP
6	αTub	High	1	102	0	45
6	αTub	High	2	103	0	43
6	αTub	High	3	110	0	42
6	αTub	High	4	101	0	43
6	αTub	High	5	89	0	25
6	αTub	Med	1	120	45	70
6	αTub	Med	2	82	10	31
6	αTub	Med	3	111	57	78
6	αTub	Med	4	100	17	41
6	αTub	Med	5	84	13	22
6	αTub	Low	1	103	50	50
6	αTub	Low	2	123	60	60
6	αTub	Low	3	109	36	63
6	αTub	Low	4	5	2	2
6	αTub	Low	5	20	12	13
6	αTub	Non	1	25	10	14
6	αTub	Non	2	124	61	88
6	αTub	Non	3	100	56	72
6	αTub	Non	4	112	51	79
6	αTub	Non	5	72	31	46

Line	Promoter	K/O	Replicate	Pupae	Body EGFP	Eye EGFP
7	αTub	High	1	0	0	0
7	αTub	High	2	72	0	24
7	αTub	High	3	0	0	0
7	αTub	High	4	79	6	9
7	αTub	High	5	90	8	20
7	αTub	Med	1	95	18	37
7	αTub	Med	2	112	47	65
7	αTub	Med	3	122	18	25
7	αTub	Med	4	0	0	0
7	αTub	Med	5	111	22	41
7	αTub	Low	1	115	46	67
7	αTub	Low	2	115	46	67
7	αTub	Low	3	102	37	57
7	αTub	Low	4	99	20	53
7	αTub	Low	5	125	48	75
7	αTub	Non	1	2	1	2
7	αTub	Non	2	115	45	77
7	αTub	Non	3	0	0	0
7	αTub	Non	4	116	63	77
7	αTub	Non	5	117	53	67

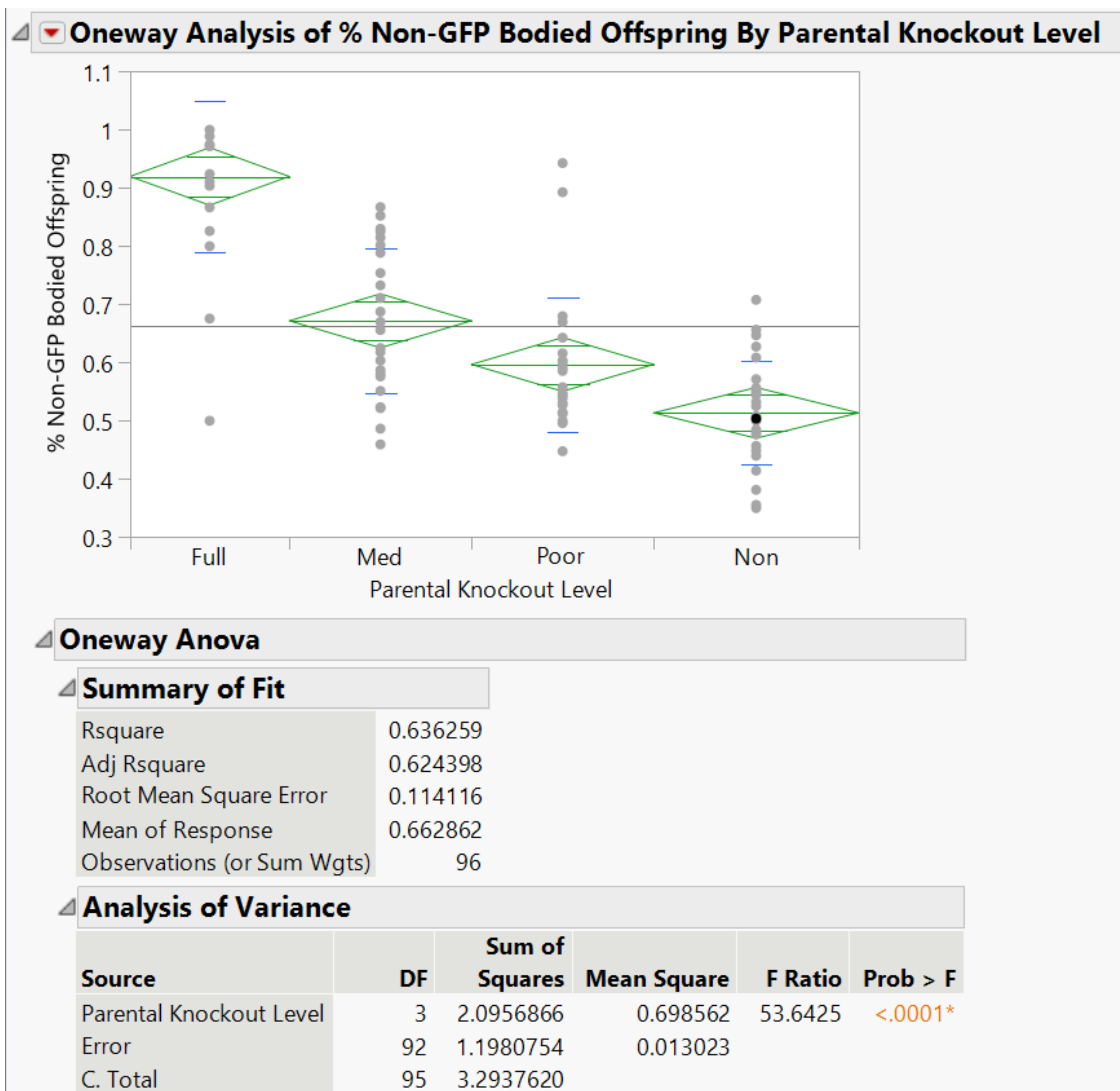
Line	Promoter	K/O	Replicate	Pupae	Body EGFP	Eye EGFP
8	αTub	High	1	2	1	1
8	αTub	High	2	0	0	0
8	αTub	High	3	0	0	0
8	αTub	High	4	0	0	0
8	αTub	High	5	0	0	0
8	αTub	Med	1	108	9	24
8	αTub	Med	2	128	37	66
8	αTub	Med	3	106	1	6
8	αTub	Med	4	126	31	58
8	αTub	Med	5	120	21	56
8	αTub	Low	1	108	47	72
8	αTub	Low	2	129	61	81
8	αTub	Low	3	4	3	3
8	αTub	Low	4	110	45	78
8	αTub	Low	5	148	67	97
8	αTub	Non	1	131	65	82
8	αTub	Non	2	101	48	72
8	αTub	Non	3	116	63	86
8	αTub	Non	4	130	67	101
8	αTub	Non	5	100	52	74

Line	Promoter	K/O	Replicate	Pupae	Body EGFP	Eye EGFP
9	αTub	High	1	103	0	57
9	αTub	High	2	144	0	81
9	αTub	High	3	41	0	19
9	αTub	High	4	119	0	66
9	αTub	High	5	7	0	5
9	αTub	Med	1	113	15	26
9	αTub	Med	2	136	61	92
9	αTub	Med	3	2	1	1
9	αTub	Med	4	149	71	111
9	αTub	Med	5	0	0	0
9	αTub	Low	1	40	18	26
9	αTub	Low	2	99	41	67
9	αTub	Low	3	89	43	63
9	αTub	Low	4	114	52	77
9	αTub	Low	5	128	41	78
9	αTub	Non	1	139	65	98
9	αTub	Non	2	116	65	87
9	αTub	Non	3	118	73	91
9	αTub	Non	4	133	69	97
9	αTub	Non	5	78	44	66

Source	K/O	Replicate	Pupae	Body EGFP
Protein	High	1	64	0
Protein	High	2	0	0
Protein	High	3	63	0
Protein	High	4	65	0
Protein	High	5	49	0
Protein	Med	1	57	0
p(bhsp68Cas9)	High	1	56	14
p(bhsp68Cas9)	High	2	93	19
p(bhsp68Cas9)	Med	1	117	15
p(bhsp68Cas9)	Med	2	96	20
p(bhsp68Cas9)	Med	3	120	52
p(bhsp68Cas9)	Med	4	111	33
p(bhsp68Cas9)	Med	5	137	49
p(bhsp68Cas9)	Low	1	80	51
p(bhsp68Cas9)	Low	2	110	41
p(bhsp68Cas9)	Low	3	82	23
p(bhsp68Cas9)	Low	4	113	47
p(bhsp68Cas9)	Low	5	103	40
p(bhsp68Cas9)	Non	1	82	52
p(bhsp68Cas9)	Non	2	86	51
p(bhsp68Cas9)	Non	3	116	47
p(bhsp68Cas9)	Non	4	85	49
p(bhsp68Cas9)	Non	5	104	54

Supplementary Material 2.9

One-way ANOVA



Oneway Anova

Summary of Fit

Rsquare	0.636259
Adj Rsquare	0.624398
Root Mean Square Error	0.114116
Mean of Response	0.662862
Observations (or Sum Wgts)	96

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Parental Knockout Level	3	2.0956866	0.698562	53.6425	<.0001*
Error	92	1.1980754	0.013023		
C. Total	95	3.2937620			

Supplementary Material 2.10

Gametic EGFP muscle knockout rates from p(bhsp68Cas9), purified protein, and the top two performing transgenic lines

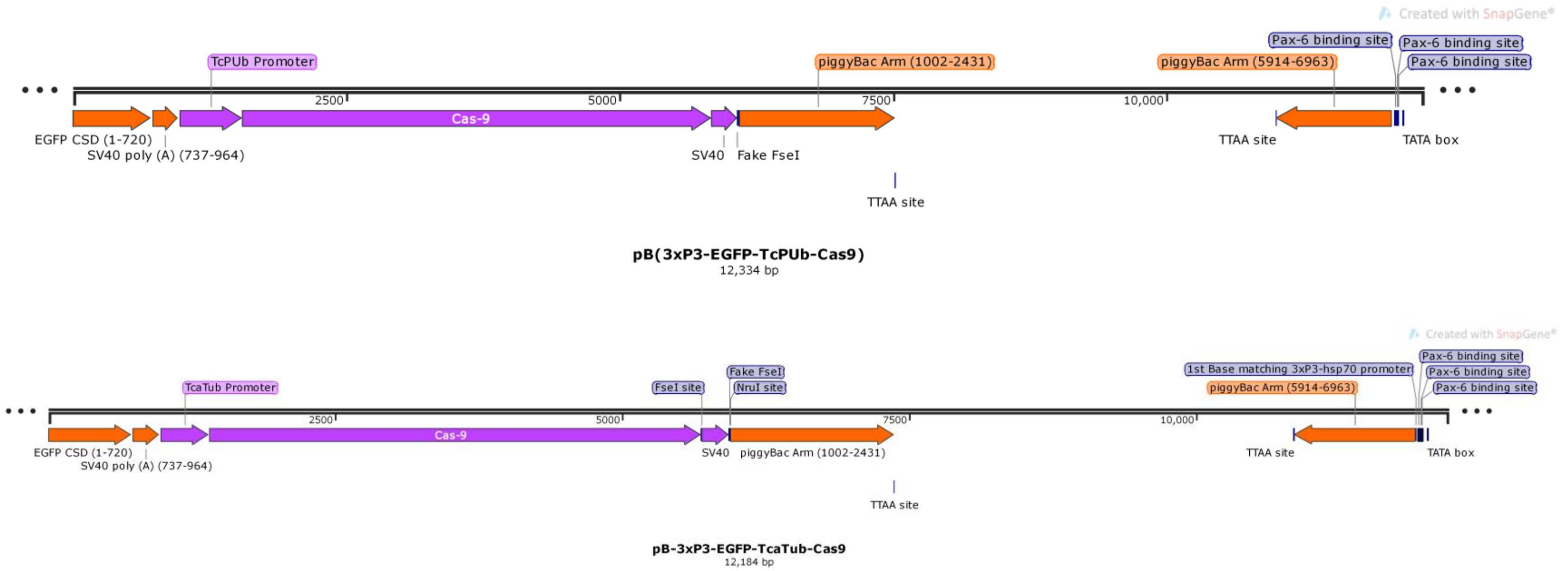
Source	K/O	Replicate	Pupae	Body EGFP
Protein	High	1	64	0
Protein	High	2	0	0
Protein	High	3	63	0
Protein	High	4	65	0
Protein	High	5	49	0
Protein	Med	1	57	0
p(bhsp68Cas9)	High	1	56	14
p(bhsp68Cas9)	High	2	93	19
p(bhsp68Cas9)	Med	1	117	15
p(bhsp68Cas9)	Med	2	96	20
p(bhsp68Cas9)	Med	3	120	52
p(bhsp68Cas9)	Med	4	111	33
p(bhsp68Cas9)	Med	5	137	49
p(bhsp68Cas9)	Low	1	80	51
p(bhsp68Cas9)	Low	2	110	41
p(bhsp68Cas9)	Low	3	82	23
p(bhsp68Cas9)	Low	4	113	47
p(bhsp68Cas9)	Low	5	103	40

Supplementary Material 2.10 *(Continued)*

Source	K/O	Replicate	Pupae	Body EGFP
p(bhsp68Cas9)	Non	1	82	52
p(bhsp68Cas9)	Non	2	86	51
p(bhsp68Cas9)	Non	3	116	47
p(bhsp68Cas9)	Non	4	85	49
p(bhsp68Cas9)	Non	5	104	54

Supplementary Material 2.11

SnapGene feature map of pB(TcPubCas9-3xP3EGFP) and *pB(TcαTubCas9-3xP3EGFP)*



Supplementary Material 2.12

>pB(TcPubCas9-3xP3EGFP)

atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttca

gcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgcc

ctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa

gtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaa

gttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagct

ggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgcca

caacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccga

caaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgacc

gccgccgggatcactctcggcatggacgagctgtacaagtaaagcggccgcgactctagatcataatcagccataccacatttgtag

aggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagc

ttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatc

aatgtatcgatacgcgtacggcgcgcctattgtcgtccgtatttacatgataaacaagactttagaacaacaacgataaggataatttttt

gtaaacaaataaattcctttcaaaattgaacagacggaaatttggcaatttcatgcaggtattaatgggtggtttatttgaaacgtttgaaa

agcataaattggaaaataaattaaaatttgaaaaataaaaaagggcgtaattccaaaacaaagaaaacagctgatcgataaagct

gacgacgtagataaaaatcagctgacatgtcaccaacaaatactggccaagtccctagaatcttgttcaattaaaagcaattttcttaat

acctgaaacattgttttgaccacgattaaccgaattggcattaaactgatagcaactaaagtttatgtataaatttttgattaaagaggacg

ttttttggataattagtcgaaaattctagaaagtgtactttgggttcgagagttaaaaaggagtgggggaagccgtccgtataaaagcgg

ttccacagctgcattcgccatattgtctgagcaaaagtggtgagacaggtcgtacttaagcgcgcctagaatggactataaggaccac

gacggagactacaaggatcatgatattgattacaaagacgatgacgataagatggccccaaagaagaagcggaaggtcggtatcc

acggagtcccagcagccgacaagaagtacagcatcggcctggacatcggcaccaactctgtgggctgggccgtgatcaccgacg

agtacaaggtgcccagcaagaaattcaaggtgctgggcaacaccgaccggcacagcatcaagaagaacctgatcggagccctg

ctgttcgacagcggcgaaacagccgaggccacccggctgaagagaaccgccagaagaagatacaccagacggaagaaccgg

atctgctatctgcaagagatcttcagcaacgagatggccaaggtggacgacagcttcttccacagactggaagagtccttcctggtgg

aagaggataagaagcacgagcggcaccccatcttcggcaacatcgtggacgaggtggcctaccacgagaagtaccccaccatct

accacctgagaaagaaactggtggacagcaccgacaaggccgacctgcggctgatctatctggccctggcccacatgatcaagttc
cggggccacttcctgatcgagggcgacctgaaccccgacaacagcgacgtggacaagctgttcatccagctggtgcagacctaca
accagctgttcgaggaaaaccccatcaacgccagcggcgtggacgccaaggccatcctgtctgccagactgagcaagagcagac
ggctggaaaatctgatcgcccagctgcccggcgagaagaagaatggcctgttcggaaacctgattgccctgagcctgggcctgacc
cccaacttcaagagcaacttcgacctggccgaggatgccaaactgcagctgagcaaggacacctacgacgacgacctggacaac
ctgctggcccagatcggcgaccagtacgccgacctgtttctggccgccaagaacctgtccgacgccatcctgctgagcgacatcctg
agagtgaacaccgagatcaccaaggcccccctgagcgcctctatgatcaagagatacgacgagcaccaccaggacctgaccctg
ctgaaagctctcgtgcggcagcagctgcctgagaagtacaaagagattttcttcgaccagagcaagaacggctacgccggctacatt
gacggcggagccagccaggaagagttctacaagttcatcaagcccatcctggaaaagatggacggcaccgaggaactgctcgtg
aagctgaacagagaggacctgctgcggaagcagcggaccttcgacaacggcagcatcccccaccagatccacctgggagagct
gcacgccattctgcggcggcaggaagatttttacccattcctgaaggacaaccgggaaaagatcgagaagatcctgaccttccgcat
cccctactacgtgggccctctggccaggggaaacagcagattcgcctggatgaccagaaagagcgaggaaaccatcaccccctg
gaacttcgaggaagtggtggacaagggcgcttccgcccagagcttcatcgagcggatgaccaacttcgataagaacctgcccaac
gagaaggtgctgcccaagcacagcctgctgtacgagtacttcaccgtgtataacgagctgaccaaagtgaaatacgtgaccgaggg
aatgagaaagcccgccttcctgagcggcgagcagaaaaaggccatcgtggacctgctgttcaagaccaaccggaaagtgaccgt
gaagcagctgaaagaggactacttcaagaaaatcgagtgcttcgactccgtggaaatctccggcgtggaagatcggttcaacgcctc
cctgggcacataccacgatctgctgaaaattatcaaggacaaggacttcctggacaatgaggaaaacgaggacattctggaagata
tcgtgctgaccctgacactgtttgaggacagagagatgatcgaggaacggctgaaaacctatgcccacctgttcgacgacaaagtga
tgaagcagctgaagcggcggagatacaccggctggggcaggctgagccggaagctgatcaacggcatccgggacaagcagtcc
ggcaagacaatcctggatttcctgaagtccgacggcttcgccaacagaaacttcatgcagctgatccacgacgacagcctgacctttа
aagaggacatccagaaagcccaggtgtccggccagggcgatagcctgcacgagcacattgccaatctggccggcagccccgcc
attaagaagggcatcctgcagacagtgaaggtggtggacgagctcgtgaaagtgatgggccggcacaagcccgagaacatcgtg
atcgaaatggccagagagaaccagaccacccagaagggacagaagaacagccgcgagagaatgaagcggatcgaagaggg
catcaaagagctgggcagccagatcctgaaagaacaccccgtggaaaacacccagctgcagaacgagaagctgtacctgtacta
cctgcagaatgggcgggatatgtacgtggaccaggaactggacatcaaccggctgtccgactacgatgtggaccatatcgtgcctca
gagctttctgaaggacgactccatcgacaacaaggtgctgaccagaagcgacaagaaccggggcaagagcgacaacgtgccct

ccgaagaggtcgtgaagaagatgaagaactactggcggcagctgctgaacgccaagctgattacccagagaaagttcgacaatct
gaccaaggccgagagaggcggcctgagcgaactggataaggccggcttcatcaagagacagctggtggaaacccggcagatca
caaagcacgtggcacagatcctggactcccggatgaacactaagtacgacgagaatgacaagctgatccgggaagtgaaagtga
tcaccctgaagtccaagctggtgtccgatttccggaaggatttccagttttacaaagtgcgcgagatcaacaactaccaccacgccca
cgacgcctacctgaacgccgtcgtgggaaccgccctgatcaaaaagtaccctaagctggaaagcgagttcgtgtacggcgactaca
aggtgtacgacgtgcggaagatgatcgccaagagcgagcaggaaatcggcaaggctaccgccaagtacttcttctacagcaacat
catgaactttttcaagaccgagattaccctggccaacggcgagatccggaagcggcctctgatcgagacaaacggcgaaaccggg
gagatcgtgtgggataagggccgggattttgccaccgtgcggaaagtgctgagcatgccccaagtgaatatcgtgaaaaagaccga
ggtgcagacaggcggcttcagcaaagagtctatcctgcccaagaggaacagcgataagctgatcgccagaaagaaggactggg
accctaagaagtacggcggcttcgacagccccaccgtggcctattctgtgctggtggtggccaaagtggaaaagggcaagtccaag
aaactgaagagtgtgaaagagctgctggggatcaccatcatggaaagaagcagcttcgagaagaatcccatcgactttctggaagc
caagggctacaaagaagtgaaaaaggacctgatcatcaagctgcctaagtactccctgttcgagctggaaaacggccggaagag
aatgctggcctctgccggcgaactgcagaagggaaacgaactggccctgccctccaaatatgtgaacttcctgtacctggccagcca
ctatgagaagctgaagggctcccccgaggataatgagcagaaacagctgtttgtggaacagcacaagcactacctggacgagatc
atcgagcagatcagcgagttctccaagagagtgatcctggccgacgctaatctggacaaagtgctgtccgcctacaacaagcaccg
ggataagcccatcagagagcaggccgagaatatcatccacctgtttaccctgaccaatctgggagcccctgccgccttcaagtacttt
gacaccaccatcgaccggaagaggtacaccagcaccaaagaggtgctggacgccaccctgatccaccagagcatcaccggcct
gtacgagacacggatcgacctgtctcagctgggaggcgacaaaaggccggcggccacgaaaaaggccggccaggcaaaaaa
gaaaaagtaaggccggcccgactctagatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctc
cccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcac
aaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcgatggccgggatctcgcgataactt
atacggttcttgtagtaagtttttgccaaagggattgaggtgaaccaattgtcacacgtaatattacgacaactaccgtgcacaggctttg
ataactccttcacgtagtattcaccgagtggtactccgttggtctgtgttcctcttcccaaataaggcattccatttatcatatacttcgtacca
ctgtcacacatcatgaggatttttattccatacttacttggcttgtttgggatatacatcctaaacggacaccgtcctctaaaaccaagtaac
tgttcatctatggtcaaatgagcccctggagtgtaattttgtatgcactgatggataaagagatcccatatttttctaacaggagtaaataca
tcgttttctcgaagtgtgggccgtatactttgtcatccattctaagacatcgtatcaaaaaatccaaaacgatccacagactcattacaga

gacgtacacattgacaaagatcgatccaaagaggtcatctgtggacatgtggttatcttttctcactgctgtcattaccagaataccaaag

aaagcatagatttcatcttcattcgtgtcacgaaatgtagcacctgtcatagattcccgacgtttcaatgatatctcagcatttgtccattttac

aatttgcgaaattatctcatcagtaaaaaatagtttgaagcataaaagtgggtcatatatattgcggcacatacgcgtcggacctctttga

gatctgacaatgttcagtgcagagactcggctaccgctcgtggactttgaagttgaccaacaatgtttattcttacctctaatagtcctctgt

ggcaaggtcaagattctgttagaagccaatgaagaacctggttgttcaataacattttgttcgtctaatatttcactacgcttcgacgttggc

tgacactcatgtacctcatctataaacgcttcttctgtatcgctctggacgtcttcacttacgtgatctgatatttcactgtcagaatcctcacc

aacaagctcgtcatcgctttgcagaagagcagagaggatatgctcatcgtctaaagaactacccattttattatatattagtcacgatatc

tataacaagaaaatatatatataataagttatcacgtaagtagaacatgaaataacaatataattatcgtatgagttaaatcttaaaagtc

acgtaaaagataatcatgcgtcattttgactcacgcggtcgttatagttcaaaatcagtgacacttaccgcattgacaagcacgcctcac

gggagctccaagcggcgactgagatgtcctaaatgcacagcgacggattcgcgctatttagaaagagagagcaatatttcaagaat

gcatgcgtcaattttacgcagactatctttctagggttaaaaaagatttgcgctttactcgacctaaactttaaacacgtcatagaatcttcgt

ttgacaaaaaccacattgtggccaagctgtgtgacgcgacgcgcgctaaagaatggcaaaccaagtcgcgcgagcgtcgactcta

gaggatccccgggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaaca

tacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttc

cagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgctt

cctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacaga

atcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcg

tttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaag

ataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgg

gaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccc c

cgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagcc

actggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagg

acagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggt

agcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacg

ctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaa

gttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatt

tcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgatacc

gcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaac

tttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgct

acaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgtt

gtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactg

cataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcga

ccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttc

ggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactt

tcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatac

tcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaa

taggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattattatcatgacattaacctataaaaataggcgt

atcacgaggccctttcgtctcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtct

gtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcat

cagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgcca

ttcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgct

gcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgccaagctttgtttaaaatataa

caaaattgtgatcccacaaaatgaagtggggcaaaatcaaataattaactagtgtccgtaaacttgttggtcttcaactttttgaggaaca

cgttggacggcaaatcgtgactataacacaagttgatttaataattttagccaacacgtcgggctgcgtgtttttgcgctctgtgtacacgt

tgattaactggtcgattaaataatttaatttttggttcttctttaaatctgtgatgaaattttttaaaataactttaaattcttcattggtaaaaaatg

ccacgttttgcaacttgtgagggtctaatatgaggtcaaactcagtaggagttttatccaaaaaagaaaacatgattacgtctgtacacg

aacgcgtattaacgcagagtgcaaagtataagagggttaaaaaatatattttacgcaccatatacgcatcgggttgatatcgttaatatg

gatcaatttgaacagttgattaacgtgtctctgctcaagtctttgatcaaaacgcaaatcgacgaaaatgtgtcggacaatatcaagtcg

atgagcgaaaaactaaaaaggctagaatacgacaatctcacagacagcgttgagatatacggtattcacgacagcaggctgaata

ataaaaaaattagaaactattatttaaccctagaaagataatcatattgtgacgtacgttaaagataatcatgcgtaaaattgacgcatgt

gtttttatcggtctgtatatcgaggtttatttattaatttgaatagatattaagttttattatatttacacttacatactaataataaattcaacaaac

aatttatttatgtttatttatttattaaaaaaaaacaaaaactcaaaatttcttctaaagtaacaaaactttaaacattctctcttttacaaaaat

aaacttattttgtactttaaaaacagtcatgttgtattataaaataagtaattagcttaacttatacataatagaaacaaattatacttattagtc

agtccagaaacaactttggcacatatcaatattatgctctcgacaaataactttttgcatttttgcacgatgcatttgcctttcgccttatttta

gaggggcagtaagtacagtaagtacgtttttcattactggctcttcagtactgtcatctgatgtaccaggcacttcatttggcaaaatatta

gagatattatcgcgcaaatatctcttcaaagtaggagcttctaaacggttacgcataaacgatgacgtcaggctcatgtaaaggtttctc

ataaatttttgcgactttgaaccttttctcccttgctactgacattatggctgtatataataaaagaatttatgcaggcaatgtttatcattccgt

acaataatgccataggccacctattcgtcttcctactgcaggtcatcacagaacacatttggtctagcgtgtccactccgcctttagtttgat

tataatacataaccatttgcggtttaccggtactttcgttgatagaagcatcctcatcacaagatgataataagtataccatcttagctggct

tcggtttatatgagacgagagtaaggggtccgtcaaaacaaaacatcgatgttcccactggcctggagcgactgtttttcagtacttccg

gtatctcgcgtttgtttgatcgcacggttcccacaatggttgaattaattcgagctcgcccggggatctaattcaattagagactaattcaatt

agagctaattcaattaggatccaagcttatcgatttcgaaccctcgaccgccggagtataaatagaggcgcttcgtctacggagcgac

aattcaattcaaacaagcaaagtgaacacgtcgctaagcgaaagctaagcaaataaacaagcgcagctgaacaagctaaacaat

cggggtaccgctagagtcgacggtaccgcgggcccgggatccaccggtcgccacc

Supplementary Material 2.13

> pB(TcαTubCas9-3xP3EGFP)

atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttca
gcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgcc
ctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa
gtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaa
gttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagct
ggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgcca
caacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccga
caaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgacc
gccgccgggatcactctcggcatggacgagctgtacaagtaaagcggccgcgactctagatcataatcagccataccacatttgtag
aggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagc
ttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatc
aatgtatcgatacgcgtacggcgcgcccacactgcagtgaacggttatgatggtgcaacgaccgggacctcagctgtataaaaaatc
tgaaaataattgataatggtcgaagaccataaggaaaaataattcgaaaattaaaggcaaagtatgaaaatgctggaaataaatcg
ggaaacaccgaatttttggtcacaggaaccaacgaatttttttttgaatattgatttgaaaagagatcgagaacaaggataacaatcttc
cgttagagaccgttggtaccgacatggcggggagccaatcagaagtctggaaatcgggcgttggttgctgacgtcaacaagcttata
aatagcaacgacagttgaaaatcgaatcaaagtcgtttggaaaaagccagagcttgtatttccgaagcgtactcccgtttttctgctctttt
gtcttaagcgcgcctagaatggactataaggaccacgacggagactacaaggatcatgatattgattacaaagacgatgacgataa
gatggccccaaagaagaagcggaaggtcggtatccacggagtcccagcagccgacaagaagtacagcatcggcctggacatcg
gcaccaactctgtgggctgggccgtgatcaccgacgagtacaaggtgcccagcaagaaattcaaggtgctgggcaacaccgaccg
gcacagcatcaagaagaacctgatcggagccctgctgttcgacagcggcgaaacagccgaggccacccggctgaagagaaccg
ccagaagaagatacaccagacggaagaaccggatctgctatctgcaagagatcttcagcaacgagatggccaaggtggacgaca
gcttcttccacagactggaagagtccttcctggtggaagaggataagaagcacgagcggcaccccatcttcggcaacatcgtggac
gaggtggcctaccacgagaagtaccccaccatctaccacctgagaaagaaactggtggacagcaccgacaaggccgacctgcg
gctgatctatctggccctggcccacatgatcaagttccggggccacttcctgatcgagggcgacctgaaccccgacaacagcgacgt

ggacaagctgttcatccagctggtgcagacctacaaccagctgttcgaggaaaaccccatcaacgccagcggcgtggacgccaag
gccatcctgtctgccagactgagcaagagcagacggctggaaaatctgatcgcccagctgcccggcgagaagaagaatggcctgt
tcggaaacctgattgccctgagcctgggcctgacccccaacttcaagagcaacttcgacctggccgaggatgccaaactgcagctg
agcaaggacacctacgacgacgacctggacaacctgctggcccagatcggcgaccagtacgccgacctgtttctggccgccaaga
acctgtccgacgccatcctgctgagcgacatcctgagagtgaacaccgagatcaccaaggcccccctgagcgcctctatgatcaag
agatacgacgagcaccaccaggacctgaccctgctgaaagctctcgtgcggcagcagctgcctgagaagtacaaagagattttcttc
gaccagagcaagaacggctacgccggctacattgacggcggagccagccaggaagagttctacaagttcatcaagcccatcctgg
aaaagatggacggcaccgaggaactgctcgtgaagctgaacagagaggacctgctgcggaagcagcggaccttcgacaacggc
agcatccccccaccagatccacctgggagagctgcacgccattctgcggcggcaggaagatttttacccattcctgaaggacaaccg
ggaaaagatcgagaagatcctgaccttccgcatcccctactacgtgggccctctggccaggggaaacagcagattcgcctggatga
ccagaaagagcgaggaaaccatcaccccctggaacttcgaggaagtggtggacaagggcgcttccgcccagagcttcatcgagc
ggatgaccaacttcgataagaacctgcccaacgagaaggtgctgcccaagcacagcctgctgtacgagtacttcaccgtgtataacg
agctgaccaaagtgaaatacgtgaccgagggaatgagaaagcccgccttcctgagcggcgagcagaaaaaggccatcgtggac
ctgctgttcaagaccaaccggaaagtgaccgtgaagcagctgaaagaggactacttcaagaaaatcgagtgcttcgactccgtgga
aatctccggcgtggaagatcggttcaacgcctccctgggcacataccacgatctgctgaaaattatcaaggacaaggacttcctggac
aatgaggaaaacgaggacattctggaagatatcgtgctgaccctgacactgtttgaggacagagagatgatcgaggaacggctga
aaacctatgcccacctgttcgacgacaaagtgatgaagcagctgaagcggcggagatacaccggctggggcaggctgagccgga
agctgatcaacggcatccgggacaagcagtccggcaagacaatcctggatttcctgaagtccgacggcttcgccaacagaaacttc
atgcagctgatccacgacgacagcctgacctttaaagaggacatccagaaagcccaggtgtccggccagggcgatagcctgcacg
agcacattgccaatctggccggcagccccgccattaagaagggcatcctgcagacagtgaaggtggtggacgagctcgtgaaagt
gatgggccggcacaagcccgagaacatcgtgatcgaaatggccagagagaaccagaccacccagaagggacagaagaacag
ccgcgagagaatgaagcggatcgaagagggcatcaaagagctgggcagccagatcctgaaagaacaccccgtggaaaacacc
cagctgcagaacgagaagctgtacctgtactacctgcagaatgggcgggatatgtacgtggaccaggaactggacatcaaccggct
gtccgactacgatgtggaccatatcgtgcctcagagctttctgaaggacgactccatcgacaacaaggtgctgaccagaagcgaca
agaaccggggcaagagcgacaacgtgccctccgaagaggtcgtgaagaagatgaagaactactggcggcagctgctgaacgcc
aagctgattacccagagaaagttcgacaatctgaccaaggccgagagaggcggcctgagcgaactggataaggccggcttcatca

agagacagctggtggaaacccggcagatcacaaagcacgtggcacagatcctggactcccggatgaacactaagtacgacgag
aatgacaagctgatccgggaagtgaaagtgatcaccctgaagtccaagctggtgtccgatttccggaaggatttccagttttacaaagt
gcgcgagatcaacaactaccaccacgcccacgacgcctacctgaacgccgtcgtgggaaccgccctgatcaaaaagtaccctaa
gctggaaagcgagttcgtgtacggcgactacaaggtgtacgacgtgcggaagatgatcgccaagagcgagcaggaaatcggcaa
ggctaccgccaagtacttcttctacagcaacatcatgaactttttcaagaccgagattaccctggccaacggcgagatccggaagcgg
cctctgatcgagacaaacggcgaaaccggggagatcgtgtgggataagggccgggattttgccaccgtgcggaaagtgctgagca
tgccccaagtgaatatcgtgaaaaagaccgaggtgcagacaggcggcttcagcaaagagtctatcctgcccaagaggaacagcg
ataagctgatcgccagaaagaaggactgggaccctaagaagtacggcggcttcgacagccccaccgtggcctattctgtgctggtg
gtggccaaagtggaaaagggcaagtccaagaaactgaagagtgtgaaagagctgctggggatcaccatcatggaaagaagcag
cttcgagaagaatcccatcgactttctggaagccaagggctacaaagaagtgaaaaaggacctgatcatcaagctgcctaagtactc
cctgttcgagctggaaaacggccggaagagaatgctggcctctgccggcgaactgcagaagggaaacgaactggccctgccctcc
aaatatgtgaacttcctgtacctggccagccactatgagaagctgaagggctcccccgaggataatgagcagaaacagctgtttgtgg
aacagcacaagcactacctggacgagatcatcgagcagatcagcgagttctccaagagagtgatcctggccgacgctaatctgga
caaagtgctgtccgcctacaacaagcaccgggataagcccatcagagagcaggccgagaatatcatccacctgtttaccctgacca
atctgggagcccctgccgccttcaagtactttgacaccaccatcgaccggaagaggtacaccagcaccaaagaggtgctggacgc
caccctgatccaccagagcatcaccggcctgtacgagacacggatcgacctgtctcagctgggaggcgacaaaaggccggcggc
cacgaaaaaggccggccaggcaaaaaagaaaaagtaaggccggcccgactctagatcataatcagccataccacatttgtagag
gttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagctta
taatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaat
gtatcgatggccgggatctcgcgataacttatacggttcttgtagtaagtttttttgccaaagggattgaggtgaaccaattgtcacacgtaa
tattacgacaactaccgtgcacaggctttgataactccttcacgtagtattcaccgagtggtactccgttggtctgtgttcctcttcccaaata
aggcattccatttatcatatacttcgtaccactgtcacacatcatgaggatttttattccatacttacttggcttgtttgggatatacatcctaaa
cggacaccgtcctctaaaaccaagtaactgttcatctatggtcaaatgagcccctggagtgtaattttgtatgcactgatggataaagag
atcccatatttttctaacaggagtaaatacatcgttttctcgaagtgtgggccgtatacttttgtcatccattctaagacatcgtatcaaaaaa
tccaaaacgatccacagactcattacagagacgtacacattgacaaagatcgatccaaagaggtcatctgtggacatgtggttatctttt
ctcactgctgtcattaccagaataccaaagaaagcatagatttcatcttcattcgtgtcacgaaatgtagcacctgtcatagattcccgac

gtttcaatgatatctcagcatttgtccattttacaatttgcgaaattatctcatcagtaaaaaatagtttgaagcataaaagtgggtcatatat

attgcggcacatacgcgtcggacctctttgagatctgacaatgttcagtgcagagactcggctaccgctcgtggactttgaagttgacca

acaatgtttattcttacctctaatagtcctctgtggcaaggtcaagattctgttagaagccaatgaagaacctggttgttcaataacattttgtt

cgtctaatatttcactacgcttcgacgttggctgacactcatgtacctcatctataaacgcttcttctgtatcgctctggacgtcttcacttacgt

gatctgatatttcactgtcagaatcctcaccaacaagctcgtcatcgctttgcagaagagcagagaggatatgctcatcgtctaaagaa

ctacccattttattatatattagtcacgatatctataacaagaaaatatatatataataagttatcacgtaagtagaacatgaaataacaat

ataattatcgtatgagttaaatcttaaaagtcacgtaaaagataatcatgcgtcattttgactcacgcggtcgttatagttcaaaatcagtg

acacttaccgcattgacaagcacgcctcacgggagctccaagcggcgactgagatgtcctaaatgcacagcgacggattcgcgcta

tttagaaagagagagcaatatttcaagaatgcatgcgtcaattttacgcagactatctttctagggttaaaaaagatttgcgctttactcga

cctaaactttaaacacgtcatagaatcttcgtttgacaaaaaccacattgtggccaagctgtgtgacgcgacgcgcgctaaagaatgg

caaaccaagtcgcgcgagcgtcgactctagaggatccccgggtaccgagctcgaattcgtaatcatggtcatagctgtttcctgtgtga

aattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactca

cattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggaga

ggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactc

aaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccagg

aaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagagg

tggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttac

cggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcggtgtaggtcgttcgctc

caagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaaga

cacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtgg

tggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctctt

gatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaag

atcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttc

acctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagt

gaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttacc

atctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagg

gccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagtt

aatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacga

tcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgca

gtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaacca

agtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaacttta

aaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgt

gcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaaggga

ataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggataca

tatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaaccattatta

tcatgacattaacctataaaaataggcgtatcacgaggccctttcgtctcgcgcgtttcggtgatgacggtgaaaacctctgacacatgc

agctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgg

gtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgta

aggagaaaataccgcatcaggcgccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctatta

cgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgac

ggccagtgccaagctttgtttaaaatataacaaaattgtgatcccacaaaatgaagtggggcaaaatcaaataattaactagtgtccgt

aaacttgttggtcttcaactttttgaggaacacgttggacggcaaatcgtgactataacacaagttgatttaataattttagccaacacgtc

gggctgcgtgtttttttgcgctctgtgtacacgttgattaactggtcgattaaataatttaatttttggttcttctttaaatctgtgatgaaatttttttaa

aataactttaaattcttcattggtaaaaaatgccacgttttgcaacttgtgagggtctaatatgaggtcaaactcagtaggagttttatccaa

aaaagaaaacatgattacgtctgtacacgaacgcgtattaacgcagagtgcaaagtataagagggttaaaaaatatattttacgcac

catatacgcatcgggttgatatcgttaatatggatcaatttgaacagttgattaacgtgtctctgctcaagtctttgatcaaaacgcaaatcg

acgaaaatgtgtcggacaatatcaagtcgatgagcgaaaaactaaaaaggctagaatacgacaatctcacagacagcgttgagat

atacggtattcacgacagcaggctgaataataaaaaaattagaaactattatttaaccctagaaagataatcatattgtgacgtacgtta

aagataatcatgcgtaaaattgacgcatgtgtttttatcggtctgtatatcgaggtttatttattaatttgaatagatattaagttttattatatttac

acttacatactaataataaattcaacaaacaatttatttatgtttatttatttattaaaaaaaaaacaaaaactcaaaatttcttctaaagtaac

aaaacttttaaacattctctcttttacaaaaataaacttattttgtactttaaaaacagtcatgttgtattataaaataagtaattagcttaactta

tacataatagaaacaaattatacttattagtcagtccagaaacaactttggcacatatcaatattatgctctcgacaaataactttttttgcatt

ttttgcacgatgcatttgcctttcgccttattttagaggggcagtaagtacagtaagtacgtttttcattactggctcttcagtactgtcatctga

tgtaccaggcacttcatttggcaaaatattagagatattatcgcgcaaatatctcttcaaagtaggagcttctaaacggttacgcataaac

gatgacgtcaggctcatgtaaaggtttctcataaatttttgcgactttgaaccttttctcccttgctactgacattatggctgtatataataaaa

gaatttatgcaggcaatgtttatcattccgtacaataatgccataggccacctattcgtcttcctactgcaggtcatcacagaacacatttg

gtctagcgtgtccactccgcctttagtttgattataatacataaccatttgcggtttaccggtactttcgttgatagaagcatcctcatcacaa

gatgataataagtataccatcttagctggcttcggtttatatgagacgagagtaaggggtccgtcaaaacaaaacatcgatgttcccact

ggcctggagcgactgtttttcagtacttccggtatctcgcgtttgtttgatcgcacggttcccacaatggttgaattaattcgagctcgcccg

gggatctaattcaattagagactaattcaattagagctaattcaattaggatccaagcttatcgatttcgaaccctcgaccgccggagtat

aaatagaggcgcttcgtctacggagcgacaattcaattcaaacaagcaaagtgaacacgtcgctaagcgaaagctaagcaaataa

acaagcgcagctgaacaagctaaacaatcggggtaccgctagagtcgacggtaccgcgggcccgggatccaccggtcgccacc

CHAPTER 3

Investigating the Role of Cadherin Receptors in Coleopteran Resistance to Bt Cry3Aa

Introduction

Genetically modified crop varieties with genes from *Bacillus thuringiensis* (Bt) account for over 80% of corn and cotton acreage in the US. However, the benefits of using this pesticide are threatened by insects developing resistance. While there are many molecular mechanisms of resistance, some are not well understood particularly for beetles. This chapter reports the first investigation of beetle resistance through germline transformation. Our goal is to validate the function of a proposed Cry3Aa toxin binding region in a cadherin receptor protein by converting a Cry3Aa-tolerant beetle (*Tribolium castaneum*) into a susceptible one through two objectives: (1) Use CRISPR/Cas9 to replace small portions of the *TcCad1* gene with the equivalent portion from a related Cry3Aa-susceptible species, *Tenebrio molitor*, and (2) Compare mortality of modified and unmodified *T. castaneum* compared to the control *Te. molitor* when reared on diets either containing Cry3Aa or not. Results from objective one are covered in this chapter and objective two will be completed in the future. Feeding bioassays for the second portion of this project will be carried out by a collaborator at the USDA-ARS. A better understanding of Cry3A mode of action and target specificity will facilitate methods for resistance management to preserve the benefits of Bt foliar sprays and transgenic crops, thereby protecting food resources in the US.

History and benefits of Bt pesticides

Over 80% of US corn and cotton is genetically modified to have insecticidal traits that protect against beetle and moth pests (USDA ERS 2017). Commonly cultivated varieties carry multiple crystal (Cry) toxin genes from the bacterium *Bacillus thuringiensis* (Bt). Cry toxin genes have been used in commercial varieties of Bt corn and cotton since 1996 (Benbrook 2012). In addition to the use of Cry genes in Bt crops, many strains of the bacterium have been used as effective biopesticides since the 1950’s. Bt pesticide products include sprays, dusts, granules, and pellets. The many strains of Bt produce a wide variety of toxins, each targeting different insect groups. Bt pesticides are favored for their selectivity, specificity, and low toxicity to vertebrates and low impact on the environment (Whalon and Wingerd 2003, EPA 1998).

Bt crops have been widely adopted in the US, and Bt foliar sprays are the most widely used microbial insecticides in the US, some used in both organic and conventional production systems. In 1998 the EPA declared insect susceptibility to Bt a ‘public good’ in recognition of the benefits of Bt pesticides in US agriculture (EPA 1998, USDA ERS 2017). In 2016 the National Academies of Sciences, Engineering, and Medicine (NASEM) examined evidence of potential benefits and negative effects of genetically engineered (GE) crops, including Bt corn and cotton. Based on metanalyses, the report found Bt crop varieties generally decrease yield losses and at the same time decrease the use of insecticides on small and large US farms in comparison with non-Bt varieties (NASEM 2016, Fernandez-Cornejo et al. 2014). Decreasing the use of insecticides provide great benefits in the health of farm workers as well as reduced effects on beneficial insects. There are also possible economic benefits, although in the US outcomes on net returns to adopters have high heterogeneity, and evidence of household labor savings is extremely weak (Fernandez-Cornejo et al. 2014, Gardner et al. 2009).

Another benefit of Bt crops is higher insect biodiversity than fields of non-Bt varieties treated with synthetic insecticides. Bt crops have similar or better biodiversity compared to untreated fields (NASEM 2016, NRC 2010). Many alternatives to using Bt are more harmful to non-target animals. Bt is non-toxic to vertebrates including humans and birds and is somewhat selective in the species of insects targeted, so in some cases can be used without harming beneficial insects, especially when incorporated into the crop.

Environmental and economic benefits are threatened by the evolution of pest resistance. By 2016, seven insect species had developed resistance to Bt transgenic crops, and the average time for evolution of resistance was only 5.2 years from first commercial planting (Tabashnik and Carrière 2017). Of these seven species, six were moths and one was a beetle.

Strategies for resistance management include crop rotation, planting refuge, pyramiding Bt traits, engineering Bt molecules with altered modes of action, and currently under development, pyramiding Bt and RNA interference (RNAi) traits (Bravo 2017, Gassmann 2021). Field-evolved

resistance is monitored and managed, and mechanisms of resistance are also being investigated, though moths are more well-studied than beetles (Jurat-Fuentes et al. 2021).

Mechanisms of resistance

There are many resistance mechanisms documented against Cry proteins in insects: alteration of toxin binding proteins (aminopeptidase-N (APN), alkaline phosphatase (ALP), cadherins (Cad), ABC transporters (ABC); alterations in protoxin processing; enhanced immunity; toxin sequestration; epithelial regeneration; and others (Jurat-Fuentes et al. 2021). Changes in the toxin binding site is the most commonly occurring mechanism leading to higher resistance levels (Ferré and Van Rie 2002). Some recent findings demonstrate Cry protein binding to ABC transporters is critical for high toxicity, while interaction with the other receptors (APN, ALP, Cad) may increase toxicity and thus play an enhancer role (Jurat-Fuentes et al. 2021). Another study demonstrated ABC and Cad receptors play independent roles in the mode of action of Cry1A toxins (Wang et al. 2018). Much more is known about resistance mechanisms to Cry1 proteins (active against Lepidoptera) than Cry3 proteins (active against Coleoptera). Some Bt modes of action and resistance mechanisms are known to be conserved across lepidopteran families and even across orders Lepidoptera and Coleoptera.

Bt corn and field-evolved resistance in western corn rootworm

Bt corn protects against multiple moth pests and a single beetle pest, Western corn rootworm (WCR; *Diabrotica virgifera virgifera*). WCR costs US farmers between $1-2 billion in yield loss and control expenditures annually and is controlled by four different Cry toxins on the market (Cry3Bb, Cry34/35Ab, mCry3A, eCry3.1Ab) (Wechsler and Smith 2018, Gray et al. 2009, Metcalf 1986). The first Bt corn that protected against beetle pests was commercially available in 2003 and produced a single trait, Cry3Bb1. The other three traits (Cry34/35Ab, mCry3A, eCry3.1Ab) were registered in 2005, 2006, and 2012, respectively. Bt corn with multiple Cry traits (pyramids) were available by 2009 and are now standard.

The first evidence of field-evolved resistance to Cry3 transgenics was in 2009 to Cry3Bb (Gassmann et al. 2011). Evidence of a single population having resistance to two traits (Cry3Bb1 and mCry3A) was found in 2011, and to three traits (Cry3Bb1, mCry3A, and eCry3.1Ab) in 2012 (Gassmann et al. 2014, Jakka, Shrestha, and Gassmann 2016). In a 2012 survey of Iowa farmers, just over one-tenth reported having Bt-resistant corn rootworm in their fields (Arbuckle, 2014). Resistance to Cry34/35Ab1 was found in multiple WCR populations in 2013 (Gassmann et al. 2016). By 2017, some populations of WCR were resistant to all four Bt traits (Gassmann et al. 2020). WCR resistance has now been documented in Iowa, Illinois, Minnesota, Nebraska and North Dakota (Gassmann 2021). The resistance in WCR differs from lepidopteran resistance because low levels (<25-fold) can allow severe feeding injury to Bt corn in the field (Jurat-Fuentes, Heckel, and Ferré 2021). Little is known about the mechanistic basis of Bt resistance in WCR; however, low level resistance has been associated with protease-mediated resistance (Herrero et al. 2001). Some resistance mechanisms are conserved across lepidopteran families which also seems to be the case in Coleoptera, so studies in other beetles can inform studies in WCR.

Mechanisms of resistance to Cry3 proteins in Coleoptera.

Yellow mealworm (*Tenebrio molitor*)

The first report of a coleopteran cadherin as a functional Cry3Aa toxin receptor was in the yellow mealworm, TmCad1 (Fabrick et al. 2009). Cry3Aa bound specifically to a peptide from the mealworm cadherin, TmCad1p, and the peptide promoted toxin oligomer formation. RNAi knockdown of a region of TmCad1 reduced Cry3Aa toxicity (Table 3.1). In this study, the authors identified similarities and differences between beetles and moths in the proposed toxin binding regions (TBR) of cadherin (Figure 3.1). Similarities between TmCad1 and lepidopteran cadherins in their domain structure and putative toxin binding regions suggest functional similarities. While there is a high degree of similarity between *TmCad1* and *TcCad1* (51% identity) predicted protein sequences and even more similarity in the predicted toxin binding region, there are three lysine residues only found in the *TcCad1*. Cry3Aa is much more effective against *Te. molitor* (100%

mortality at 42 mg kg−1) than *T. castaneum* (no significant mortality with doses as high as 2100 mg kg−1) (Oppert, Morgan, and Kramer 2011). They hypothesized the lysine residues that carry positive charges could disrupt toxin binding, and therefore may explain the lower toxicity of Cry3Aa in *T. castaneum* compared to *Te. molitor*. Cry34/35 is inactive against both *T. castaneum* and *Te. molitor* (Oppert, Ellis, and Babcock 2010). This is possibly because Cry34Ab1 and Cry35Ab1 are binary toxins and therefore structurally different from most other Cry toxins (including Cry3Bb1, mCry3A and eCry3) which have three domains.

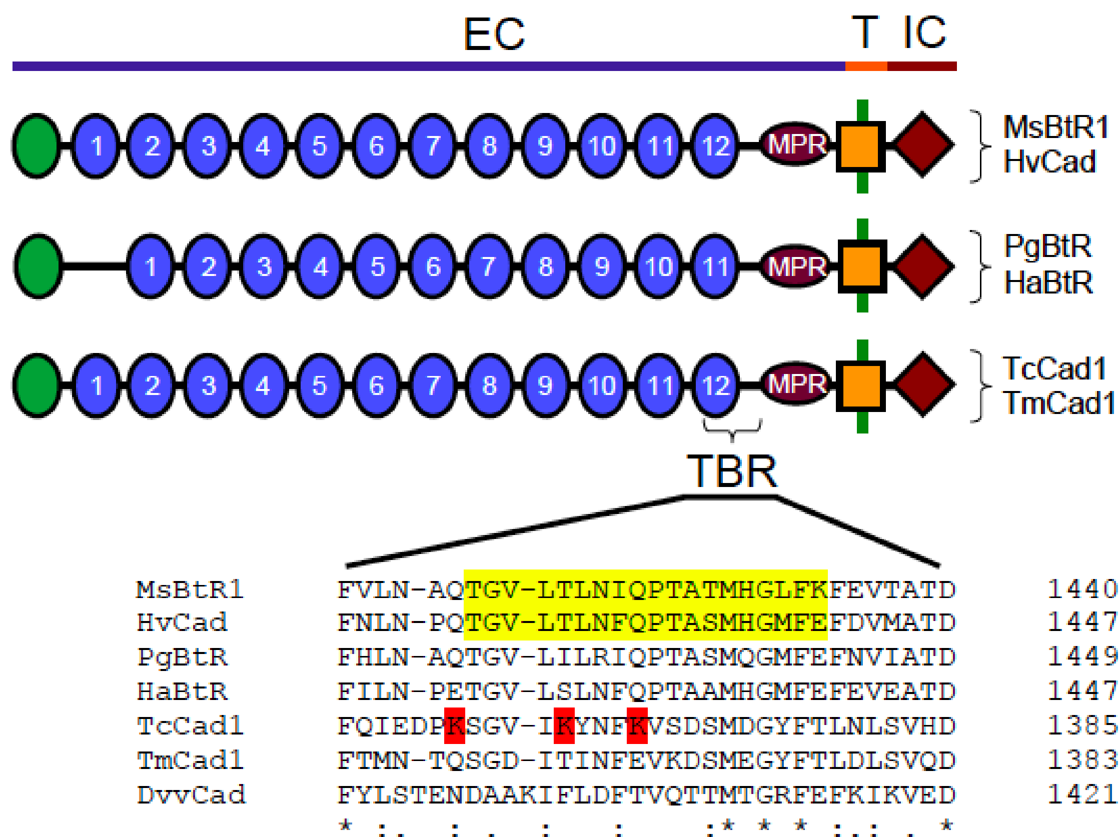
TcCad1 and TmCad1 share 51% identity, 21-24% identity to lepidopteran cadherins, and 25-27% identify to WCR DvvCad1 (Fabrick et al. 2009). Additionally, the Cry3Aa TBR in TmCad1 differs from DvvCad1. Based on alignment, the key residues within the Cry3Aa TBR are not conserved in DvvCad as they are in TcCad, TmCad, and lepidopteran cadherins (Figure 3.1). In TcCad1 and TmCad1, the proposed TBR is found in CR 12 next to the membrane-proximal region. The TBR in TmCad1 corresponds to the reported membrane-proximal region in DvvCad. Several studies have found DvvCad CR 8–10 important for toxicity (Sayed et al. 2007, Park et al. 2009, Park et al. 2014) which corresponds to CR 9–11 in TmCad1 (Fabrick et al. 2009). Given these differences, Fabrick and colleagues suggest beetle cadherins may possess different or multiple Cry3 toxin binding domains. So, not all beetles are affected by Cry3, but those that are may have different TBD. These unique sites may contribute to differences in susceptibility to different Cry3 toxins. The role of the TmCad1 TBR is further substantiated by a study that fed three vegetable pests (spotted asparagus beetle (*Crioceris quatuordecimpunctata*), cabbage leaf beetle (*Colaphellus bowringi*) and daikon leaf beetle (*Phaedon brassicae*)) a TmCad1-TBR peptide fragment combined with Cry3Aa toxin and observed up to a 15.3-fold increase in larval mortality (Gao et al. 2011).

In addition to cadherins, alkaline phosphatases (ALP) are also involved in binding of Cry3Aa toxin in *Te. molitor* (Zúñiga-Navarrete et al. 2013). Transcriptome profiling of the

intoxication response in *Te. molitor* to Cry3Aa revealed TmCad1 expression was not affected while ALP transcripts were only found in intoxicated larvae (Oppert et al. 2012).

Figure 3.1

Structure and sequence of insect cadherin Cry toxin receptors (Figure 2A Fabrick et al., 2009)



Note. The upper panel depicts the conserved domain structure, including the extracellular (EC), transmembrane (T), and intracellular (IC) domains of insect cadherins, including the proposed toxin binding region (TBR) in the cadherin repeat region (numbered) closest to the membrane-proximal-region (MPR). The lower panel is an alignment of the proposed TBR; MsBtR1 (*Manduca sexta*, AAG37912, HvCad (*Heliothis virescens*, AAK85198), PgBtR (*Pectinophora gossypiella*, AAP30715), HaBtR (*Helicoverpa armigera*, ABF69362), TmCad1 (*Tenebrio molitor* (ABL86001), TcCad1 (*Tribolium* castaneum, XP_971388). The Cry1A TBR in *M. sexta* and *H. virescens* cadherins are highlighted in yellow; lysine residues in the corresponding region of TcCad1 that are potentially disruptive to toxin binding are highlighted in red. Figure and highlights from Fabrick et al. (2009).

Western corn rootworm (*Diabrotica virgifera virgifera*)

The first WCR cadherin (GenBank accession # CN497270) was identified by expressed sequence tags from midgut tissue, and phylogenetic analysis indicated dipteran, coleopteran and lepidoperan cadherins are derived from a common ancestor (Siegfried et al 2005). The same cadherin (*DvvCad*) was later found to be a functional Cry toxin receptor (Sayed et al. 2007). The DvvCad amino acid sequence shares common architectural themes with lepidopteran cadherins including a signal peptide, cadherin repeats (CR), a membrane proximal region, transmembrane repeats, and a cytosolic region. DvvCad expression is limited to midgut tissues and transcripts are the most abundant in first, second, and third larval instars. One study demonstrated a partial fragment spanning cadherin repeats (CR) 8-10 bound Cry3Bb and significantly enhanced the toxicity of Cry3Bb in WCR, southern corn rootworm, and Colorado potato beetle (Park et al. 2009).

DvvCad-CR8-10 also increased toxicity of Cry3Aa in Colorado potato beetle and southern corn rootworm but was not tested for WCR. The same DvvCad fragment (CR 8-10) was found to enhance Cry3Aa, Cry3Bb, and Cry8Ca toxicity to lesser mealworm (Park et al. 2014). In WCR, knockdown of *DvvCad* by oral RNAi did not affect sensitivity to Cry3Aa and Cry34/35Ab1 (Tan et al. 2016). However, the percent mortality in the control treatments are much lower than other studies on effect of RNAi knockdown of binding receptors on sensitivity to Cry3 toxins (Table 3.1).

In WCR, a few potential Cry binding proteins and resistance genes have been identified. One study compared gene expression between Cry3Bb1-susceptible and Cry3Bb1-resistant WCR in the presence and absence of Cry3Bb1 (Rault et al. 2018). Differentially expressed transcripts following Cry3Bb1 treatment included ABC transporters, proteases and α-amylases, metalloproteases, cadherins and signaling proteins. Differentially expressed transcripts were inspected for polymorphisms that could interfere with Cry3Bb1 mode of action. Polymorphisms were detected in three transcripts encoding cadherins. The authors acknowledge cadherin polymorphisms as a possible resistance mechanism, with the caveat that this role would need to be confirmed because not much is known about cadherin interactions with Cry3 proteins in WCR.

That being said, changes in receptor binding are cited as the most frequent mechanism of Bt resistance, and in lepidopterans there is ample evidence of the role of cadherins in Cry1 resistance (Nagamatsu et al. 1998, Gahan, Gould, and Heckel 2001, Morin et al. 2003). One mCry3A-selected WCR colony exhibited reduced binding of mCry3A to midgut tissue and high resistance after six generations of selection (Zhao et al. 2016).

Table 3.1

Sensitivity to Cry3 toxin following RNAi knockdown of Cry toxin binding proteins

Species	Toxin	RNAi	Mortality (control)	Mortality (RNAi)	Reference
Alphitobius diaperinus	Cry3Bb	*AdCad1*	97%	8-11%	Hua et al. 2014
Diabrotica virgifera virgifera	Cry3Aa Cry34/35Ab1	*DvvCad*	10-20%	10-13%	Tan et al. 2016
Diabrotica virgifera virgifera	IP3-H9 (Cry3Aa-like)	*DvvABCB1*	52%	1%	Niu et al. 2020
Tenebrio molitor	Cry3Aa	*TmCad1*	33%	12%	Fabrick et al. 2009
Tribolium castaneum	Cry3Ba	*TcCad1*	40%	5%	Contreras et al. 2013

Note. Observed mortalities on Cry toxin following RNAi of a potential Cry binding protein (cadherin (CAD) or ABC transporter (ABC)).

ABC transporters are also Cry binding receptors and linked to resistance in WCR. *DvvABCB1* was shown to be genetically linked to resistance in the Hopkinton Cry3Bb1-resistant strain (HopR) and multiple field populations in the US (Flagel et al. 2015). *DvvABCB1* is close to CRW1862, one of two markers (the other is CRW1683) associated with Cry3Bb1 resistance across 80 field populations throughout the US Corn Belt (Flagel et al. 2015). The marker CRW1862 tags a C/T SNP polymorphism and was corroborated by a later study which showed the T allele responds to Cry3Bb1 selection across 16 WCR populations (Willse, Flagel, and Head 2021). However, both studies reported CRW1862 is not the causal allele and hypothesized

ABCB1 as a candidate gene approximately 7 cM from the marker. Heterologous expression of *DvvABCB1* in Sf9 and HEK293 cells confers sensitivity to Cry3Aa and knockdown by RNAi renders WCR larvae insensitive to Cry3Aa (Niu et al. 2020). Furthermore, reduced expression and alternatively spliced transcripts of *DvvABCB1* were identified in a mCry3A resistant strain of WCR (Niu et al. 2020).

Lesser mealworm (*Alphitobius diaperinus*)

In the lesser mealworm, AdCad1 was identified as a functional Cry3Bb receptor (Hua, Park, and Adang 2014). AdCad1 has nine cadherin repeats of which CR9 bound Cry3Bb toxin. RNAi of *AdCad1* resulted in highly reduced susceptibility to Cry3Bb which was restored when larvae were fed CR9 peptide (Table 3.1). AdCad1 CR9 has high amino acid sequence similarity to the TBR in TmCad1-CR12. A fragment of the WCR cadherin DvvCad CR 8-10 was found to enhance Cry3Aa, Cry3Bb, and Cry8Ca toxicity to lesser mealworm (Park et al. 2014).

Red flour beetle (*Tribolium castaneum*)

In red flour beetle (*Tribolium castaneum*), E-cadherin (TcCad1; ortholog of TmCad1) was identified by ligand blot as a Cry3Ba binding protein along with aminopeptidase N (TcAPN-I) and sodium solute symporter (TcSSS) (Contreras et al. 2013). RNAi knockdown of TcCad1 and TcSSS both resulted in decreased susceptibility to Cry3Ba toxin (Table 3.1). TcSSS proteins have cadherin repeats in its amino acid sequence, further supporting the role of cadherin-like proteins in Cry3 toxicity (Contreras et al. 2013, Hua et al. 2014).

Colorado potato beetle (*Leptinotarsa decemlineata*)

In Colorado potato beetle (CPB) several Cry3Aa binding proteins (mCry3A in Bt corn is a modified version of Cry3Aa) have been reported. Prohibitin-1 binds Cry3Aa and has two predicted cadherin motifs (CADHERIN2 and CADHERIN6) which align to the CR8-CR9 sequence of WCR (aa:1088-1125) (Ochoa-Campuzano et al. 2013). Cry3Aa resistance was found to correlate with decreased toxin binding and changes in the composition and activity of digestive enzymes, especially elevated aminopeptidase (APN) activity (Loseva et al. 2002). Another study showed

Cry3Aa toxin is cleaved by metalloproteases which drastically reduce pore forming activity (Rausell et al. 2007).

Leaf beetle pest of poplar trees (*Chrysomela tremula*)

In one resistant strain of *C. tremula* Cry3Aa resistance is linked to a mutation in an ABC protein, *CtABCB1*, member of the B subfamily homologous to P-glycoprotein (Pauchet et al. 2016). *CtABCB1* is a homolog of *DvABCB1*, linked to resistance in WCR (Niu et al. 2020, Willse, Flagel, and Head 2021).

Germline transformation to investigate resistance mechanisms

The recent advancement in gene editing technology CRISPR/Cas has facilitated insecticide resistance research that uses germline transformation to identify and functionally characterize resistance genes and mutations. Thus far, the majority of these types of studies have used *Drosophila* to characterize resistance mechanisms identified in other insect pests (See Douris et al. 2020 for a detailed review). Of the studies carried out in pest species, the majority have created gene knockouts of toxin receptors. Only five studies in three species of Lepidoptera have created knock-ins of specific resistance mutations (Table 3.2). All five created single amino acid substitutions in binding receptors of susceptible strains of beet armyworm (*Spodoptera exigua*), diamondback moth (*Plutella xylostella*), or cotton bollworm (*Helicoverpa armigera*). To date, there have not been any germline transformation studies investigating Cry3 toxins in beetles.

Table 3.2

Resistance ratio following CRISPR/Cas9 knock-in of proposed resistance mutations

Species	Pesticide	CRISPR/Cas9 knockin	Gene	Strain	Resistance Ratio	Reference
Helicoverpa armigera	Cry1Ac	T92C	*HaTSPAN1*	SCD	125	Jin et al. 2018
Plutella xylostella	dieldrin	A301S	*PxGABARα1*	VB	1.4	Guest et al. 2019
	fipronil				2.7	
	endosulfan				1.4	
Plutella xylostella	chlorantraniliprole	I4790M	*PxRyR*	IPP-S	6.0	Wang et al. 2020
	cyantraniliprole				7.7	
	flubendiamide				40.5	
	indoxacarb				0.97	
	β-cypermethrin				1.33	

Note. All knockins are single amino acid substitutions. Resistance ratio is the fold change in resistance, calculated LC50 of the resistant (edited) strain / LC50 of the susceptible strain. Strain listed is the background line.

Table 3.2 *(Continued)*

Species	Pesticide	CRISPR/Cas9 knockin	Gene	Strain	Resistance Ratio	Reference
Spodoptera exigua	chlorantraniliprole	G4946E	*SeRyR*	WH-S	223	Zuo et al.,
	cyantraniliprole				336	2017
	flubendiamide				1000	
Spodoptera exigua	chlorantraniliprole	I4743M	*SeRyR*	WH-S	21	Zuo et al.
	cyantraniliprole				25	2019
	flubendiamide				22	

Note. All knockins are single amino acid substitutions. Resistance ratio is the fold change in resistance, calculated LC50 of the resistant (edited) strain / LC50 of the susceptible strain. Strain listed is the background line.

Many CRISPR/Cas9 knockouts in pest species and one knock-in have investigated resistance mechanisms against Cry1 and Cry2 toxins (Douris et al. 2020, Jin et al. 2018). Cry toxins are active against specific insect groups, so resistance cannot be investigated heterologously in *Drosophila* unlike many other insecticides. Knockouts of Cry1 and Cry2 binding proteins (Cad1, ABC1, ABC2, ABC3, APN1, APN2, APN5, TSPAN1) have been created in multiple moths (*P. xylostella*, *H. armigera*, *T. ni*, *S. frugiperda*, *O. furnacalis*, *H. armigera*, *S. exigua*) (Douris et al. 2020). Three studies have investigated the role of cadherins in resistance through gene knockouts (Table 3.3). Cadherin knockouts in *H. armigera* and *T. ni* confirmed its role in Cry1Ac resistance (Wang et al. 2016, Wang et al. 2018). The *HaCad* knockout line had 549-fold resistance, while the *TnCad* knockout line only had 1.5 to 3.2-fold resistance. The authors also found the *TnABCC2* knockout to confer resistance and based on their assays conclude TnCad and TnABCC2 play independent roles in the mode of action of Cry1A toxins. Knockout of *SeCad1* in *S. exigua* conferred resistance to Cry1Ac and Cry1Fa, but not Cry1Ca (Huang et al. 2020).

The only CRISPR/Cas9 knockin of a Cry resistance mutation thus far was in cotton bollworm and constituted a single amino acid substitution (T92C) in a tetraspanin gene (*HaTSPAN1*) in the resistant strain AY2 (Jin et al. 2018). There have yet to be any cadherin or ABC transporter knockins using CRISPR/Cas9, though one study used a transposase transformation system to knockin an ABC transporter. The role of an ABC transporter in Cry1Ab resistance was confirmed by converting a resistant silkworm (Bombyx mori) strain into a susceptible one by using germline transformation (Atsumi et al. 2012). The authors used *piggyBac* to introduce a copy of the gene from a susceptible strain (Rin; ABC gene 007792–93) into a resistant strain (w1-pnd) which made them highly susceptible to Cry1Ab toxin.

Table 3.3

Resistance ratio following CRISPR/Cas9 knockout of cadherin binding receptors

Species	Pesticide	CRISPR/Cas9 knockout line	Gene	Strain	Resistance Ratio	Reference
Helicoverpa armigera	Cry1Ac	SCD-Cad	*HaCad*	SCD	549	Wang et al. 2016
Spodoptera exigua	Cry1Ac	SeCad1-KO	*SeCad1*	WH-S	3.3	Huang et al. 2020
	Cry1Fa				3.2	
	Cry1Ca				0.4	
Trichoplusia ni	Cry1Ac	Tn-cad-2	*TnCad*	Cornell	1.7	Wang et al. 2018
	Cry1Aa				3.2	
	Cry1Ab				1.5	

Note. Knockout line shows the names given by the authors. Strain shows the background line used to create the knockout line. Resistance ratio is the fold change in resistance.

In this chapter I report the first knockin in a beetle to investigate a Cry3 protein resistance mechanism. From our literature search we identified cadherins and ABC transporters as the best candidate genes. The most compelling evidence to us was the conserved TBR in TmCad1 and TcCad1 and the divergent lysine residues hypothesized to confer tolerance in *T. castaneum* (Fabrick et al. 2009). Additionally, TmCad1-CR12 enhanced the toxicity of Cry3Aa in spotted asparagus beetle, cabbage leaf beetle, daikon leaf beetle (Gao et al. 2011) and DvvCad-CR8-10 enhanced the toxicity of Cry3Aa and Cry3Bb in WCR, southern corn rootworm, and Colorado potato beetle (Park et al. 2009). From a broader impact perspective, we are most interested in resistance in WCR, but studies investigating the genetic mechanism of resistance in WCR are difficult due to the highly repetitive and noncontiguous genome sequence, further compounded by long generation times, high-maintenance lab rearing, and a lack of gene editing tools for the species. On the other hand, there are several advantages to using red flour beetle. Tribolium is a worldwide pest of stored grains, and considered a more representative model of pest insects than the highly specialized model insect, *Drosophila melanogaster* (Brown et al. 2010). Tribolium is a robust system; it is a model for studying molecular biology and gene function in insects, easy to rear, and has a high quality assembled genome and transgenic tools.

Thus, we use CRISPR/Cas9 to replace a portion of the cadherin gene (*TcCad1*) in a Cry3Aa-tolerant beetle species, *T. castaneum*, with the orthologous region (*TmCad1*) from a Cry3Aa-susceptible species, *Te. molitor*, so we can determine the effects of Cry3Aa on both native beetles and altered *T. castaneum*. This builds on previous studies that use RNAi knockdown to confirm the role of cadherins in resistance by identifying a functional domain and confirming the location of the TBR in CR12 of TmCad and TcCad. The mechanism of resistance is hypothesized to be differences within a region of *TcCad1* and *TmCad1* that are orthologous to the Cry1A toxin-binding region in moth cadherins (Fabrick et al 2009). By replacing this portion of *TcCad1* with *TmCad1*, we hope to knock in Cry3Aa susceptibility, and therefore identify a

functional domain critical for Bt toxicity. The method may have application in the control of field pests to drive susceptibility and/or maintain susceptibility to Cry toxins into a pest population.

Methods

Insect strains

The following strains were used: *vermillion*^{white} (*v*^{W}), a white-eyed mutant strain described in Lorenzen et al (2002); Georgia2 (GA2) a wild-type strain originally collected in 1980 from Georgia, USA and described in Haliscak and Beeman (1983); and a laboratory colony of *Tenebrio molitor* for obtaining gDNA.

Homology-directed repair donor templates

The plasmids created were *p(TcCadHDRa-3xP3EGFP)* and *p(TcCadHDRb-PubEGFP)*. Total genomic DNA was extracted from whole *T. castaneum* and *Te. molitor* pupae using a DNeasyTissue kit (Qiagen, Valencia, CA). All construct components were amplified in 50 µl PCR reactions with 5 µl 10X PCR buffer, 4 µl dNTPs, 2 µl of each 25 µM primer, 0.25 µl ExTaq DNA polymerase (Takara Bio, Kusatsu, Shiga, Japan) and 15 ng gDNA template. Florescent marker sequences followed the same protocol and used 2 ng of previously built plasmids as template. Conditions followed manufacturer suggestions with an initial denaturation at 98°C for 2 min and final elongation at 72°C for 5 min. See Table 3.1 and Table 3.2 for primer sets, annealing temperatures, and elongation times. Products were purified using Monarch PCR & DNA Clean-up Kit (New England Biolabs, Ipswich, MA, USA).

To assemble *p(TcCadHDRa-3xP3EGFP)*, PCR products from Tribolium and Tenebrio were assembled by PCR-mediated recombination. The first round of this two-step PCR used purified PCR products of three fragments (*TcCad1* homology arms and *TmCad1*) as template. These products had been amplified with chimeric primers and had 19-bp overhangs which acted as primers in the following PCR (See Supplementary Material 3.1 for chimeric primer design). The first round PCR had a 50 µl PCR reaction volume with 5 µl 10X PCR buffer, 4 µl dNTPs, 0.25 µl ExTaq DNA polymerase (Takara Bio) and three templates: 10 ng purified PCR product of TmCad1

and 50 ng of each homology arm. First round conditions were 98°C for 2 min, nine cycles of 98°C for 10 sec, 60°C for 30 sec, 72°C for 30 sec, and a final elongation of 72°C for 2 min. The second round PCR had a 50 µl PCR reaction volume with 5 µl 10X PCR buffer, 4 µl dNTPs, 1 µl of 25 µM primers Chimeric Cad F1 and Chimeric Cad R4, 0.25 µl ExTaq DNA polymerase (Takara Bio) and a 1:10 dilution of round one product. Second round conditions were 98°C for 2 min, 34 cycles of 98°C for 10 sec, 60°C for 30 sec, 72°C for 2 min, and a final elongation of 72°C for 2 min. Second round product was purified with the Monarch Gel Extraction Kit (New England Biolabs) and stored at -20°C.

To insert the marker gene, we used an EcoRI in an intron of *TcCad1*. To remove the existing EcoRI sites from the desired pGEM-T Easy cloning vector (Promega, Durham, NC, USA), we digested with NotI (New England Biolabs) and purified it with the Monarch Gel Extraction Kit (New England Biolabs). Blunting of the digested pGEM-T Easy cloning vector was then done with T4 DNA polymerase (New England Biolabs) and dephosphorylated with Shrimp Alkaline Phosphatase (rSAP) (New England Biolabs). PCR product (without A overhangs) was then inserted into the blunted vector by blunt-end ligation using T4 DNA ligase (New England Biolabs) for 2 hours at room temperature. The plasmid was purified using QIAprep spin mini prep kit (Qiagen, Valencia, CA, USA) then digested with EcoRI (New England Biolabs) and purified using Monarch PCR & DNA Clean-up Kit (New England Biolabs).

The TA cloned florescent marker *pGEM(3xP3-EGFP)* was digested with EcoRI (New England Biolabs) and purified with the Monarch Gel Extraction Kit (New England Biolabs). The fluorescent marker gene was then ligated into the homology construct using T4 DNA ligase (New England Biolabs) for 2 hours at room temperature. The final plasmid was purified using a Plasmid DNA Midi Kit (Qiagen) and eluted into microinjection grade water.

Table 3.4

Primers for building p(TcCadHDRa-3xP3EGFP)

Name	Sequence	Anneal (°C)	Extension (sec)
Chimeric Cad 1F	TGATTTCTTTGACCAGGATAA	60	60
Chimeric Cad 2F	ACTCTTCAGGTCTATCAAACCTGGACATAAAATCTGCT	60	
Chimeric Cad 2R	AGCAGATTTTATGTCCAGGTTTGATAGACCTGAAGAGT	60	60
Chimeric Cad 3F	TCAAAGACAGCATGGAGGGCTATTTTACACTGAACCTC	60	
Chimeric Cad 3R	GAGGTTCAGTGTAAAATAGCCCTCCATGCTGTCTTTGA	60	60
Chimeric Cad 4R	TTTTGTATCCGTTGGTTGAGA	60	
3xP3 F	AGCTCGCCCGGGGATCTAATTCA	56	80
SV40 (EGFP) R	ATCGATACATTGATGAGTTT	56	

Note. See Supplementary Material 3.1 for chimeric primer design for PCR-mediated recombination

To assemble *p(TcCad1HDRb-PubEGFP)*, Tenebrio and Tribolium PCR products were each TA cloned into the pGEM-T Easy vector (Promega, Durham, NC, USA) with an overnight ligation at 4°C and purified using QIAprep spin mini prep kit (Qiagen, Valencia, CA, USA). Each vector was restriction digested with sticky-end cutters then sequentially ligated together. The right homology arm vector became the backbone for the rest of the construct and was digested with SphI and BsmBI (New England Biolabs) then purified using Monarch PCR & DNA Clean-up Kit (New England Biolabs). BsmBI cleaves outside of its recognition sequence and in this case created a BsaI overhang. The left homology arm vector and *TmCad1* vector were both digested with BsaI and SphI then purified using the QIAprep Gel Extraction Kit (Qiagen). The *TmCad1* fragment was ligated into the right homology arm vector using T4 DNA ligase (New England Biolabs) for 2 hours at room temperature. This ligation introduced a new BsmBI site on the 5' side of *TmCad1*. The *TmCad1* and right homology arm vector was purified using the QIAprep spin mini prep kit (Qiagen), digested with SphI and BsmBI (New England Biolabs), and then purified using Monarch PCR & DNA Clean-up Kit (New England Biolabs). This digest created a new BsaI overhang. Finally, the left homology arm fragment was ligated into the vector and purified using QIAprep spin mini prep kit (Qiagen).

An AscI restriction site was added to intronic sequence using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). The TA cloned florescent marker *p(TcPub-EGFP)* was digested with AscI (New England Biolabs) and purified with the Monarch Gel Extraction Kit (New England Biolabs). The fluorescent marker gene was then ligated into the homology construct using T4 DNA ligase (New England Biolabs) for 2 hours at room temperature. The final plasmid was purified using a Plasmid DNA Midi Kit (Qiagen) and eluted into microinjection grade water.

Table 3.5

Primers for building p(TcCadHDRb-PubEGFP)

Name	Sequence	Anneal (°C)	Extension (sec)
TcCad 5F	ACCGATAAAGATGTTGGAGAC	50	80
TcCad 6R (BsaI)	TCGGTCTCTCAGAGTGGCTTCCTTCTGCGAGTC	50	
TmCad 6F (BsmBI)	CTCGTCTCTTCTGAGCCTATTAGGAGACAATTT	55	60
TmCad 7R (BsaI)	TCGGTCTCTGGATGGTACAATTCACTTCGTAAG	55	
TcCad 7F (BsmBI)	CTCGTCTCTATCCAAATTGATTTCTTTGACCAG	50	60
TcCad 8R	AGGTTCAGTGTAAAATAGCCA	50	
Q5SDM_F	CGCGCCATAACTTCCATACTAATTTTTTAC	56	165
Q5SDM_R	CCAAGTAGTTCTTCTTTTG	56	
AscI-TcPub F	TAGGCGCGCCTATTGTCGTCCGTATTTACA	56	90
EGFP(SV40)-AscI R	GGCGCGCCATCGATACATTGATGAGTTT	56	

Table 3.6

Primers for sequencing and bacterial colony PCR

Name	Sequence	Anneal (°C)
TcCad1 F1	TCAGGGCGAGGATAAGGATGC	55
TcCad1 R1	TGGTCTCTTCGTTGGCATTGG	55-59
TcCad1 F2	ACGAAAGTCTGAAAAATTGCG	55-59
TcCad1 F3	GATGTGAACGATGAACTGCC	55
TcCad1 R3	TACGGTGTGTTTGGAAGTCA	55
TcCad F3	ACCCAAAACAAAGACTCATCG	55
TcCad R3	TGTTTTGGGTCTGGAGTGTG	55-57
TcCad F10	GACATCTCGCCAAAGCTAGG	57
TcCad R10	CCCTCATGGTCGTTGAAAAC	57
TcCad F11	ATGCCAACGAAGAGACCAAG	57
TcCad R11	TCAGGGACTGGAATTTGTCC	57
TcCad R8	ACGAGGCATTTTCGACTTTG	54
TmCad F8	TGTCCACATTGGTTCTTCCA	54
TcCad R4	TAGATTTGAGCGGTGGCTCT	55

CRISPR/Cas9 single guide RNAs (sgRNAs)

Nine *TcCad1* guides (Table 3.7) were designed based on off- and on-target scores using the Benchling CRISPR Guide RNA design tool (available at https://www.benchling.com/crispr/). Guides 1-3 were designed to target the middle of the 84bp region to be substituted by HDRa. Guides 4-6 were designed to target the outer flanks of the 680bp region to be substituted by HDRb. Synthetic sgRNAs were ordered (Synthego, Redwood City, CA, USA) and stored at -80°C. See Supplementary Material 3.5, 3.6, and 3.7 for plasmid sequences and feature maps.

Table 3.7

CRISPR/Cas9 guide RNAs for targeting TcCad1

Name	Sequence
TcCad1 Guide1	AAAUUGAGGACCCGAAAAG
TcCad1 Guide2	ACUUGAUGACACCGCUUUU
TcCad1 Guide3	GCUUUUCGGGUCCUCAAUU
TcCad1 Guide4	GUCGUGGUUCAAUACGAGAG
TcCad1 Guide5	AAUUGUCCCGCAAAUCGUCG
TcCad1 Guide6	UAUUGAACCACGACGAUUUG
TcCad1 Guide7	UAAAACUACUCGUGCUCAUC
TcCad1 Guide8	GAUGAGCACGAGUAGUUUUA
TcCad1 Guide9	UUUACGGCAAAACCUUACA

Microinjection

Transgenic *T. castaneum* strains with modified *TcCad1* (XP_971388.2) were established by CRISPR/Cas9-mediated knock-in and components delivered into preblastodermal eggs via microinjection. Two hour old eggs were washed with 2.5% bleach, aligned on a glass cover slip, then injected with ribonucleoprotein (RNP) complex. RNP complex was created by incubating injection mix on ice for 30 minutes before injections.

Injection mix for HDRa was 0.5 µl Cas9 protein (Fisher TrueCut Cat#A36498), 2.1 µl sgRNAs (mix of guides 1-3; final concentration of 400 ng/µl each), 0.4 µl 10X injection buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0), and 1.0 µl *p(TcCadHDRa-3xP3EGFP)* at a final concentration of 500 ng/µl. Injection mix for HDRb was 0.5 µl Cas9 protein (Fisher TrueCut Cat#A36498), 2.1 µl sgRNAs (mix of guides 4,6,7,8 or 5,6,8,9; final concentration of 400 ng/µl each), 0.4 µl 10X injection buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0), and 1.0 µl *p(TcCadHDRb-PubEGFP)* at a final concentration of 500 ng/µl. HDR constructs were injected as circular dsDNA donor plasmids as suggested by Gratz et al. 2014 and evidenced as more effective than linearized plasmids in Beumer et al. 2008. Injection mix was stored on ice within a refrigerator for up to 48 hours, otherwise it was made fresh.

Florescent marker screening

The fluorescent protein genes used to mark the homology constructs are driven by different promoters and were screened for at different insect life stages. Injected G_0 individuals from HDRa were segregated by sex at the pupal stage and individually mated (single-pair crosses) with a v^W mate. The G_1 offspring were screened at the pupal stage for green eyes from *3xP3-EGFP* expression. Injected G_0 individuals from HDRb were screened at the pupal stage for green bodies from *Pub-EGFP* expression.

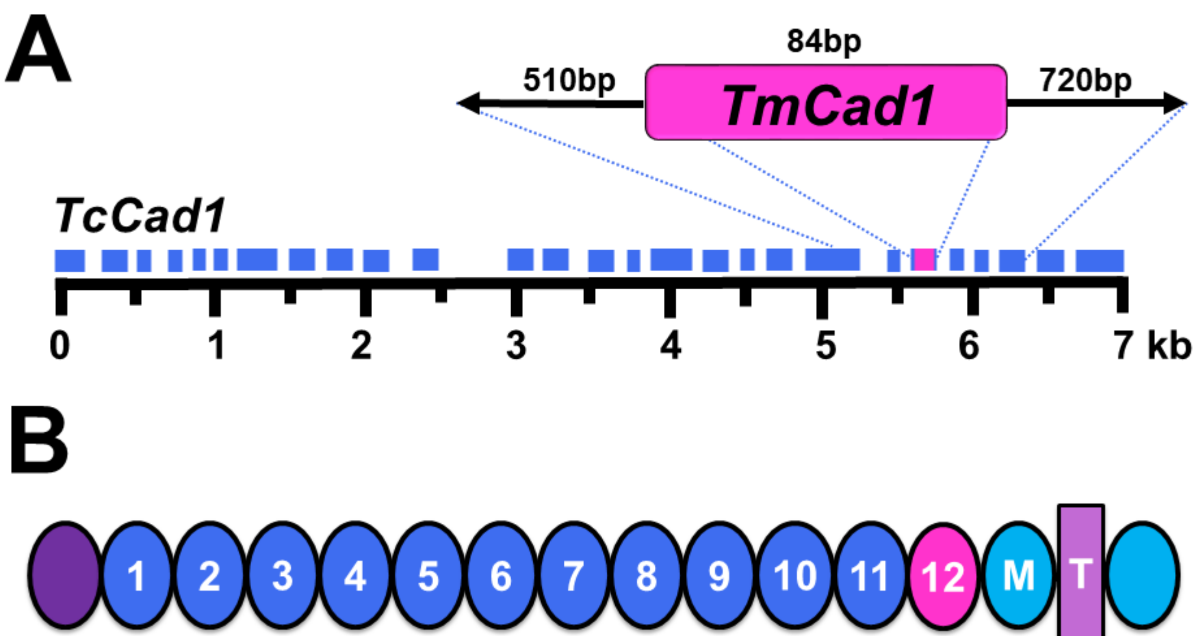
Results

The first homology construct (HDRa) replaces 84bp of *TcCad1* with the corresponding sequence from *TmCad1*. This substitution is within CR12 and replaces 17 amino acids which results in a shared identity of 1609/1626 (99.0%) (Figure 3.2). The second homology construct (HDRb) replaces 680bp of *TcCad1* with the corresponding sequence from *TmCad1*. This substitution is within CR8-10 and replaces 81 amino acids which results in a shared identity of 1545/1626 (95.0%) (Figure 3.3). The change to the modified TcCad is even less when similarity between amino acids is considered. Pairwise Sequence Alignment using EMBOSS Needle shows an amino acid sequence similarity of 1615/1626 (99.3%) for HDRa and 1575/1626 (96.9%) for

HDRb. See Supplementary Material 3.2 for an amino acid alignment of all three sequences. For both constructs we placed a marker gene within an intron. HDRa included *3xP3-EGFP* in the intron immediately upstream from the single exon being substsituted with *TmCad1* (Supplementary Material 3.3). HDRb included *Pub-EGFP* in the third intron of the four exons being substituted with *TmCad1* (Supplementary Material 3.4).

Figure 3.2

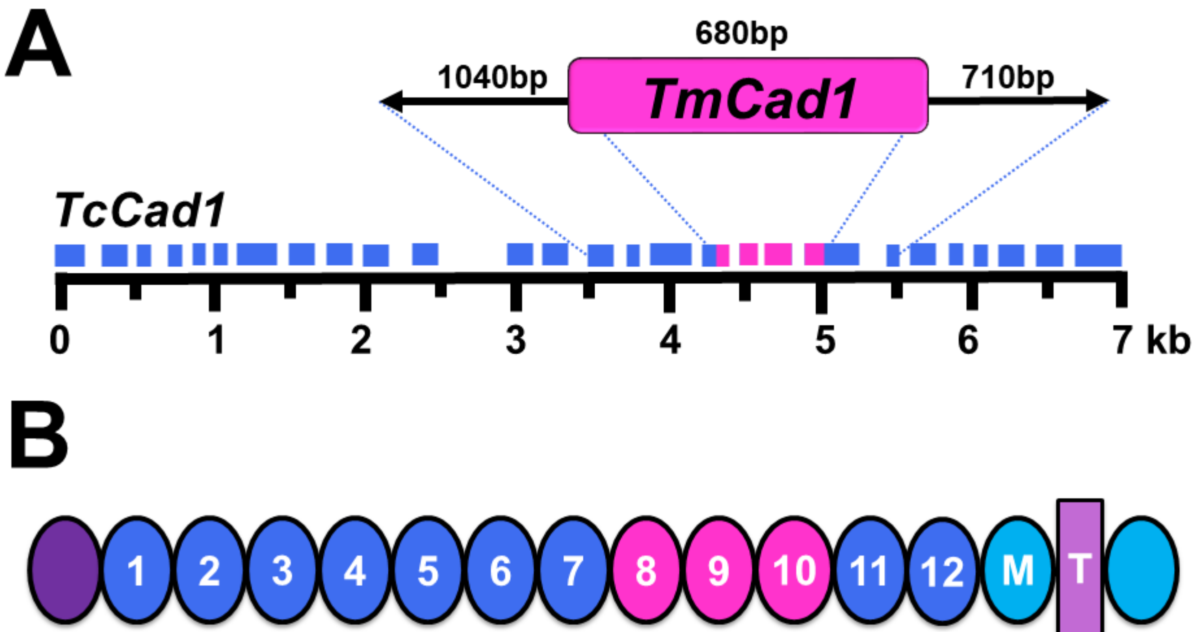
CRISPR/Cas9-mediated substitution of CR12 region of TcCad1 gene with TmCad1



Note. A) *TcCad1* gene with exons (blue) and portion of exon to be substituted with *TmCad1* (pink). Homology directed repair construct (HDRa) with *TmCad1* toxin binding region (pink) and homology arms (black arrows). B) Domain structure of TcCad1 with cadherin repeats (CR 1-12), membrane proximal region (M) and transmembrane domain (T). HDRa substitutes amino acids within CR12 (pink).

Figure 3.3

CRISPR/Cas9-mediated substitution of CR8-10 region of TcCad1 gene with TmCad1



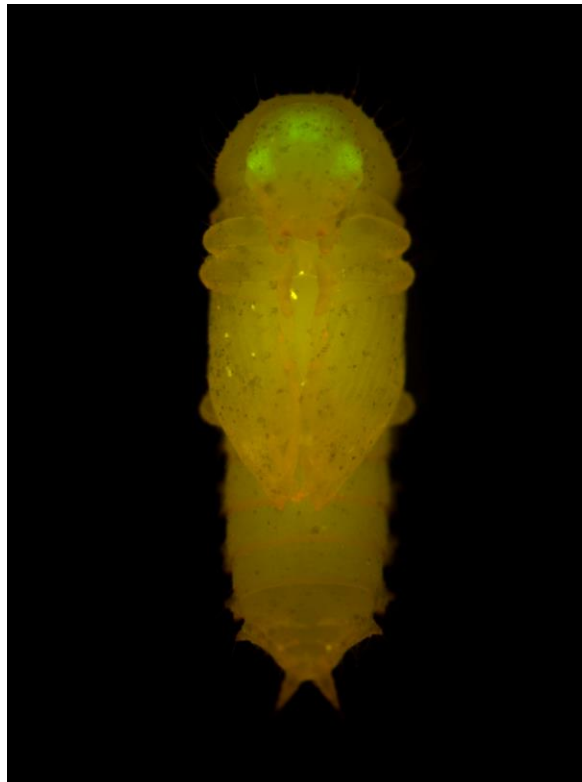
Note. A) *TcCad1* gene with exons (blue) and portion of exon to be substituted with *TmCad1* (pink). Homology directed repair construct (HDRb) with *TmCad1* toxin binding region (pink) and homology arms (black arrows). B) Domain structure of TcCad1 with cadherin repeats (CR 1-12), membrane proximal region (M) and transmembrane domain (T). HDRb substitutes amino acids within CR8-10 (pink).

Five individuals from HDRa injections were found to express the fluorescent marker (Table 3.8, Figure 3.4), of which three were male and two were female. Each pupae was allowed to eclose before setting up mating vials. One male died two days after eclosing, but before it had been given a mate. The remaining four individuals were set up in two single mate-pair crosses (vial one with Male1 X Female1 and vial two with Male2 X Female2). After ten days, males were extracted from the vials and put into two new vials, each with two female v^W mates (Male1 X v^W and Male1 X v^W). Seamless substitution of the *TcCad1* gene with the HDRa construct was

confirmed by Sanger sequencing off of PCR products from 20 G_1 offspring from all four mating vials (Figure 3.5). Additional G_1 offspring were used to create homozygous lines that will be tested later to compare mortality of edited and unedited beetles when reared on Cry3Aa.

Figure 3.4

Fluorescent marker gene expression in G_1 offspring of HDRa-injected parents



Note. Pupa showing EGFP fluorescence in eyes from marker gene *3xP3-EGFP*. There is additional expression in the head vertex.

Table 3.8

CRISPR/Cas9-mediated homology directed repair results

HDR Plasmid	Injected	Hatched	Survival Rate	EGFP G_0 Pupae	Adults	Fertile Adults	Crosses w/ EGFP	EGFP G_1 Pupae	Transformation Rate
p(TcCadHDRa-3xP3EGFP)	760	187	24.6%	n/a	116	109	1	5	0.9%
p(TcCadHDRb-PubEGFP)	963	310	32.2%	0	n/a	n/a	n/a	n/a	n/a

Note. Transformation rate was calculated as the number of fertile crosses that produced transgenic offspring.

Figure 3.5

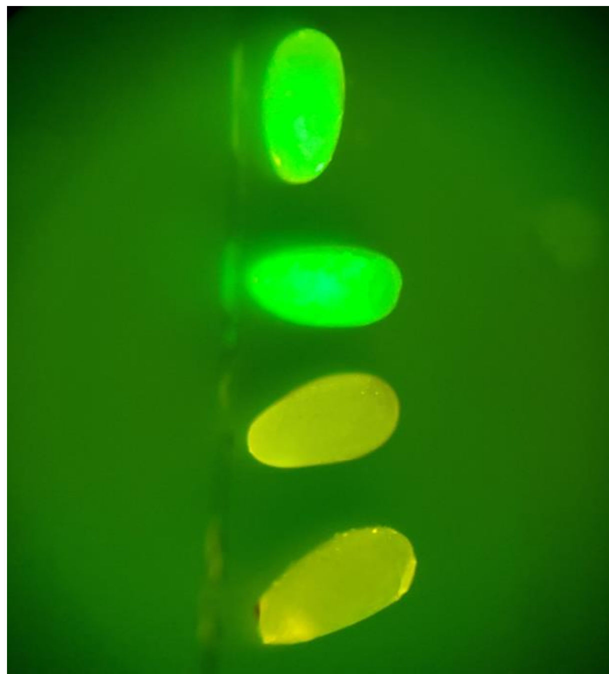
Sanger sequencing to confirm seamless substitution

Note. SnapGene screenshots showing alignment of sequenced PCR products from TcCad1-HDRa individuals. Both Male1 and Female1 had the correct substitution. There are two questionable base calls in the *3xP3* promoter, but the marker was functional so they were not an issue.

None of the G_0 pupae from HDRb injections expressed the fluorescent marker (Table 3.8). To test if this was from the CRISPR/Cas9-HDR failing or from a problem with our construct, we did test injections and screened for transient expression of EGFP from the plasmid. We injected GA2 with *p(TcCadHDRb-PubEGFP)* and as a positive control used another plasmid with *Pub-EGFP* from Chapter 4 *pB(TcPubEGFP-ATPsynthCDUF1703)*. After 24 hours, eggs injected with *pB(TcPubEGFP-ATPsynthCDUF1703)* were fluorescing while the p(TcCadHDRb-PubEGFP) injected eggs were not (Figure 3.6).

Figure 3.6

Transient EGFP from injected plasmids



Note. Top two eggs were injected with *pB(TcPubEGFP-ATPsynthCDUF1703)* and the bottom two eggs with *p(TcCadHDRb-PubEGFP)*.

Next the marker gene portion of *p(TcCadHDRb-PubEGFP)* was assessed by sequencing. Previously, we found no mutations in the plasmid *p(PubEGFP)* which was digested and used to

generate *p(TcCadHDRb-PubEGFP)*. The junction sites of *PubEGFP* within the injection-ready plasmid *p(TcCadHDRb-PubEGFP)* were also confirmed before injections. However, the CDS of *EGFP* within the final injection plasmid *p(TcCadHDRb-PubEGFP)* had not been sequenced. Sequencing revealed two missense mutations which caused S73T and K108N. These mutations were not present in *p(PubEGFP)*, so must have been produced by *E. coli* during the final cloning step and not during PCR to produce fragments for cloning. Serine (S) and Threonine (T) are both polar amino acids and considered similar in amino acid alignments. Lysine (K) is basic and Asparagine (N) is polar. There is some variation in naturally occurring *GFP* genes at position 108 (K/C) and position 73 (S/V/A). However, the residues responsible for green fluorescence, called the chromophore, was not disrupted. The EGFP chromophore consists of residues T65, Y66, and G67 (GFP is SYG).

There were also a couple of point mutations in the promoter sequence. Analysis using the Neural Network Promoter Prediction tool (Available at https://www.fruitfly.org/seq_tools/promoter.html) showed one fewer predicted promoter sequences, though this was the predicted promoter sequence furthest from the start codon and did not disrupt the TATA box (Figure 3.7).

During the cloning process for building both plasmids, plasmid DNA was purified from multiple bacterial colonies and sequenced. There were often multiple point mutations that differed between colonies (miniprep samples) (Supplementary Material 3.8). Therefore, if these sequences are cloned again to obtain an error free HDRb plasmid, I suggest trying a different competent cell line (I used TOP10).

Figure 3.7

Predicted promoter sequences within the promoter TcPub

Desired HDRb plasmid sequence:

Start	End	Score	Promoter Sequence
182	232	0.99	atttgaaaaataaaaaagggcgtaattccaaaacaaagaaaacagctgat
243	293	0.93	acgacgtagataaaaatcagctgacatgtcaccaacaaatactggccaag
466	516	0.99	ggttcgagagttaaaaaggagtgggggaagccgtccgtataaaagcggtt
493	543	0.98	aagccgtccgtataaaagcggttccacagctgcattcgccatattgtctg

HDRb plasmid sequence with mutations:

Start	End	Score	Promoter Sequence
242	292	0.94	gacgacgtagataaaaatcagctgacatgtcaccaacaactactggcaca
466	516	1.00	gggttcgagattaaaaaggagtgggggaagccgtccgtataaaagcggtt
493	543	0.98	aagccgtccgtataaaagcggttccacagctgcattcgccatattgtctg

Note. TATA box highlighted in yellow.

Discussion

Our hypothesis for the Bt bioassay is that the TcCad1-CR12 edited strain will have a significantly higher mortality rate than an un-edited strain of the same genetic background. We will need multiple controls to confirm that death is due to the altered cadherin, and not because the new strain is somehow weaker or has differences in cadherin expression. Fitness and expression levels of *TcCad1* will need to be comparable to the parental strain. While we expect them to be equivalent, the knockin contains an intronic marker (Supplementary Material 3.3) that could possibly impact gene expression. Once we have confirmed equivelant *TcCad1* expression levels we can be confident in our assessment of the effect of TcCad1-CR12 on Cry3Aa toxicity.

However, if mortality rates are not significantly different, then either a) the TBR is located in another CR (possibly CR8-10), b) TcCad1 plays an enhancer role in susceptibility and other genes would also need to be altered (possibly *TcABC*) in order to see an effect, or C) TcCad1

does not play a role in resistance. The review by Jurat-Fuentes et al. (2021) on mechanisms of Bt-resistance states that ABC transporters are possibly more critical for high toxicity, while interaction with the other receptors (APN, ALP, Cad) may increase toxicity and thus play an enhancer role. Many of the CRISPR/Cas9 knockout studies in Lepidoptera have knocked out *ABCC and ABCA* genes, and they are also common targets for RNAi. In Tribolium, RNAi knockdown targeting members of the gene subfamily *ABCC* was not lethal, unlike many other subfamilies (Broehan et al. 2013). Therefore, *TcABCC* would also make a good target for a similar CRISPR/Cas9 knock-in assay. The hypothesis that ABC transporters are critical and cadherins are enhancers could be tested by creating an analogous CRISPR/Cas9 *TmABCC* knockin in our TcCad1-CR12 edited strain and in a wild-type strain. If both are responsible for Cry3Aa-tolerance in Tribolium and cadherin acts as an enhancer, then unedited Tribolium should have the highest Cry3Aa LC50 followed by *TcCad1*-edited, then *TcABCC*-edited, and the two gene edited being the lowest.

Conclusion

We used CRISPR/Cas9-mediated gene editing to replace a small portion of the *TcCad1* gene with the equivalent portion from a related Cry3Aa-susceptible species, *Tenebrio molitor*. We were successful in making a strain of Tribolium with an edited TcCad CR12. However, we were unable to create a strain with an edited TcCad CR8-10 due to an issue with the marker gene. Next, the TcCad CR12 edited line will be sent to a collaborator at the USDA-ARS for bioassays on Cry3Aa. Our goal is to validate the function of a proposed Cry3Aa toxin binding region in a cadherin receptor protein by converting a tolerant beetle (*Tribolium castaneum*) into a susceptible one.

This is the first gene editing of a Cry3 toxin binding receptor. Previous studies have focused on Cry1 receptors in Lepidoptera and most knocked out a Cry1 toxin binding receptor while a few created single amino acid substitutions. These studies all measured resistance ratios following gene editing with CRISPR/Cas9. Besides our current work, only one other study used

germline transformation to increase susceptibility. Atsumi et al. (2012) used *piggyBac* to introduce a copy of a *BmABC* gene from a susceptible silkworm strain into a resistant strain, which made them highly susceptible to Cry1Ab toxin.

REFERENCES

Arbuckle, J.R. (2014). Farmer Perspective and Pesticide Resistance. Ames: Iowa State University Extension and Outreach. Available online at http://www.soc.iastate.edu/extension/ifrlp/PDF/PM3070.pdf. Accessed February 22, 2016.

Atsumi, S., Miyamoto, K., Yamamoto, K., Narukawa, J., Kawai, S., Sezutsu, H., Kobayashi, I., Uchino, K., Tamura, T., Mita, K. and Kadono-Okuda, K., Wadaa, S., Kandad,K., Goldsmithe, M. R., & Noda, H. (2012). Single amino acid mutation in an ATP-binding cassette transporter gene causes resistance to Bt toxin Cry1Ab in the silkworm, Bombyx mori. Proceedings of the National Academy of Sciences, 109(25), E1591-E1598.

Benbrook, C. M. (2012). Impacts of genetically engineered crops on pesticide use in the US--the first sixteen years. *Environmental Sciences Europe*, *24*(1), 1-13.

Beumer, K. J., Trautman, J. K., Bozas, A., Liu, J. L., Rutter, J., Gall, J. G., & Carroll, D. (2008). Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. *Proceedings of the National Academy of Sciences*, *105*(50), 19821-19826.

Bravo, A., Pacheco, S., Gómez, I., Garcia-Gómez, B., Onofre, J., & Soberón, M. (2017). Insecticidal proteins from Bacillus thuringiensis and their mechanism of action. In *Bacillus thuringiensis and Lysinibacillus sphaericus* (pp. 53-66). Springer, Cham.

Broehan, G., Kroeger, T., Lorenzen, M., & Merzendorfer, H. (2013). Functional analysis of the ATP-binding cassette (ABC) transporter gene family of Tribolium castaneum. *BMC genomics*, *14*(1), 1-19.

Brown, S. J., Shippy, T. D., Miller, S., Bolognesi, R., Beeman, R. W., Lorenzen, M. D., Bucher, G., Wimmer, E.A. & Klingler, M. (2009). The red flour beetle, Tribolium castaneum (Coleoptera): a model for studies of development and pest biology. *Cold Spring Harbor Protocols*, *2009*(8), pdb-emo126.

Contreras, E., Schoppmeier, M., Real, M. D., & Rausell, C. (2013). Sodium solute symporter and cadherin proteins act as Bacillus thuringiensis Cry3Ba toxin functional receptors in Tribolium castaneum. *Journal of Biological Chemistry*, *288*(25), 18013-18021.

Douris, V., Denecke, S., Van Leeuwen, T., Bass, C., Nauen, R., & Vontas, J. (2020). Using CRISPR/Cas9 genome modification to understand the genetic basis of insecticide resistance: Drosophila and beyond. *Pesticide Biochemistry and Physiology*, 104595.

EPA (Environmental Protection Agency) The Environmental Protection Agency's White Paper on Bt Plant-Pesticide Resistance Management. Washington, DC (1998)

Fabrick, J., Oppert, C., Lorenzen, M. D., Morris, K., Oppert, B., & Jurat-Fuentes, J. L. (2009). A novel Tenebrio molitor cadherin is a functional receptor for Bacillus thuringiensis Cry3Aa toxin. *Journal of Biological Chemistry*, *284*(27), 18401-18410.

Fernandez-Cornejo, J., Wechsler, S., Livingston, M., & Mitchell, L. (2014). Genetically engineered crops in the United States. *USDA-ERS Economic Research Report*, (162).

Ferré, J., & Van Rie, J. (2002). Biochemistry and genetics of insect resistance to Bacillus thuringiensis. *Annual review of entomology*, *47*(1), 501-533.

Flagel, L. E., Swarup, S., Chen, M., Bauer, C., Wanjugi, H., Carroll, M., ... & Goldman, B. S. (2015). Genetic markers for western corn rootworm resistance to Bt toxin. G3: Genes, Genomes, Genetics, 5(3), 399-405.

Gahan, L. J., Gould, F., & Heckel, D. G. (2001). Identification of a gene associated with Bt resistance in Heliothis virescens. *Science*, *293*(5531), 857-860.

Gao, Y., Jurat-Fuentes, J. L., Oppert, B., Fabrick, J. A., Liu, C., Gao, J., & Lei, Z. (2011). Increased toxicity of Bacillus thuringiensis Cry3Aa against Crioceris quatuordecimpunctata, Phaedon brassicae and Colaphellus bowringi by a Tenebrio molitor cadherin fragment. *Pest management science*, *67*(9), 1076-1081.

Gardner, J. G., Nehring, R. F., & Nelson, C. H. (2009). Genetically modified crops and household labor savings in US crop production.

Gassmann, A. J., Petzold-Maxwell, J. L., Keweshan, R. S., & Dunbar, M. W. (2011). Field-evolved resistance to Bt maize by western corn rootworm. *PloS one*, *6*(7), e22629.

Gassmann, A. J., Petzold-Maxwell, J. L., Clifton, E. H., Dunbar, M. W., Hoffmann, A. M., Ingber, D. A., & Keweshan, R. S. (2014). Field-evolved resistance by western corn rootworm to multiple Bacillus thuringiensis toxins in transgenic maize. *Proceedings of the National Academy of Sciences*, *111*(14), 5141-5146.

Gassmann, A.J., Shrestha, R.B., Jakka, S.R., Dunbar, M.W., Clifton, E.H., Paolino, A.R., Ingber, D.A., French, B.W., Masloski, K.E., Dounda, J.W. and St. Clair, C.R.. (2016). Evidence of resistance to Cry34/35Ab1 corn by western corn rootworm (Coleoptera: Chrysomelidae): root injury in the field and larval survival in plant-based bioassays. *Journal of economic entomology*, *109*(4), pp.1872-1880.

Gassmann, A. J., Shrestha, R. B., Kropf, A. L., St Clair, C. R., & Brenizer, B. D. (2020). Field-evolved resistance by western corn rootworm to Cry34/35Ab1 and other Bacillus thuringiensis traits in transgenic maize. *Pest management science*, *76*(1), 268-276.

Gassmann, A. J. (2021). Resistance to Bt Maize by Western Corn Rootworm: Effects of Pest Biology, the Pest–Crop Interaction and the Agricultural Landscape on Resistance. *Insects*, *12*(2), 136.

Gray, M. E., Sappington, T. W., Miller, N. J., Moeser, J., & Bohn, M. O. (2009). Adaptation and invasiveness of western corn rootworm: intensifying research on a worsening pest. *Annual review of entomology*, *54*, 303-321.

Guo, Z., Sun, D., Kang, S., Zhou, J., Gong, L., Qin, J., ... & Zhang, Y. (2019). CRISPR/Cas9-mediated knockout of both the PxABCC2 and PxABCC3 genes confers high-level resistance to Bacillus thuringiensis Cry1Ac toxin in the diamondback moth, Plutella xylostella (L.). *Insect biochemistry and molecular biology*, *107*, 31-38.

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., & O'Connor-Giles, K. M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics*, *196*(4), 961-971.

Guest, M., Goodchild, J. A., Bristow, J. A., & Flemming, A. J. (2019). RDL A301S alone does not confer high levels of resistance to cyclodiene organochlorine or phenyl pyrazole insecticides in Plutella xylostella. *Pesticide biochemistry and physiology*, *158*, 32-39.

Haliscak, J. P., & Beeman, R. W. (1983). Status of malathion resistance in five genera of beetles infesting farm-stored corn, wheat, and oats in the United States. *Journal of Economic Entomology*, *76*(4), 717-722.

Herrero, S., Oppert, B., & Ferré, J. (2001). Different Mechanisms of Resistance toBacillus thuringiensis Toxins in the Indianmeal Moth. *Applied and environmental microbiology*, *67*(3), 1085-1089.

Hua, G., Park, Y., & Adang, M. J. (2014). Cadherin AdCad1 in Alphitobius diaperinus larvae is a receptor of Cry3Bb toxin from Bacillus thuringiensis. Insect biochemistry and molecular biology, 45, 11-17.

Huang, J., Xu, Y., Zuo, Y., Yang, Y., Tabashnik, B. E., & Wu, Y. (2020). Evaluation of five candidate receptors for three Bt toxins in the beet armyworm using CRISPR-mediated gene knockouts. *Insect biochemistry and molecular biology*, *121*, 103361.

Jakka, S. R., Shrestha, R. B., & Gassmann, A. J. (2016). Broad-spectrum resistance to Bacillus thuringiensis toxins by western corn rootworm (Diabrotica virgifera virgifera). *Scientific reports*, *6*(1), 1-9.

Jin, L., Wang, J., Guan, F., Zhang, J., Yu, S., Liu, S., Xue, Y., Li, L., Wu, S., Wang, X., Yang, Y., Abdelgaffar, H., Jurat-Fuentes, J. L., Tabashnik, B. E., & Wu, Y. (2018). Dominant point mutation in a tetraspanin gene associated with field-evolved resistance of cotton bollworm to transgenic Bt cotton. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(46), 11760-11765. https://doi.org/10.1073/pnas.1812138115

Jurat-Fuentes, J. L., Heckel, D. G., & Ferré, J. (2021). Mechanisms of Resistance to Insecticidal Proteins from Bacillus thuringiensis. *Annual Review of Entomology*, *66*, 121-140.

Li, H., Olson, M., Lin, G., Hey, T., Tan, S. Y., & Narva, K. E. (2013). Bacillus thuringiensis Cry34Ab1/Cry35Ab1 interactions with western corn rootworm midgut membrane binding sites. *PLoS One*, *8*(1), e53079.

Loseva, O., Ibrahim, M., Candas, M., Koller, C. N., Bauer, L. S., & Bulla Jr, L. A. (2002). Changes in protease activity and Cry3Aa toxin binding in the Colorado potato beetle: implications for insect resistance to Bacillus thuringiensis toxins. *Insect Biochemistry and Molecular Biology*, *32*(5), 567-577.

Metcalf, R. L. (1986). Foreword. In J. L. Krysan & T. A. Miller (Eds.), Methods for the study of pest Diabrotica (pp. vii–xv). New York, NY: Springer-Verlag

Morin, S., Biggs, R. W., Sisterson, M. S., Shriver, L., Ellers-Kirk, C., Higginson, D., ... & Tabashnik, B. E. (2003). Three cadherin alleles associated with resistance to Bacillus thuringiensis in pink bollworm. *Proceedings of the National Academy of Sciences*, *100*(9), 5004-5009.

Nagamatsu, Y., Toda, S., Yamaguchi, F., Ogo, M., Kogure, M., Nakamura, M., ... & Katsumoto, T. (1998). Identification of Bombyx mori midgut receptor for Bacillus thuringiensis insecticidal CryIA (a) toxin. *Bioscience, biotechnology, and biochemistry*, *62*(4), 718-726.

National Academies of Sciences, Engineering, and Medicine. (2016). *Genetically engineered crops: experiences and prospects*. National Academies Press.

Niu, X., Kassa, A., Hasler, J., Griffin, S., Perez-Ortega, C., Procyk, L., Zhang, J., Kapka-Kitzman, D.M., Nelson, M.E. & Lu, A. (2020). Functional validation of DvABCB1 as a receptor of Cry3 toxins in western corn rootworm, Diabrotica virgifera virgifera. *Scientific reports*, *10*(1), 1-13.

NRC (National Research Council). (2010). The Impact of Genetically Engineered Crops on Farm Sustainability in the United States. Washington, DC: National Academies Press.

Ochoa-Campuzano, C., Martínez-Ramírez, A. C., Contreras, E., Rausell, C., & Real, M. D. (2013). Prohibitin, an essential protein for Colorado potato beetle larval viability, is relevant to Bacillus thuringiensis Cry3Aa toxicity. Pesticide biochemistry and physiology, 107(3), 299-308

Oppert, B., Ellis, R. T., & Babcock, J. (2010). Effects of Cry1F and Cry34Ab1/35Ab1 on storage pests. *Journal of Stored Products Research*, *46*(3), 143-148.

Oppert, B., Morgan, T. D., & Kramer, K. J. (2011). Efficacy of Bacillus thuringiensis Cry3Aa protoxin and protease inhibitors against coleopteran storage pests. *Pest management science*, *67*(5), 568-573.

Oppert, B., Dowd, S. E., Bouffard, P., Li, L., Conesa, A., Lorenzen, M. D., Toutges, M., Marshall, J., Huestis, D.L., Fabrick, J., Oppert, C., & Jurat-Fuentes, J. L. (2012). Transcriptome profiling of the intoxication response of Tenebrio molitor larvae to Bacillus thuringiensis Cry3Aa protoxin. *PloS one*, *7*(4), e34624.

Park, Y., Abdullah, M. A. F., Taylor, M. D., Rahman, K., & Adang, M. J. (2009). Enhancement of Bacillus thuringiensis Cry3Aa and Cry3Bb toxicities to coleopteran larvae by a toxin-binding fragment of an insect cadherin. *Applied and Environmental Microbiology*, *75*(10), 3086-3092.

Park, Y., Hua, G., Taylor, M. D., & Adang, M. J. (2014). A coleopteran cadherin fragment synergizes toxicity of Bacillus thuringiensis toxins Cry3Aa, Cry3Bb, and Cry8Ca against lesser mealworm, Alphitobius diaperinus (Coleoptera: Tenebrionidae). *Journal of invertebrate pathology*, *123*, 1-5.

Pauchet, Y., Bretschneider, A., Augustin, S., & Heckel, D. G. (2016). A P-glycoprotein is linked to resistance to the Bacillus thuringiensis Cry3Aa toxin in a leaf beetle. *Toxins*, *8*(12), 362.

Rault, L. C., Siegfried, B. D., Gassmann, A. J., Wang, H., Brewer, G. J., & Miller, N. J. (2018). Investigation of Cry3Bb1 resistance and intoxication in western corn rootworm by RNA sequencing. *Journal of Applied Entomology*, *142*(10), 921-936.

Rausell, C., Ochoa-Campuzano, C., Martínez-Ramírez, A. C., Bravo, A., & Real, M. D. (2007). A membrane associated metalloprotease cleaves Cry3Aa Bacillus thuringiensis toxin reducing pore formation in Colorado potato beetle brush border membrane vesicles. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1768*(9), 2293-2299.

Sayed, A., Nekl, E. R., Siqueira, H. A. A., Wang, H. C., Ffrench-Constant, R. H., Bagley, M., & Siegfried, B. D. (2007). A novel cadherin-like gene from western corn rootworm, Diabrotica virgifera virgifera (Coleoptera: Chrysomelidae), larval midgut tissue. *Insect molecular biology*, *16*(5), 591-600.

Siegfried, B. D., Waterfield, N., & Ffrench-Constant, R. H. (2005). Expressed sequence tags from Diabrotica virgifera virgifera midgut identify a coleopteran cadherin and a diversity of cathepsins. *Insect molecular biology*, *14*(2), 137-143.

Tabashnik, B. E., & Carrière, Y. (2017). Surge in insect resistance to transgenic crops and prospects for sustainability. *Nature Biotechnology*, *35*(10), 926.

Tan, S. Y., Rangasamy, M., Wang, H., Vélez, A. M., Hasler, J., McCaskill, D., ... & Siegfried, B. D. (2016). RNAi induced knockdown of a cadherin-like protein (EF531715) does not affect toxicity of Cry34/35Ab1 or Cry3Aa to Diabrotica virgifera virgifera larvae (Coleoptera: Chrysomelidae). *Insect biochemistry and molecular biology*, *75*, 117-124.

USDA ERS. (2021). “Recent Trends in GE Adoption.” *Economic Research Service U.S. Department of Agriculture*, 2021, www.ers.usda.gov/data-products/adoption-of-genetically-engineered-crops-in-the-us/recent-trends-in-ge-adoption.aspx.

Wang, J., Zhang, H., Wang, H., Zhao, S., Zuo, Y., Yang, Y., & Wu, Y. (2016). Functional validation of cadherin as a receptor of Bt toxin Cry1Ac in Helicoverpa armigera utilizing the CRISPR/Cas9 system. *Insect biochemistry and molecular biology*, *76*, 11-17.

Wang, J., Wang, H., Liu, S., Liu, L., Tay, W. T., Walsh, T. K., ... & Wu, Y. (2017). CRISPR/Cas9 mediated genome editing of Helicoverpa armigera with mutations of an ABC transporter

gene HaABCA2 confers resistance to Bacillus thuringiensis Cry2A toxins. *Insect biochemistry and molecular biology*, *87*, 147-153.

Wang, S., Kain, W., & Wang, P. (2018). Bacillus thuringiensis Cry1A toxins exert toxicity by multiple pathways in insects. *Insect biochemistry and molecular biology*, *102*, 59-66.

Wang, X., Cao, X., Jiang, D., Yang, Y., & Wu, Y. (2020). CRISPR/Cas9 mediated ryanodine receptor I4790M knockin confers unequal resistance to diamides in Plutella xylostella. *Insect Biochemistry and Molecular Biology*, *125*, 103453.

Wechsler, S., & Smith, D. (2018). Has resistance taken root in US corn fields? Demand for insect control. *American Journal of Agricultural Economics*, *100*(4), 1136-1150.

Whalon, M. E., & Wingerd, B. A. (2003). Bt: mode of action and use. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, *54*(4), 200-211.

Willse, A., Flagel, L., & Head, G. (2021). Estimation of Cry3Bb1 Resistance Allele Frequency in Field Populations of Western Corn Rootworm using a Genetic Marker. *G3 Genes| Genomes| Genetics*.

Zhao, J. Z., Oneal, M. A., Richtman, N. M., Thompson, S. D., Cowart, M. C., Nelson, M. E., ... & Yamamoto, T. (2016). mCry3A-selected western corn rootworm (Coleoptera: Chrysomelidae) colony exhibits high resistance and has reduced binding of mCry3A to midgut tissue. *Journal of economic entomology*, *109*(3), 1369-1377.

Zúñiga-Navarrete, F., Gómez, I., Peña, G., Bravo, A., & Soberón, M. (2013). A Tenebrio molitor GPI-anchored alkaline phosphatase is involved in binding of Bacillus thuringiensis Cry3Aa to brush border membrane vesicles. *Peptides*, *41*, 81-86.

Zuo, Y., Wang, H., Xu, Y., Huang, J., Wu, S., Wu, Y., & Yang, Y. (2017). CRISPR/Cas9 mediated G4946E substitution in the ryanodine receptor of Spodoptera exigua confers high levels of resistance to diamide insecticides. *Insect biochemistry and molecular biology*, *89*, 79-85.

Zuo, Y. Y., Ma, H. H., Lu, W. J., Wang, X. L., Wu, S. W., Nauen, R., Wu, Y.D., & Yang, Y. H. (2020). Identification of the ryanodine receptor mutation I4743M and its contribution to diamide insecticide resistance in Spodoptera exigua (Lepidoptera: Noctuidae). *Insect science*, *27*(4), 791-800.

SUPPLEMENTARY MATERIAL

Supplementary Material 3.1

Chimeric primer design for building HDRa

Part 1: (Tribolium)

tgatttctttgaccaggataagacagatcctgttttcaacgaccatgagggcactgccgacttttacgaa
aattccttgacggagttttacaaagtcgaaaatgcctcgtatcctgatgttaacattaccgggctggata
tttactatttgttgcgcgagggcaacgagaagattttcgatctggacataaagagcgggaatgtaacact
taagacggagcttgattacgaaacagtgaaaagttacactttggtcatacagtcgtccaataaagacacc
attaatgccaatgccaacgaagagaccaagttccagcttacgataaatgtaaaacatttgtgtgaacttt
gcaacaattataagttttcttcaggtccaagatgtgaacgatgaactgccagtgtttgaccagaaggagt
acatttcaattctcacgtccggcacaggcgacgtccttaaaatctctgtaagtattttagtctgaattcg
taaaattggtaaaactgatcggctaggccaccgatatcgatacaacgagccaaggacagctaacttatgc
catagattccatagaatcacactcttcaggtctatcaaa

Part 2: (Tenebrio)

Cctggacataaaatctgcttttaccatgaatacgcaatccggagatattacaattaattttgaagtcaaa
gacagcatggaggg

Part 3: (Tribolium)

Ctattttacactgaacctcagtgttcatgacgaaggtaaaatttttggcacagtaataagtttctaatat
cccttgtctcaaacagacccaactcacatagacagagccaccgctcaaatctacctggtgacttccaaac
acaccgtaaatttcaaatttgaaaacaagcaagaagttgtacaaaatgcgaccgcgtctgtaagtaccga
cccgtagcacaaaatccaattttgtaattttgttcagattaaaaccatactggaagcccagtttaact
actcaacaaccgtcgaagaccctaagaaagactcagacgctgacgaaaacaccgaccaaacaatcgtccc
tgttttcttcctcgacacaaaatcacaccaaccgaaagaagcaacggaaattttaaagtaagtgtacaat
tacacgttacgaaatttaaattaaagtttttagacaagtcaccaaagtggacaaattccagtccctgaaa
gaggagttcctaaaagtgggcttgctcctgatgagtttttcctccgattcggacactaccgaaaacctgg
aagagacgctgaaagcgtggctgatcggagtatcggtggtcctgggggccctctgtctcattctcttgat
tgcttttatcataagaacgcgagcgtaagtcgatgaacgcgtgtgcgaaataccaaaagtttggttttag
tctcaaccaacggatacaaaa

Chimeric Cad 1 (F):tgatttctttgaccaggataa

Chimeric Cad 2 (F):actcttcaggtctatcaaacctggacataaaatctgct

Chimeric Cad 2 (R): agcagattttatgtccaggtttgatagacctgaagagt

Chimeric Cad 3 (F):tcaaagacagcatggagggctattttacactgaacctc

Chimeric Cad 3 (R): gaggttcagtgtaaaatagccctccatgctgtctttga

Chimeric Cad 4 (R): ttttgtatccgttggttgaga

Supplementary Material 3.2

CLUSTAL O(1.2.4) multiple sequence alignment between TcCad1 and the edited versions (HDRa in CR12 and HDRb in CR8-10)

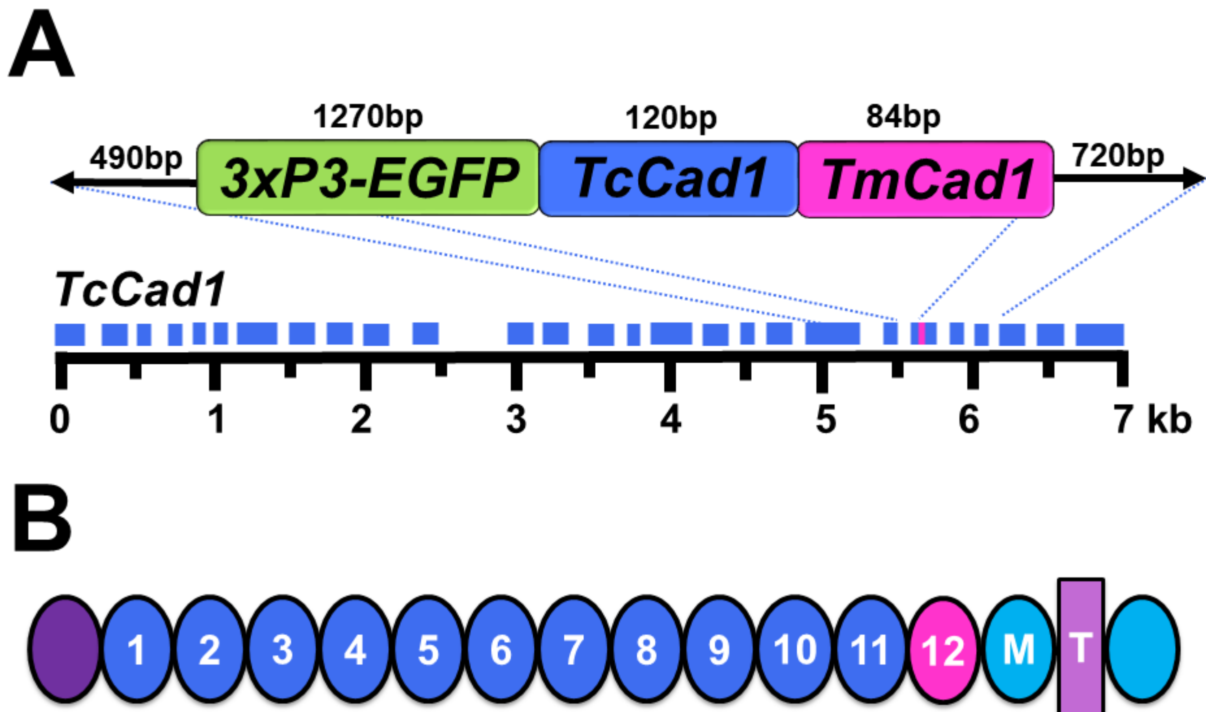
```
HDRB      MRFTFLVFLGILRYVSSFDFKDAQSKGVHFDEETPSGQPVTVDEDNDGPRVVAVLVVNVQ   60
TcCad1    MRFTFLVFLGILRYVSSFDFKDAQSKGVHFDEETPSGQPVTVDEDNDGPRVVAVLVVNVQ   60
HDRA      MRFTFLVFLGILRYVSSFDFKDAQSKGVHFDEETPSGQPVTVDEDNDGPRVVAVLVVNVQ   60
          ************************************************************

HDRB      NSLQINDYDSEKKLDAKSEPGDNANEFKITIKRLDYEKVDSKFFPLEIMDGGSRRTLIVG   120
TcCad1    NSLQINDYDSEKKLDAKSEPGDNANEFKITIKRLDYEKVDSKFFPLEIMDGGSRRTLIVG   120
HDRA      NSLQINDYDSEKKLDAKSEPGDNANEFKITIKRLDYEKVDSKFFPLEIMDGGSRRTLIVG   120
          ************************************************************

HDRB      IKNLDDESPTLTTSNCEMDELTDYTVENSKCIFQVSDPDGFLDKMNFDNIVGTRGEKDIF   180
TcCad1    IKNLDDESPTLTTSNCEMDELTDYTVENSKCIFQVSDPDGFLDKMNFDNIVGTRGEKDIF   180
HDRA      IKNLDDESPTLTTSNCEMDELTDYTVENSKCIFQVSDPDGFLDKMNFDNIVGTRGEKDIF   180
          ************************************************************

HDRB      EFKYKETPTSTATSSDVYLTLKKQLNYDDAILYSFHLEAKDGAGHETTPVKSVPIIVQVN   240
TcCad1    EFKYKETPTSTATSSDVYLTLKKQLNYDDAILYSFHLEAKDGAGHETTPVKSVPIIVQVN   240
HDRA      EFKYKETPTSTATSSDVYLTLKKQLNYDDAILYSFHLEAKDGAGHETTPVKSVPIIVQVN   240
          ************************************************************

HDRB      DIPNKPPVWEKFQSAITIPEKQPYTTQVSARDGDYGINNDISYSVKDTSGYVKIDSNDGT   300
TcCad1    DIPNKPPVWEKFQSAITIPEKQPYTTQVSARDGDYGINNDISYSVKDTSGYVKIDSNDGT   300
HDRA      DIPNKPPVWEKFQSAITIPEKQPYTTQVSARDGDYGINNDISYSVKDTSGYVKIDSNDGT   300
          ************************************************************

HDRB      ITIAPIDRDAGTDQITFDVTATEVGDEESKTTDSILLIIEDIDDNIPKIYIDDSGKIREL   360
TcCad1    ITIAPIDRDAGTDQITFDVTATEVGDEESKTTDSILLIIEDIDDNIPKIYIDDSGKIREL   360
HDRA      ITIAPIDRDAGTDQITFDVTATEVGDEESKTTDSILLIIEDIDDNIPKIYIDDSGKIREL   360
          ************************************************************

HDRB      NVTIDEGATEIDTKISVSDIDFGENAIYTTQLTSDKADFLTAFSIIPGRGYENTTLTLTI   420
TcCad1    NVTIDEGATEIDTKISVSDIDFGENAIYTTQLTSDKADFLTAFSIIPGRGYENTTLTLTI   420
HDRA      NVTIDEGATEIDTKISVSDIDFGENAIYTTQLTSDKADFLTAFSIIPGRGYENTTLTLTI   420
          ************************************************************

HDRB      LDSKKLDFEQPEWRIITMELHTKGTKNSSMADMVPIRISLNDINDETPKFGDIDVIKVSE   480
TcCad1    LDSKKLDFEQPEWRIITMELHTKGTKNSSMADMVPIRISLNDINDETPKFGDIDVIKVSE   480
HDRA      LDSKKLDFEQPEWRIITMELHTKGTKNSSMADMVPIRISLNDINDETPKFGDIDVIKVSE   480
          ************************************************************

HDRB      DVPLGYVITFIQATDADAEDIAAGLKHELLGTLANNILTIEALGGNITTKVDDAFDYEKQ   540
TcCad1    DVPLGYVITFIQATDADAEDIAAGLKHELLGTLANNILTIEALGGNITTKVDDAFDYEKQ   540
HDRA      DVPLGYVITFIQATDADAEDIAAGLKHELLGTLANNILTIEALGGNITTKVDDAFDYEKQ   540
          ************************************************************

HDRB      NEVFIQIMATDLVNHTATAQLTIEVQDVNDESPKLVVSSSIEIDENQDDGTELEATITAS   600
TcCad1    NEVFIQIMATDLVNHTATAQLTIEVQDVNDESPKLVVSSSIEIDENQDDGTELEATITAS   600
HDRA      NEVFIQIMATDLVNHTATAQLTIEVQDVNDESPKLVVSSSIEIDENQDDGTELEATITAS   600
          ************************************************************

HDRB      DEDTTADLEFSIDWSKSYATKNSRRIKDEDFAKYHCVNVETVAGTTNHEASAKLTIIETQ   660
TcCad1    DEDTTADLEFSIDWSKSYATKNSRRIKDEDFAKYHCVNVETVAGTTNHEASAKLTIIETQ   660
HDRA      DEDTTADLEFSIDWSKSYATKNSRRIKDEDFAKYHCVNVETVAGTTNHEASAKLTIIETQ   660
          ************************************************************

HDRB      KGNTPDYELFDSLYVYLMVTDKNQAKNDGSDFALVAITVGDKNDNSPEFSGNANDKKQVT   720
```

```
TcCad1      KGNTPDYELFDSLYVYLMVTDKNQAKNDGSDFALVAITVGDKNDNSPEFSGNANDKKQVT     720
HDRA        KGNTPDYELFDSLYVYLMVTDKNQAKNDGSDFALVAITVGDKNDNSPEFSGNANDKKQVT     720
            ************************************************************

HDRB        ENSATGVVIGSITATDKDVGDNITYSIKPVDEKTPNWVAIDKETGSLFVSLQEGDVIDSD     780
TcCad1      ENSATGVVIGSITATDKDVGDNITYSIKPVDEKTPNWVAIDKETGSLFVSLQEGDVIDSD     780
HDRA        ENSATGVVIGSITATDKDVGDNITYSIKPVDEKTPNWVAIDKETGSLFVSLQEGDVIDSD     780
            ************************************************************

HDRB        KPEGIFTYNVTYTVTASDAVQHNTSIDITIDIKDQNDISPKLGNQETEIFEDNGLDDHSD     840
TcCad1      KPEGIFTYNVTYTVTASDAVQHNTSIDITIDIKDQNDISPKLGNQETEIFEDNGLDDHSD     840
HDRA        KPEGIFTYNVTYTVTASDAVQHNTSIDITIDIKDQNDISPKLGNQETEIFEDNGLDDHSD     840
            ************************************************************

HDRB        QNGAVVLELKPTDGDRDEPFHTVKCSFSSTTSADVTNRFKIDDQNRITVNLSSGATLDRE     900
TcCad1      QNGAVVLELKPTDGDRDEPFHTVKCSFSSTTSADVTNRFKIDDQNRITVNLSSGATLDRE     900
HDRA        QNGAVVLELKPTDGDRDEPFHTVKCSFSSTTSADVTNRFKIDDQNRITVNLSSGATLDRE     900
            ************************************************************

HDRB        TNPEFTFDLRCIDDPTEQGPISNPVDPQPKVTIKLKDVNDNYPTVVTTTLPKLTENTKKG     960
TcCad1      TNPEFTFDLRCIDDPTEQGPISNPVDPQPKVTIKLKDVNDNYPTVVTTTLPKLTENTKKG     960
HDRA        TNPEFTFDLRCIDDPTEQGPISNPVDPQPKVTIKLKDVNDNYPTVVTTTLPKLTENTKKG     960
            ************************************************************

HDRB        PLSTQLQGEDKDAGDNGKMNFAIAAIERCQDESLKNCEKTSDDLFTIETADSQKEATLSL     1020
TcCad1      PLSTQLQGEDKDAGDNGKMNFAIAAIERCQDESLKNCEKTSDDLFTIETADSQKEATLVL     1020
HDRA        PLSTQLQGEDKDAGDNGKMNFAIAAIERCQDESLKNCEKTSDDLFTIETADSQKEATLVL     1020
            ********************************************************** *

HDRB        LGDNLSGLWGRLEATIDVTDKGTPALNGQNVVKIDVAKYNFKEPIFDFPKQGDSYYFKTI     1080
TcCad1      NHDDLRDNYGFLKFTINTTDLGDPALSSESTVSVEVSKFNFEEPKFKFPEDGTKYYLKTT     1080
HDRA        NHDDLRDNYGFLKFTINTTDLGDPALSSESTVSVEVSKFNFEEPKFKFPEDGTKYYLKTT     1080
              *:* . :* *: **:.** * ***..:..*.::*:*:**:** *.**::* .**:**

HDRB        QDKDAPLKLWNGGNMDNVKANDQQGNKYSIKFDVVEDSSDQNLFKVAYLGNSQGQLQLTN     1140
TcCad1      QNKDSSLKMWNDDVLNNLAATDQQKDKYAMLFKLVEDSSGENLFKVSNLDGSQAQLQMST     1140
HDRA        QNKDSSLKMWNDDVLNNLAATDQQKDKYAMLFKLVEDSSGENLFKVSNLDGSQAQLQMST     1140
            *:**: **:**.. ::*: *.*** :**:: *.:*****.:*****: *..**.***::.

HDRB        ADFVPKPYTVTLRASLDVDNS-PANGEASYEVNCTIQIDFFDQDKTDPVFNDHEGTADFY     1199
TcCad1      SSFTAKPYTLTVMAYLDADESQMAPGEKPYSKNCTFQIDFFDQDKTDPVFNDHEGTADFY     1200
HDRA        SSFTAKPYTLTVMAYLDADESQMAPGEKPYSKNCTFQIDFFDQDKTDPVFNDHEGTADFY     1200
            :.*. ****:*: * **.*:*  * **  *. ***:************************

HDRB        ENSLTEFYKVENASYPDVNITGLDIYYLLREGNEKIFDLDIKSGNVTLKTELDYETVKSY     1259
TcCad1      ENSLTEFYKVENASYPDVNITGLDIYYLLREGNEKIFDLDIKSGNVTLKTELDYETVKSY     1260
HDRA        ENSLTEFYKVENASYPDVNITGLDIYYLLREGNEKIFDLDIKSGNVTLKTELDYETVKSY     1260
            ************************************************************

HDRB        TLVIQSSNKDTINANANEETKFQLTINVQDVNDELPVFDQKEYISILTSGTGDVLKISAT     1319
TcCad1      TLVIQSSNKDTINANANEETKFQLTINVQDVNDELPVFDQKEYISILTSGTGDVLKISAT     1320
HDRA        TLVIQSSNKDTINANANEETKFQLTINVQDVNDELPVFDQKEYISILTSGTGDVLKISAT     1320
            ************************************************************

HDRB        DIDTTSQGQLTYAIDSIESHSSGLSNDSVKDYFQIEDPKSGVIKYNFKVSDSMDGYFTLN     1379
TcCad1      DIDTTSQGQLTYAIDSIESHSSGLSNDSVKDYFQIEDPKSGVIKYNFKVSDSMDGYFTLN     1380
HDRA        DIDTTSQGQLTYAIDSIESHSSGLSNLDIKSAFTMN-TQSGDITINFEVKDSMEGYFTLN     1379
            ************************** .:*. * ::  :** *. **:*.***:******

HDRB        LSVHDEDPTHIDRATAQIYLVTSKHTVNFKFENKQEVVQNATASIKTILEAQFNYSTTVE     1439
TcCad1      LSVHDEDPTHIDRATAQIYLVTSKHTVNFKFENKQEVVQNATASIKTILEAQFNYSTTVE     1440
HDRA        LSVHDEDPTHIDRATAQIYLVTSKHTVNFKFENKQEVVQNATASIKTILEAQFNYSTTVE     1439
            ************************************************************
```

```
HDRB       DPKKDSDADENTDQTIVPVFFLDTKSHQPKEATEILKQVTKVDKFQSLKEEFLKVGLLLM    1499
TcCad1     DPKKDSDADENTDQTIVPVFFLDTKSHQPKEATEILKQVTKVDKFQSLKEEFLKVGLLLM    1500
HDRA       DPKKDSDADENTDQTIVPVFFLDTKSHQPKEATEILKQVTKVDKFQSLKEEFLKVGLLLM    1499
           ************************************************************

HDRB       SFSSDSDTTENLEETLKAWLIGVSVVLGALCLILLIAFIIRTRALNQRIQKLSNTKFGSQ    1559
TcCad1     SFSSDSDTTENLEETLKAWLIGVSVVLGALCLILLIAFIIRTRALNQRIQKLSNTKFGSQ    1560
HDRA       SFSSDSDTTENLEETLKAWLIGVSVVLGALCLILLIAFIIRTRALNQRIQKLSNTKFGSQ    1559
           ************************************************************

HDRB       ESGLNRQGLAAPTTNKHAIEGSNPVYNNEVDTKEVDRRSVTSGDSDLIGVEDDEKFNYDN    1619
TcCad1     ESGLNRQGLAAPTTNKHAIEGSNPVYNNEVDTKEVDRRSVTSGDSDLIGVEDDEKFNYDN    1620
HDRA       ESGLNRQGLAAPTTNKHAIEGSNPVYNNEVDTKEVDRRSVTSGDSDLIGVEDDEKFNYDN    1619
           ************************************************************

HDRB       TKEAFD        1625
TcCad1     TKEAFD        1626
HDRA       TKEAFD        1625
           ******
```

Supplementary Material 3.3

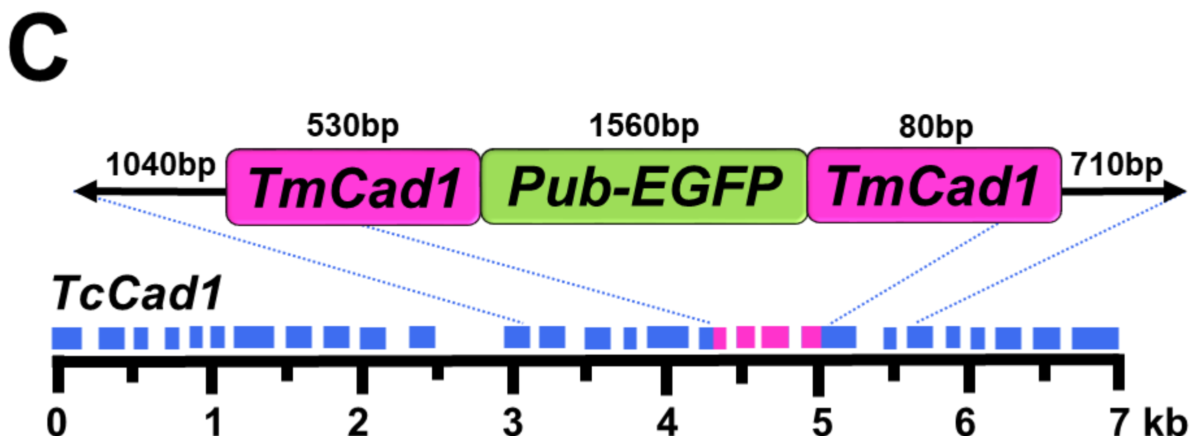
Substitution of CR12 region of TcCad1 gene with TmCad1 and marker gene

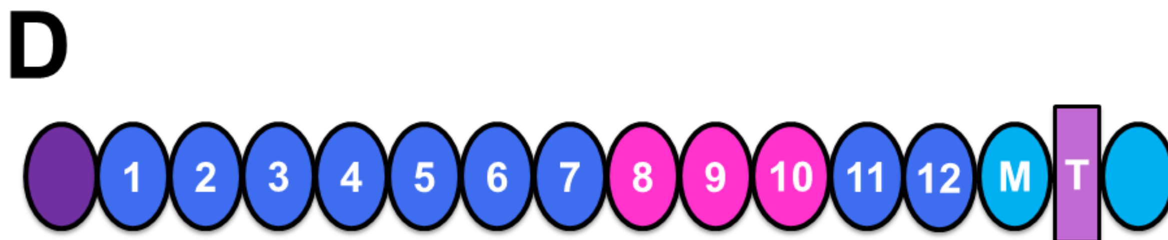


Note. A) *TcCad1* gene with exons (blue) and portion of exon to be substituted with *TmCad1* (pink). Marker gene is shown in green and is located in the intron directly upstream to the exon being changed. Homology directed repair construct (HDRa) with *TmCad1* toxin binding region (pink) and homology arms (black arrows). B) Domain structure of TcCad1 with cadherin repeats (CR 1-12), membrane proximal region (M) and transmembrane domain (T). HDRa substitutes amino acids within CR12 (pink).

Supplementary Material 3.4

Substitution of CR8-10 region of TcCad1 gene with TmCad1 and marker gene





Note. A) *TcCad1* gene with exons (blue) and portion of exon to be substituted with *TmCad1* (pink). Marker gene is shown in green and is located in the third intron between the third and fourth exons being changed. Homology directed repair construct (HDRa) with *TmCad1* toxin binding region (pink) and homology arms (black arrows). B) Domain structure of TcCad1 with cadherin repeats (CR 1-12), membrane proximal region (M) and transmembrane domain (T). HDRa substitutes amino acids within CR12 (pink).

Supplementary Material 3.5

SnapGene feature maps of p(TcCadHDRa-3xP3EGFP) and *p(TcCadHDRb-PubEGFP)*

pGEM T Easy HDR with EGFP
5662 bp

pGEM T Easy HDR Part B
7012 bp

Supplementary Material 3.6

>p(TcCadHDRa-3xP3EGFP)

gggcgaattgggcccgacgtcgcatgctcccggccgccatggctgatttctttgaccaggataagacagatcctgttttcaacgaccat

gagggcactgccgacttttacgaaaattccttgacggagttttacaaagtcgaaaatgcctcgtatcctgatgttaacattaccgggctgg

atatttactatttgttgcgcgagggcaacgagaagattttcgatctggacataaagagcgggaatgtaacacttaagacggagcttgatt

acgaaacagtgaaaagttacactttggtcatacagtcgtccaataaagacaccattaatgccaatgccaacgaagagaccaagttcc

agcttacgataaatgtaaaacatttgtgtgaactttgcaacaattataagttttcttcaggtccaagatgtgaacgatgaactgccagtgttt

gaccagaaggagtacatttcaattctcacgtccggcacaggcgacgtccttaaaatctctgtaagtattttagtctgaattcactagtgatt

ataggcgcgcagctcgcccggggatctaattcaattagagactaattcaattagagctaattcaattaggatccaagcttatcgatttcg

aaccctcgaccgccggagtataaatagaggcgcttcgtctacggagcgacaattcaattcaaacaagcaaagtgaacacgtcgcta

agcgaaagctaagcaaataaacaagcgcagctgaacaagctaaacaatcggggtaccgctagagtcgacggtaccgcgggccc

gggatccaccggtcgccaccatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcga

cgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcacc

accggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacat

gaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactaca

agacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggc

aacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaag

gtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgac

ggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatg

gtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaaagcggccgcgactctagatcataat

cagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattgtt

gttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagt

tgtggtttgtccaaactcatcaatgtatcgatccggcgagacgactaatcgaattcgtaaaattggtaaaactgatcggctaggccaccg

atatcgatacaacgagccaaggacagctaacttatgccatagattccatagaatcacactcttcaggtctatcaaacctggacataaa

atctgcttttaccatgaatacgcaatccggagatattacaattaattttgaagtcaaagacagcatggagggctattttacactgaacctc

agtgttcatgacgaaggtaaaatttttggcacagtaataagtttctaatatcccttgtctcaaacagacccaactcacatagacagagcc

accgctcaaatctacctggtgacttccaaacacaccgtaaatttcaaatttgaaaacaagcaagaagttgtacaaaatgcgaccgcgt

ctgtaagtaccgacccgtagcacaaaatccaatttttgtaatttttgttcagattaaaaccatactggaagcccagtttaactactcaacaa

ccgtcgaagaccctaagaaagactcagacgctgacgaaaacaccgaccaaacaatcgtccctgttttcttcctcgacacaaaatca

caccaaccgaaagaagcaacggaaattttaaagtaagtgtacaattacacgttacgaaatttaaattaaagtttttagacaagtcacca

aagtggacaaattccagtccctgaaagaggagttcctaaaagtgggcttgctcctgatgagtttttcctccgattcggacactaccgaaa

acctggaagagacgctgaaagcgtggctgatcggagtatcggtggtcctgggggccctctgtctcattctcttgattgcttttatcataaga

acgcgagcgtaagtcgatgaacgcgtgtgcgaaataccaaaagtttggttttagtctcaaccaacggatacaaaaggccgcctgcag

gtcgaccatatgggagagctcccaacgcgttggatgcatagcttgagtattctatagtgtcacctaaatagcttggcgtaatcatggtcat

agctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaat

gagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggcc

aacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcga

gcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggc

cagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcg

acgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgtt

ccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttc

ggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttga

gtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgcta

cagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagttaccttcggaa

aaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaa

aaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagat

tatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagtt

accaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacga

tacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaacc

agccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagt

aagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagc

tccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcaga

agtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactg

gtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgcc

acatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttc

gatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatg

ccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgt

ctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctgatgc

ggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggaaattgtaagcgttaatattttgttaaaattcgcgttaaatttt

tgttaaatcagctcattttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgtt

gttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggccc

actacgtgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgattt

agagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggc

aagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtccattcgccattcaggctg

cgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagt

tgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgaattgtaatacgactcactata

Supplementary Material 3.7

>p(TcCadHDRb-PubEGFP)

gggcgaattgggcccgacgtcgcatgctcccggccgccatggcggccgcgggaattcgattaccgataaagatgttggagacaac

atcacttattccataaagtgagataatttagtgttgaacaacgaaaaaaagagtgttttaggcccgttgacgagaaaacgcccaactgg

gttgcaattgataaagaaacagggagtctgtttgtcagtttgcaagaaggtgatgttatagattcggacaaacctgaagggattttcacct

acaatgttacttacaccgtcacggcaagtgatgccgtccagcacaacacctcaattgatgtaataaatacttcgttgattgttcattcaac

acagcaataattttagataactatcgacatcaaagaccaaaatgacatctcgccaaagctaggcaatcaggaaacagagatattcg

aagacaatgggcttgacgaccacagcgaccagaacggagctgtggtgcttgagctaaaaccaaccgatggcgacagagacggta

aaaacttccccacattctagtttgaactcaaacaaaatattttttagaaccgtttcacaccgtaaaatgttcgttttcgtccaccacaagtg

cggacgtgacaaaccgatttaaaattgacgaccagaacaggattacagtcaacttatcatctggagccaccttggaccgagagaca

aatcccgagtttacgtttgatcttagatgcatagacgatcccactgagcaaggaccaatctcaaaccccgtggacccacaaccaaaa

gtaacgatcaaactgaaagacgtcaatgataattacccaactgttgttaccacaactttgccaaaactgaccgagaacactaaaaaa

gtaacacttgcgttgcgattttacgtttaattataataaaatctagggcccacttagcacacagcttcagggcgaggataaggatgcggg

cgataacggaaaaatgaactttgcgatcgctgccatcgaacgctgccaggacgaaagtctgaaaaattgcgaaaaaacctccgatg

atctcttcactattgaaactgcagactcgcagaaggaagccactctgagcctattaggagacaatttgagcggcctgtggggtcgtctg

gaagcgacgattgatgtacgtaccgaaatgcaaagtagcccatcataagttcatttcgtttcaggtaactgacaaggggacgcctgca

ctaaacggtcaaaacgtcgtcaaaattgatgtcgccaagtacaatttcaaagagcctatattcgacttccccaaacagggagattcgta

ctacttcaagacggtattcttgtccacattggttcttccatagagttaacacaaatggaattgcagatacaagataaagacgccccgctc

aaactctggaacggaggcaatatggacaacgtgaaagcgaacgatcaacaaggcaacaaatacagtataaaatttgacgttgttg

aggatagcagcgaccagaatttgtttaaagttgcctatctcggaaattcgcaaggacagcttcagctaactaatgccgactttgtgccga

aaccttacacggtaattactggtcaaaagaagaactacttggcgcgccatcgatacattgatgagtttggacaaaccacaactagaat

gcagtgaaaaaaatgctttatttgtgaaatttgtgatgctattgctttatttgtaaccattataagctgcaataaacaagttaacaacaacaa

ttgcattcattttatgtttcaggttcagggggaggtgtgggaggttttttaaagcaagtaaaacctctacaaatgtggtatggctgattatgat

ctagagtcgcggccgctttacttgtacagctcgtccatgccgagagtgatcccggcggcggtcacgaactccagcaggaccatgtgat

cgcgcttctcgttggggtctttgctcagggcggactgggtgctcaggtagtggttgtcgggcagcagcacggggccgtcgccgatggg

ggtgttctgctggtagtggtcggcgagctgcacgctgccgtcctcgatgttgtggcggatcttgaagttcaccttgatgccgttcttctgcttg

tcggccatgatatagacgttgtggctgttgtagttgtactccagcttgtgccccaggatgttgccgtcctccttgaagtcgatgcccttcagct

cgatgcggttcaccagggtgtcgccctcgaacttcacctcggcgcgggtcttgtagttgccgtcgtccttgaagaagatggtgcgctcct

ggacgtagccttcgggcatggcggacttgaagaagtcgtgctgcttcatgtggtcggggtagcggctgaagcactgcacgccgtaggt

cagggtggtcacgagggtgggccagggcacgggcagcttgccggtggtgcagatgaacttcagggtcagcttgccgtaggtggcat

cgccctcgccctcgccggacacgctgaacttgtggccgtttacgtcgccgtccagctcgaccaggatgggcaccaccccggtgaaca

gctcctcgcccttgctcaccatccataatcgaattcccgcggccgccatggtacgacctgtctcaccacttttgctcagacaatatggcg

aatgcagctgtggaaccgcttttatacggacggcttcccccactcctttttaactctcgaacccaaagtacactttctagaattttcgactaa

ttatccaaaaaacgtcctctttaatcaaaaatttatacataaactttagttgctatcagtttaatgccaattcggttaatcgtggtcaaaacaa

tgtttcaggtattaagaaaattgcttttaattgaacaagattctagggacttggccagtatttgttggtgacatgtcagctgatttttatctacgt

cgtcagctttatcgatcagctgttttctttgttttggaattacgcccttttttatttttcaaattttaatttattttccaatttatgcttttcaaacgtttcaa

ataaaccacccattaatacctgcatgaaattgccaaatttccgtctgttcaattttgaaaggaatttatttgtttacaaaaaattatccttatcg

ttgttgttctaaagtcttgtttatcatgtaaatacggacgacaataggcgcgccataacttccatactaatttttttaccagtcgacggtggtac

ttcaaaaaatgatgaattgtgttccaggtgactctgcgggcatctttggacgtcgacaacagtcccgccaatggagaggcatcttacga

agtgaattgtaccatccaaattgatttctttgaccaggataagacagatcctgttttcaacgaccatgagggcactgccgacttttacgaa

aattccttgacggagttttacaaagtcgaaaatgcctcgtatcctgatgttaacattaccgggctggatatttactatttgttgcgcgagggc

aacgagaagattttcgatctggacataaagagcgggaatgtaacacttaagacggagcttgattacgaaacagtgaaaagttacact

ttggtcatacagtcgtccaataaagacaccattaatgccaatgccaacgaagagaccaagttccagcttacgataaatgtaaaacattt

gtgtgaactttgcaacaattataagttttcttcaggtccaagatgtgaacgatgaactgccagtgtttgaccagaaggagtacatttcaatt

ctcacgtccggcacaggcgacgtccttaaaatctctgtaagtattttagtctgaattcgtaaaattggtaaaactgatcggctaggccacc

gatatcgatacaacgagccaaggacagctaacttatgccatagattccatagaatcacactcttcaggtctatcaaatgactcagtgaa

ggattacttccaaattgaggacccgaaaagcggtgtcatcaagtacaatttcaaagtcagtgacagcatggatggctattttacactga

acctaatcactagtgaattcgcggccgcctgcaggtcgaccatatgggagagctcccaacgcgttggatgcatagcttgagtattctat

agtgtcacctaaatagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagcc

ggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcggg

aaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctca

ctgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcagggg

ataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccata

ggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccagg

cgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtg

gcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagc

ccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaa

caggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtattt

ggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtg

gtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtgg

aacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatc

aatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcc

atagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacc

cacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcc

tccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatc

gtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaa

aagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattct

cttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttg

ctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcga

aaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccag

cgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactc

ttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggtt

ccgcgcacatttccccgaaaagtgccacctgatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggaaa

ttgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcattttttaaccaataggccgaaatcggcaaaatcccttata

aatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaa

agggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagc

actaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaag

aaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcg

ccgctacagggcgcgtccattcgccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagct

ggcgaaagggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtg

aattgtaatacgactcactata

Supplementary Material 3.8

Point mutations from cloning with E. coli

Note. Replication in *E. coli* normally has a very low error rate. However, we observed point mutations that differed between colonies. Top alignment shows mutations in the SV40 terminator following *EGFP* and the bottom alignment shows *EGFP*.

CHAPTER 4

Medea Gene Drive as Transgene in *Drosophila* and Non-*Medea Tribolium*: Hurdles and a Path Forward

Introduction

The red flour beetle, *Tribolium castaneum*, harbors selfish genetic elements that have a unique pattern of maternal-effect lethality. Maternal-Effect Dominant Embryonic Arrest (*Medea*) elements alter the genotypic outcomes of a mating by eliminating any non-*Medea* offspring of a heterozygous *Medea* mother. *Medea* elements are hypothesized to function via a toxin-antidote mechanism whereby *Medea* females produce a toxin that leads to the death of all offspring that fail to inherit the antidote (Beeman et al. 1992). The toxin and antidote functions of *Medea* elements are produced by two tightly-linked loci (Beeman and Friesen 1999). Despite the acquisition of *Medea*-related sequence (Lorenzen et al. 2008), the genes responsible for the *Medea* phenotype have remained a mystery. If we identify the *Medea* genes then they can be used as gene drivers in genetic pest management.

The biology and prevalence of Medea

Medea was initially discovered through a screen for hybrid dysgenesis between geographically diverse *Tribolium* populations (Beeman et al. 1992). When F_1 hybrid females derived from a cross between beetles from Singapore and the US were backcrossed to US males, 50-80% of the progeny died. Lethality was only observed when test crossing heterozygous females, not males, so the authors called this female-specific semi-sterility. Beeman and Friesen (1999) later demonstrated the existence of at least four *Medea* elements in the *Tribolium* genome. All four *Medea* elements have the same phenotype of maternal-effect lethality but have distinct genetic loci. *Medea* elements are widespread in wild populations, with *Medea1* (M^1) and *Medea4* (M^4) being the most prevalent worldwide (Beeman and Friesen 1999). Out of samples collected between 1985 and 1995 from 23 countries, M^4 was found in 14 countries and M^1 in nine. In US samples collected between 1993 and 1995, M^4 was found to be fixed above the 33rd parallel and mixed or absent below it (Beeman 2003).

Recent data offers new insights into the distribution and spread of *Medea* elements within the US (Cash 2016, Cash et al. 2019, Cash et al. 2020). In samples from 2011 to 2014, M^4 was

found to be fixed in many populations above and below the 33rd parallel, and only absent in one sampled population. This indicates the spread and increased distribution of M^4 in the US over the last 20 years. Distribution of M^4 no longer follows a latitudinal pattern. Likewise, M^1 has seen an increased distribution across the US. Samples from 2004 to 2007 indicated fixation of M^1 in the southern US, and absence in the north. Samples from 2011 to 2014 indicate widespread but uneven distribution of M^1, with clusters of high and low frequencies. Each *Medea* loci is unlinked, though M^1 is only found with M^4 in wild populations in the US (Cash et al. 2019). Cash and colleagues hypothesize M^1 arose in an M^4 background and has been tied to the M^4 element's maternal-effect lethality ever since. M^1 and M^4 do not cross rescue, meaning inheriting M^4 does not protect offspring of M^1 mothers. In a laboratory strain, M^1 was successfully segregated from M^4 and purified M^1 stocks are stable. Understanding the spatial and temporal dynamics of *Medea* elements are important for assessing their potential usefulness in applications for pest management. Thus, Cash et al. (2019) call for future studies at a finer geographical scale and highlight the need for molecular markers for the multiple *Medea* elements.

Medea gene candidates

As mentioned above the most promising lead in the search for *Medea* genes came 12 years ago when the DNA sequence associated with *Medea1* (M^1) was discovered and described (Lorenzen et al. 2008). The 21.5-kb insertion found in the M^1 genome was reported to contain many defective genes, and the most promising link to the M^1 phenotype was an intact prokaryotic gene containing a domain of unknown function (DUF). Upon surveying the Tribolium genome, multiple genes possessing this particular DUF (DUF1703) were found. Further insight into the possible functional components of M^1 came 8 years ago when Rapid Amplification of cDNA Ends (RACE) was used to analyze the DUF1703 gene family in Tribolium (Klobasa 2012). Phylogenetic analysis showed there are seven subfamilies of Tribolium DUF1703 genes in Tribolium, and there is a high level of conservation within each subfamily. Subfamilies 1 and 2 are unique in that DUF1703 is fused to a partial *ATP Synthase Subunit C* (*ATPsynthC*) gene. Subfamily 1 includes

the copy found in the M^1-associated insertion and has four exons. Subfamily 2 has a slightly different coding sequence and five exons. To investigate the functionality of the DUF1703 family, transcripts from each subfamily were individually knocked down using RNA interference (RNAi). Out of the seven subfamilies, RNAi was only lethal when targeting subfamily 1 (Klobasa 2012). This means that of all the DUF1703 subfamilies, only the one which includes the M^1- associated copy plays a vital role in Tribolium. Individual DUF1703 genes within each subfamily could not be targeted with RNAi due to the extreme degree of sequence conservation. Regardless, this suggests that the M^1-associated DUF1703 gene may be a functional component of M^1.

To gain insight into what functional role the *ATPSynthC-DUF1703* fusion gene may play, we took a deeper dive into its genomic context, and the various motifs found in the deduced amino acid sequence. Importantly, the gene contains the upstream regulatory region, promoter and first coding exons found in the functional copy of *ATPsynthC*. Therefore, the protein product likely possesses a mitochondrial targeting peptide (MTP) presumably resulting in its transport from the cytoplasm into the mitochondria. It also encodes a truncated transmembrane alpha helix. If this alpha helix functions as an insoluble transmembrane domain, then *ATPSynthC-DUF1703* would likely anchor to the mitochondrial inner membrane.

There is less clarity concerning the DUF1703 portion of the fusion gene. This region within ATPSynthC-DUF1703 was assigned to the DUF1703 family (pfam08011) due to conservation of key amino acid residues (protein motifs) with this family and other members of the PD-(D/E)XK nuclease superfamily (Lorenzen et al 2008). Members of this superfamily are difficult to identify due to little sequence similarity, with only a few conserved active site residues and a common core fold (Knizewski et al. 2007). These restriction endonuclease-like proteins are important for various cellular processes that require nucleic acid cleavage. Within this superfamily, DUF1703 is unique; in addition to the PD-(D/E)XK nuclease domain it has a second protein module categorized as an AAA-ATPase (ATPases Associated with diverse cellular Activities)-like domain (pfam09820), part of the AAA+ superfamily (Knizewski et al. 2007).

The DUF1703 AAA+ module has two domains: an α/β domain and a small anti-parallel four-helix bundle domain. The α/β domain has two characteristic motifs, Walker A (P loop) and Walker B, which together coordinate and stabilize the Mg^{2+} required for ATP hydrolysis (delToro et al. 2016). The AAA+ superfamily couples ATP hydrolysis to a wide variety of cellular functions, the diversity of which is achieved by additional domains or insertions within the domain itself (Miller and Enemark 2016). Given this information, the M^1-associated DUF1703 is likely an ATP dependent nuclease.

The DUF1703 family groups together many hypothetical proteins from various bacterial species (Knizewski et al. 2007, Steczkiewicz et al. 2012) and DUF1703 is thought to have entered the Tribolium genome via an ancient horizontal gene transfer event (Klobasa 2012). Some bacterial endosymbionts of arthropods are considered selfish genetic elements (Hurst 1998) for inducing selfish reproductive manipulation in their host to increase their own transmission. These endosymbionts commonly use toxin-antidote operons and DUF1703 genes have recently been identified in reproductive parasites Wolbacia, Rickettsia, and Cardinium (Beckmann et al. 2017, Siozios et al. 2019).

In this chapter, we test the hypothesis that *ATPsynthC-DUF1703* is the selfish genetic element responsible for M^1 maternal-effect lethality. One potential explanation of the maternal-effect specificity is if the toxin is inherited through the mitochondria or maternal preloading (cytoplasmic inheritance) and the antidote through nuclear inheritance. This is also a possible explanation for how a single gene could act as both toxin and antidote. Specifically, since RNAi targeting subfamily 1 DUF1703 genes is lethal, it is possible that once the M^1 version of *ATPSynthC-DUF1703* is incorporated into the mitochondria it becomes the required isoform. If this is the case, then offspring from heterozygous *Medea* females would be greatly disadvantaged if they failed to inherit the M^1-associated insertion. We test the hypothesis by inserting *ATPsynthC-DUF1703* into a non-*Medea* strain of *Tribolium castaneum*, as well as *Drosophila melanogaster.*

If *ATPsynthC-DUF1703* functions like M[1], then we would expect around 50% mortality of offspring from crosses between *ATPsynthC-DUF1703* heterozygous mothers and wild-type fathers.

Methods

Insect strains

The strains used in this work are as follows: Tribolium M26, a transgenic helper strain that expresses *piggyBac transposase* (described in Lorenzen et al. 2007) that does not carry the M[1]-associated insertion; *vermillion*^{white} (v^W), a white-eyed mutant strain described in Lorenzen et al (2002) that also lacks the M[1]-associated insertion; Drosophila attp40w which has a PhiC31 docking site located on the 2nd chromosome (described in Pfeiffer et al. 2010); Drosophila *y w*, a white-eyed and yellow-bodied mutant strain.

Generation of donor plasmids

Two plasmids were used in this work: a *piggyBac* donor plasmid *pB(TcPubEGFP-ATPsynthCDUF1703)* and an attB-containing donor plasmid *p(attB-DmPubEGFP-ATPsynthCDUF1703)*.

To generate the Tribolium donor plasmid *pB(TcPubEGFP-ATPsynthCDUF1703)*, *ATPsynthCDUF1703* was inserted into *piggyBac* vector *pB(TcPubEGFP-MCS)* which has a multiple cloning site (MCS) SpeI-NcoI-AscI-AflII-FseI-BglII and short piggyBac arms L2 and R2. A few studies in *D. melanogaster* and *B. mori* reported short *piggyBac* arm (L2 and R2) constructs to have higher mobilization efficiency in comparison to constructs with wild-type *piggyBac* arms L1 and R1 (Li et al. 2005, Zhuang et al. 2010, Long et al. 2015). For ease of cloning, construct components were amplified using gene specific primers with an additional overhang of a restriction site on the 5’ end (Table 4.1). Primer restriction sites matched the sites in the MCS.

Table 4.1

Primers for making a *piggyBac* donor plasmid *pB(TcPubEGFP-ATPsynthCDUF1703)*

Name	Sequence	Anneal (°C)	Extension (sec)
MCS Adapter F	CTAGTCCATGGGGCGCGCCCTTAAGGGCCGGCCA	n/a	n/a
MCS Adapter R	GATCTGGCCGGCCCTTAAGGGCGCGCCCCATGGA		
SpeI-TcPUb F	ACTAGTTATTGTCGTCCGTATTTACA	60	30
TcPUb-NcoI R	CCATGGTACGACCTGTCTCACCACTT		
NcoI-EGFP F	CCATGGATGGTGAGCAAGGGCGAGGA	56	30
EGFP-AscI R	GGCGCGCCATCGATACATTGATGAGTTT		
AscI-TcATPSynC F1	TAGGCGCGCCATCAAGCTCCAATTTAGTCTA	59	480
TcDUF-FseI R1	TAGGCCGGCCCAATCCTGAATACCCCGAAAT		
pB short check F	CGGTCTGTATATCGAGGTTT	56	10
pB short check R	CGCATGATTATCTTTTACGTG		

Note. All primers for amplifying plasmid components expect pB short check F and R which are for bacterial colony PCR.

Marker components *TcPub* and *EGFP* were amplified in 50 µl PCR reactions with 5 µl 10X PCR buffer, 4 µl dNTPs, 2 µl of each 25 µM primer, 0.25 µl ExTaq DNA polymerase (Takara Bio, Kusatsu, Shiga, Japan) and 1 ng plasmid template. A previously built plasmid containing this marker was used as template. Conditions followed manufacturer suggestions with an initial denaturation at 98°C for 2 min and final elongation at 72°C for 5 min. See Table 4.1 for annealing temperature and elongation times. Products were purified using QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA) and TA cloned into the pGEM-T Easy vector (Promega, Durham, NC, USA) with a 1 hr ligation at room temperature. TOP10 competent cells were transformed with either *pGem(TcPub)* or *pGem(EGFP)* and the plasmid DNAs purified using ZymoPURE Plasmid Miniprep Kit (Zymo Research, Irvine, CA, USA). The *ATPsynthCDUF1703* transgene was amplified using the same protocol with 2 ng BAC DNA as template (BAC = M^1 bacterial artificial chromosome P22J23 (discussed in Lorenzen et al. 2008)). The PCR product was purified and TA cloned following the same procedure, this time with a 4 hr ligation at room temperature and cloned using SURE supercompetent cells (Agilent, Santa Clara, CA, USA). Overnight cultures were grown at the standard 37ºC except for *pGEM(ATPsynthCDUF1703)* which was grown at 30ºC to reduce recombination.

To generate *pB(MCS)* all existing cargo was excised from *pB(Lchsp83-ZsGreen-attP)* gifted by M. Scott and described in Concha et al. (2011). To remove the unneeded cargo and prepare the plasmid for ligation with a multiple cloning site (MCS) adapter, it was digested with SpeI and BglII (New England Biolabs, Ipswich, Massachusetts, USA) and purified using the Monarch Gel Extraction Kit (New England Biolabs). The adapter SpeI-NcoI-AscI-AflII-FseI-BglII was created with forward and reverse oligonucleotides (Integrated DNA Technologies, Coralville, Iowa, USA) (Table 4.1). To anneal the forward and reverse oligos, 10µl of 2x annealing buffer (20mM Tris, 2mM EDTA, 100 mM NaCl, pH 8.0) was mixed with 5µl each 100µM oligo in a 20 µl reaction volume and the following PCR conditions: 98°C for 1min, 98-88°C for 5 s, decrease 0.1°C/cycle x99 cycles, 88-38°C for 10 s, decrease 0.1°C/cycle x499 cycles, 38-18°C for 1 s,

decrease 0.1°C/cycle x99 cycles, and 18°C indefinitely. Annealed oligos were then phosphorylated using T4 Polynucleotide Kinase (New England Biolabs) with 1µl annealed oligos in a 20µl reaction volume. Phosphorylated oligos were diluted with 30 µl ddH_20. The adapter oligos (1 µl of 5X diluted) were ligated into the vector (GenBank: MT453110, SpeI and BglII digested and purified) using T4 DNA ligase (New England Biolabs) in a total reaction volume of 10 µl for 2 hours at room temperature. The plasmid was transformed into TOP10 competent cells and purified with ZymoPURE Plasmid Miniprep Kit (Zymo Research, Irvine, CA, USA).

The components *TcPub*, *EGFP*, and *ATPsynthCDUF1703* were excised from their respective pGEM plasmids using the appropriate NEB restriction enzymes (see corresponding primers). Components were separated via gel electrophoresis and isolated using the Monarch DNA Gel Extraction Kit (New England Biolabs). Each component was sequentially inserted into the *pB(MCS)* plasmid using the corresponding cloning sites. Each round, the *piggyBac* vector was double digested, cleaned with QIAquick PCR Purification Kit (Qiagen), before the component ligated with T4 DNA Ligase (Promega) into the growing transformation construct. Once the final plasmid was made, ZymoPURE II Plasmid Midiprep (Zymo Research) was used to prepare injection grade plasmid DNA with endotoxin levels of ≤ 0.025 EU/µg DNA. A sample was sent to the Genomic Sciences Laboratory (NCSU, Raleigh, NC, USA) for sequencing with primers from Table 4.2.

Table 4.2

Primers for sequencing piggyBac donor plasmid *pB(TcPubEGFP-ATPsynthCDUF1703)*

Name	Sequence
TcPUb-Seq-R	AATTGCCAAATTTCCGTCTG
TcATPSynC F3	CAGGGGCGTAGAAAACGTAG
TcATPSynC F4	GAGATTCAGTGTCCCAACTCTCC
TcATPSynC R4	TACGTTTTCCACACGAGACG
TcATPSynC F5	CACATTGCTATGGCAACCACT
TcATPSynC F6	GAGTCTGACATAACCGTATAC
TcATPSynC F7	GATTGTGGTTTGATTGGTGCG
TcATPSynC R3	TTTCTACGCCCCTGGTTAAA
TcATPSynC R5	CATTGCCATGGACAAAAATG
TcATPSynC R6	CGCAGGCGAACATTTTATTT
TcATPSynC R7	TCACTGCTGGGAGAAGTGTG
DUF1703 F1	GTTACAGGCAGTTTGTTTATT
DUF1703 R2	GCATTGTTTCGAGCCATTTA
DUF1703 F2	CCAAGATTTCTGTGGATTCACTG
DUF1703 F3	GATAAATCTGGGATAATAATGG
DUF1703 R3	CCATTGTTCGCAATCCTTTT
DUF1703 F4	GGTCTGATGGGTTTTAACGG

To generate the Drosophila plasmid *p(attB-DmPubEGFP-ATPsynthCDUF1703)*, *DmATPsynthCDUF1703* was synthesized by GENEWIZ (Morrisville, NC, USA) then inserted into vector *p(PUbEGFP-attB)*. The vector *p(PUbEGFP-attB)* was gifted by S. Webster and was an intermediate plasmid product from building *p(UASGal80-PUbEGFP-attB)* described in Webster et al. (2020). Five µg of *p(PUbEGFP-attB)* was sent to GENEWIZ for custom cloning with the synthesized insert at the midi-scale. Restriction site ApaI was used for ligation. DNA was confirmed to be endotoxin free and a sample was sequenced (Table 4.3)

Table 4.3

Primers for sequencing attB-containing donor plasmid p(attB-DmPubEGFP-ATPsynthCDUF1703)

Name	Sequence
DmATPSynC F1	CGGGCTCTTAGCGAGTGCA
DmATPSynC F2	GCAGATACACCTTCCTATAGA
DmATPsynC R1	TAGGATCCAGCGACACCGA
DmATPsynC F3	GACGCAGTTTCGTGGGAAAGG
DmATPsynC F4	GTCCGAATTGTGCGAGATAAAG
DmATPsynC F5	ACAACAAACACAGCAGCCAAAG
DmMedea F1	CGCAAATTTAAAGCGCTGAT
DmMedea F2	CGTTGCCAGATAGCTCCATT
DmMedea F3	CCAATTGACATTTGCAGCAG
DmMedea F4	TGCACACACTCGCATTATGA
DmMedea F5	CCTGAAGTTCATCTGCACCA

Microinjections

For transgenesis in Tribolium, *piggyBac* donor plasmid (final concentration 500ng/µl) was injected into *piggyBac* helper line M26. Microinjections followed standard protocols (Berghammer et al. 2009). For transgenesis in Drosophila, attB plasmid (25 µg) was sent to Rainbow Transgenic Flies (Camarillo, CA, USA) where they injected pre-blastoderm embryos (less than 20 mins old) of fly line attP40 with donor and PhiC31 RNA. Surviving flies were crossed with yellow-white (yw) mutant flies, which have yellow bodies and white eyes (parental line of attP40w).

Confirmation of piggyBac excision

To confirm the excision of the *piggyBac* arms and cargo from the plasmid vector, test injections were performed on 100 eggs. DNA was isolated from eggs 24hrs post injection using the E.Z.N.A. Tissue DNA Kit (Omega BioTek) following the manufacturer protocol and eluting into 30µl molecular grade water. These DNA preps contained both gDNA and plasmid DNA. A nested PCR was used to amplify across the TTAA-target site in the plasmid. The primary PCR had a total volume 50 µl with 10 µl 5X PCR buffer, 1 µl each of pB Excision F1 and R1 (25 µM), 0.25 µl MyTaq DNA polymerase (Bioline, Toronto, Canada) and 18 ng template. Primary PCR product was diluted 1:20 and 5 µl used as template in a nested PCR using pB Excision F2 and R2 in a 50 µl reaction volume and the same recipe. See Table 4.4 for primers, annealing temperature, and elongation times. Products were purified using QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA) and sequenced with pB Excision F2 and R2.

Table 4.4

Primers for piggyBac excision assay

Name	Sequence	Anneal (°C)	Extension (sec)
pB Excision F1	TTTTGAACCCGTGGAGGAC	55	10
pB Excision R1	AGCTTGGGCGATGACTGA		
pB Excision F2	TTTACAGCGTGATGGAGCAG	56	10
pB Excision R2	ACCCGACGATATGATCCTGA		

Note. Outer primers are F1 and R1, inner/nested primers are F2 and R2.

Screening for ATPSynthCDUF1703 insertion in beetles

Injected beetle (G_0) eggs were observed for hatch rate and larvae were screened for EGFP from *TcPubEGFP*. Individuals surviving to the pupal stage were sorted by sex, then placed into vials with 5g flour diet. Because of the low number of adults in vials, topper was used to prevent desiccation. Topper allows beetles that flip on their backs to right themselves. Each G_0 adult was given two v^W mates. Offspring from these crosses (F_1) were screened for green fluorescence at the larval stage.

Screening for ATPSynthCDUF1703 insertion in flies

Injected flies were screened by Rainbow Transgenic Flies (Camarillo, CA, USA) for EGFP from *DmPubEGFP*. Adult G_0 flies were outcrossed to *y w* mutants for a total of 24 crosses. For each cross, a single injected male fly was given 3 *y w* virgins, and 2-3 injected virgin females were given 4-6 *y w* male flies.

Results

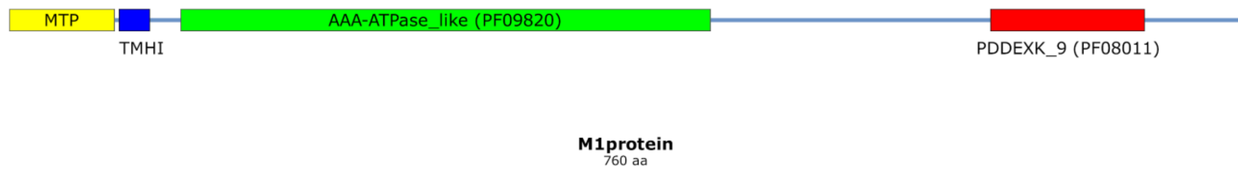
Generation of Tribolium and Drosophila ATPsynthC-DUF1703 transgenes

While the Tribolium *ATPsynthC-DUF1703* transgene could be generated directly from a bacterial artificial chromosome possessing the M^1 locus, production of the Drosophila-specific version first required identification of the Drosophila *ATPsynthC* promoter and mitochondrial targeting peptide (MTP) on the 5' end of the coding sequence (CDS). We used the Neural Network Promoter Prediction tool (Available at https://www.fruitfly.org/seq_tools/ promoter.html) to determine the likely promoter sequence. We found an annotation of the MTP in DmATPSynC (Lovero et al. 2018) which is of particular importance since the MTP directs newly synthesized protein to the mitochondria.

We then used the annotated MTP of DmATPSynC (NP_651852.1) and sequence homology to TcATPSynC (XP_008198899.1) to determine if there is a MTP in TcATPSynC. The MTP region in DmATPSynC and TcATPSynC share 51% identity, TcATPSynC and ATPsynthC-DUF1703 share 98% identity. The prediction tools InterProScan (available at https://www.ebi.ac.uk/interpro/search/sequence/) and iPSORT (available at http://ipsort.hgc.jp/) also predict the first 15-30 amino acids of ATPsynthC-DUF1703 to be signal peptides (Figure 4.1).

Figure 4.1

Predicted domain architecture of ATPsynthC-DUF1703



Note. Predicted domain features: Mitochondrial targeting peptide (MTP; yellow) based on alignment to DmATPSynC and prediction tools InterProScan and iPSORT; Truncated transmembrane helix I from ATPSynC (TMHI; blue) based on alignment to HsATPSyncC); AAA-ATPase-like domain (green) based on PfamScan; PD-(D/E)XK nuclease superfamily 9 (red) based on PD-(D/E)XK recognition web server.

Insertion of the ATPsynthC-DUF1703 transgene in Tribolium

Due to the relatively large size of *ATPsynthC-DUF1703* (7.3kb, and combined with the marker gene 8.9kb) we decided against using CRISPR/Cas9 to insert our transgene. To date there have only been a few CRISPR/Cas9 knockins in Tribolium: 1.2kb (Gilles et al. 2015), 1.3kb (Rylee et al. 2018) and 1.5kb (Chapter 3). Instead we selected to insert our construct into the *Tribolium* genome using the *piggyBac transposase* transformation system (Cary et al. 1989, Tamura et al. 2000). Together with the short *piggyBac* arms L2 and R2, the *ATPsynthC-DUF1703* transgene cassette totaled 9.6kb. There was a mutation (C47T) within the left inverted terminal repeat (Figure 4.2) but the arm was still functional as evidenced by an excision assay (Figure 4.3) and somatic insertion (Figure 4.4).

Figure 4.2

Shortened piggyBac arms and mutation

```
piggyBac arm R2 (238bp)
        TRL                                                 IRL
1    CCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGACGC 60
61   ATGTGTTTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTT 120
121  TATTATATTTACACTTACATACTAATAATAAATTCAACAAACAATTTATTTATGTTTATT 180
181  TATTTATTAAAAAAAAACAAAAACTCAAAATTTCTTCTATAAAGTAACAAAACTTTTA   238

piggyBac arm L2 (309bp)

1     GATATCTATAACAAGAAAATATATATATATAATAAGTTATCACGTAAGTAGAACATGAAATA  60
61    ACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCATGCGTCA  120
121   TTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGCAC  180
181   GCCTCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCG  240
                                                      IRR             TRR
241   CTATTTAGAAAGAGAGAGCAATATTTCAAGAATGCATGCGTCAATTTTACGCAGACTATC  300
         TRR
301   TTTCTAGGG   309
```

Note. Left terminal repeat (TR_L), left inverted repeat right terminal repeat (IR_L), right inverted repeat (IR_R), and right terminal repeat (TR_R). Names R2 and L2 and lengths based on Long et al. (2015) and repeats based on Cary et al. (1989). Mutation C47T in red.

Test injections to assess the functionality of the *piggyBac* arms were successful. Sequencing from plasmid DNA retrieved from injected embryos confirmed the excision of the cargo, with the “TTAA” site being the only remnant (Figure 4.3) Test injections also confirmed the functionality of our EGFP marker (Figure 4.4).

Figure 4.3

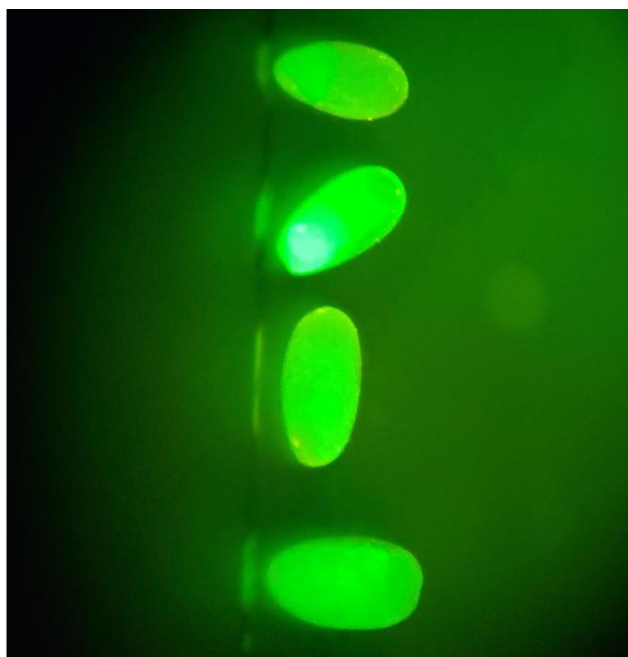
Sequence alignment for PiggyBac excision assay

```
E>TGAAGATGCTCGACACGCTGCAGAACACGCAGCTAGATTAATCTAGCTGCATCAGGA>92

1>~~~~~~~~~~~~~~~~~~~~~~~~ACACGCAGCTAGATTAATCTAGCTGCATCAGGA>47

54<GAAGATGCTCGACACGCTGCAGAACACGCAGC~~~~~~~~~~~~~~~~~~~~~~~~<1
```

Note. TTAA-target site is highlighted in pink. Top line is expected plasmid sequence following excision of the *piggyBac* recognition arms and cargo. Bottom two lines are sequencing results.

Figure 4.4

Transient expression of Pub-EGFP from PiggyBac plasmid



Microinjection of *piggyBac*-based helper/donor constructs aimed at adding the *ATPSynthC-DUF1703* transgene to the Tribolium genome resulted in over 200 G_0 injectees possessing patches of EGFP expression (Figure 4.5, Table 4.5) however, no transgenic G_1 offspring were recovered.

Figure 4.5

Somatic transformation in G_0 beetles.

Note. Far left individual shows no knock-in and is categorized "Non-EGFP". All other individuals show various levels of somatic knock-in and are categorized "EGFP". See Table 4.5 for numbers of each.

Table 4.5

Survival and somatic knock-in of Medea (TcPubEGFP- ATPsynthCDUF1703) in injected beetles

Date	Egg	Larva		Pupa		Adult		Adult*	
	Injected	Hatched	EGFP	Pupated	EGFP	Eclosed	EGFP	Total	EGFP
Subtotal	354	233 (65.8%)	153 (65.7%)	201 (86.3%)	126 (62.7%)	192 (95.5%)	124 (64.6%)	177	118 (66.7%)
Subtotal	289	201 (69.6%)	50 (24.9%)	141 (70.1%)	34 (24.1%)	127 (90.1%)	29 (22.8%)	80	25 (31.3%)
Total	643	434 (67.5%)	203 (46.8%)	342 (78.8%)	160 (46.8%)	319 (93.3%)	153 (48.0%)	257	143 (55.6%)

Note. *Adults that were alive with progeny; Indicates the total number of adults that survived after eclosing (many died) and generated G_1 progeny (some were sterile).

For comparison, previously reported transformation frequencies in Tribolium when using *piggyBac transposase* helper plasmids was 24% for a 8.2-kb donor plasmid and 22% for a 12.4-kb donor plasmid (Lorenzen et al. 2003). The M26 helper line, when used without additional helper plasmid, was reported to have an average 13.8% transformation efficiency (Lorenzen et al. 2007). Therefore, a 9.6-kb donor would be expected to have a transformation efficiency of around 20%. In Chapter 2, the M26 line was used to create lines carrying a *Cas9* transgene. For *pB(Cas9)* injections, survival from larva to adult was 83.4%, death rate of adults was 3.5%, and transformation efficiency was 2.5%. Transformation efficiency of *pB(Cas9)* was lower than expected, because our plasmid concatenated, resulting in a larger than expected insertion. For *pB(Medea)* injections, survival from larva to adult was 73.5%, death rate of adults was 38.9%, and transformation efficiency was 0%. Between *pB(Cas9)* and *pB(Medea)*, survival from larva to adult was comparable, but early adult death rate was significantly different. G_0 individuals injected with *pB(Medea)* had a higher death rate by a factor of 11. This unusual finding suggests that *ATPsynthC-DUF1703* is lethal to Tribolium that do not already possess the 21.5-kb M^1-associated insertion.

Table 4.6

Efficiency of *piggyBac*-mediated germline transformation in *Tribolium castaneum* with two different donors

Donor	Eggs	Larvae	Adults (Total)	Adults (Dead)	Adults (Sterile)	Adults (Viable)	Transfor-mants	Early Adult Death	Survival (Larva to Adult)	Tranformation Efficiency
pB(Cas9)	1108	415	346	12	52	282	7	3.5%	83.4%	2.5%
pB(Medea)	643	434	319	124	20	175	0	38.9%	73.5%	0.0%

Note. Both donors are reported in this dissertation: *pB(Cas9)* from Chapter 2 and *pB(Medea)* from this chapter.

Insertion of the ATPsynthC-DUF1703 transgene in Drosophila

Since there may be complicating factors when adding the *ATPsynthC-DUF1703* fusion to the Tribolium genome, we sought to add the transgene to the Drosophila genome which is completely devoid of DUF1703-related genes. For insertion into the Drosophila genome, we used the Phi31C transformation system (Groth et al. 2004). This gene editing method uses PhiC31 integrase to insert the desired sequence into the insect genome at a known location by using two attachment sequences, attP in the genome and attB in the construct. We selected line attp40 for its high expression level (Pfeiffer et al. 2010).

Despite using a highly reputable service to generate transgenic flies, this too was unsuccessful. It should be noted that transformation frequencies using the phiC31 system in Drosophila are usually around 40%. For example, one study had an average transformation frequency of ~38% for multiple attP lines and integrase constructs (Bischof et al. 2007) while another had 43-55% (Groth et al. 2004). However, with the injections of the *ATPsynthC-DUF1703* fusion we had 0% transformation, despite high numbers of injections (Table 4.7)

The survival rate of injected flies from the larval to adult stage (8.5%) was also abnormally low. Groth et al. 2004 report a survival rate of 57.9% from larva to adult. The extremely low adult survival rate combined with the failure to create transformants suggests that *ATPsynthC-DUF1703* is lethal to flies. Additionally the fly transgenesis experts at Rainbow Gene who performed the injections suggested that the construct might be lethal and would likely not produce transgenics even if further injections were done (personal communication).

Table 4.7

Efficiency of PhiC31-mediated germline transformation in Drosophila melanogaster

Injection Date	Larvae	Adults	Male	Female	Survival from larva to adult	Crosses	Sterile Crosses	Transformants
10/10/2019	75	3	1	2	4.0%	2	1	0
10/10/2019	120	7	2	5	5.8%	4	1	0
10/31/2019	105	17	7	10	16.2%	11	6	0
10/31/2019	145	11	5	6	7.6%	7	2	0
Total	445	38	15	23	8.5%	24	10	0

Further, *ATPsynthC-DUF1703* also seems to be lethal in *E. coli*. During cloning and culture prep of the *ATPsynthC-DUF1703* plasmid construct, dozens of plasmids were obtained that had recombined *ATPsynthC-DUF1703* sequence. Plasmids had large deletions a couple of kilobases long, sections that were inverted, and many point mutations. Once we switched from using Top10 competent cells to SURE cells (a strain engineered to have fewer recombinases) we were finally able to obtain the desired plasmid; yet even this took dozens of minipreps until we obtained the correct plasmid. After several failed cloning attempts with overnight cultures at 37ºC which had no colony growth, a suitable culture was obtained by growing cells at 30ºC. Culturing at the midi-prep scale was another challenge, and cell density and plasmid concentration were consequently much lower than usual.

Despite the lack of transgenic offspring, somatic insertion of the transgene cassette (Figure 4.5) suggests that under some circumstances, *ATPSynthC-DUF1703* may not be toxic. However, even in the naturally occurring system non-M^1 progeny survive from the late embryonic stage (3 days post egg-lay) to as late as the second-larval instar (Beeman et al. 1992) and they would presumably have lots of preloaded toxin or "incompatible" mitochondria. It is possible that since our injectees only had a small percentage of cells in their body with the M^1-associated DUF1703 that they can survive longer. Even so, we observed significantly more mortality from larva to adult and after eclosion than normal. It is also possible that every somatic insertion of the transformation construct was concomitant with inactivating rearrangements within the *ATPSynthC-DUF1703* transgene, similar to what we found after passage through *E. coli*.

Discussion

Below we consider possible evolutionary origins of *Medea* and ways it may function, especially those that conform to the traditional toxin-antidote system. The concept of maternal lethality and zygotic protection has been validated via synthetic *Medea* systems (Chen et al. 2007, Hay et al. 2010, Akbari et al. 2014, Buchman et al. 2018) and several toxin-antidote systems in bacteria have been functionally validated. Of particular interest is the recent identification of

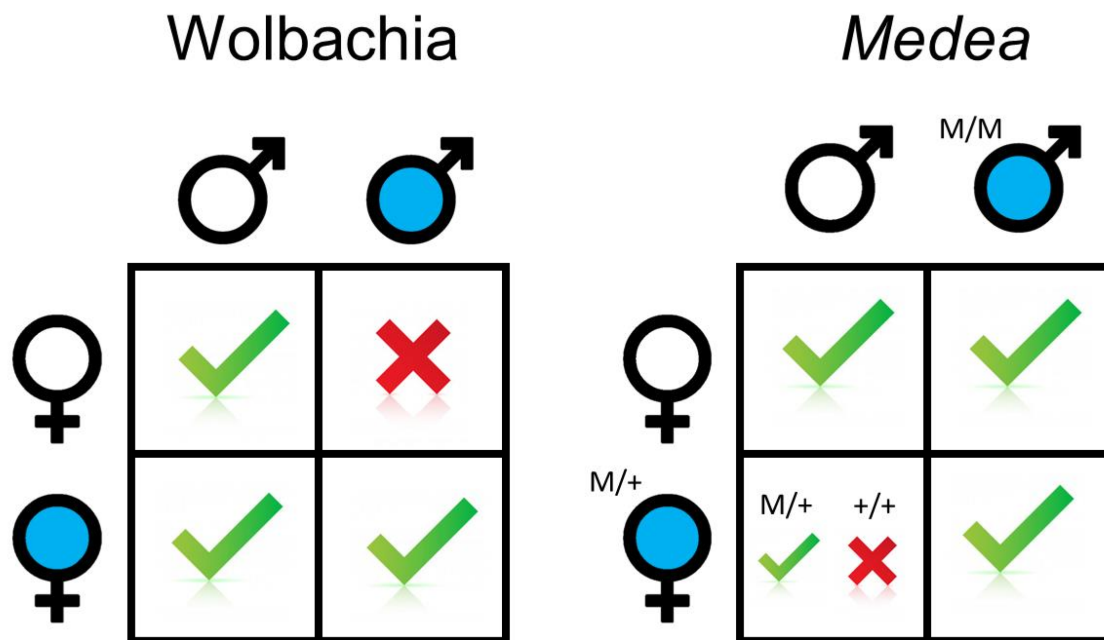
DUF1703 genes in the insect reproductive parasites Wolbacia, Rickettsia, and Cardinium (Beckmann et al. 2017, Siozios et al. 2019). However, the functional role of DUF1703 in a toxin-antidote system has only been made in Wolbachia (Beckmann et al. 2017, Lindsey et al 2018).

Using Wolbachia as a model

If we use the Wolbachia toxin-antidote system as a model, *Medea* can be loosely conceptualized as female-specific cytoplasmic incompatibility (CI). Wolbachia causes post-zygotic male sterility, while *Medea* causes post-zygotic female semi-sterility (Figure 4.6). In CI induced by Wolbachia, sperm carry a toxin and are unable to form viable offspring unless an egg has the antidote. In *Medea*, we hypothesize eggs carry a toxin either within their mitochondria (maternally inherited) or within their cytoplasm (as a maternally preloaded transcript), and the antidote can come from either the sperm or the egg (nuclear inheritance). *Medea* is also similar to cytonuclear incompatibility which results from lethal interactions between genes in the mitochondrial genome (located in the cytoplasm) and nuclear genome, with the caveat that *Medea* toxin is nuclear encoded and later shuttled into the mitochondria.

Figure 4.6

Differences in sterility caused by Wolbachia versus Medea



Note. Only a single possible genotype combination is shown for Medea.

As an example of how this might work, all M^1 offspring of non-M^1 mothers could be born with wild-type mitochondria which are later converted to M^1 mitochondria. M^1 males and females both express the M^1 toxin (nuclease) and antidote (nuclease inhibitor) which get hauled into the mitochondria. M^1 individuals also express the antidote and are safe, but M^1 females pass on their toxic mitochondria. M^1 mothers therefore have female semi-sterility in cases where offspring do not inherit the M^1 antidote in their genome. Wolbachia sperm are thought to carry the toxin and antidote together, but the antidote is unstable and is lost faster than the toxin (Beckmann et al. 2019). This might also be the case with *Medea* and would explain why non-M^1 offspring die, despite presumably receiving some antidote along with the toxin.

Wolbachia CI factors, or cif genes create a toxin-antidote system that vary in their molecular mechanism (Chen et al. 2019). The more specific names cid or cin are used when the enzymatic function of the toxin is known or strongly predicted. The Cid operon disrupts embryonic

nuclear division. CidB is a deubiquitylating enzyme and cleaves ubiquitin from substrates and CidA binds CidB which inactivates the toxin. A paralogous operon involves a nuclease CinB and its binding partner CinA. CinB was reported to be a putative DUF1703 nuclease, though it lacks the characteristic AAA-ATPase domain and has two PD-(D/E)XK nuclease domains. Both nuclease domains are required for CinB activity and toxicity in yeast and flies. CinB is not toxic to *E. coli*, unlike our experience with the M^1-associated DUF1703. Transgenic expression of the *cin* operon in male flies induced male sterility and transgenic CinA in female flies rescued defects in egg-hatch rates (Chen 2019). Expression of CinA also rescued CinB-induced growth defects in yeast.

At the cellular level for both cid and cin systems, insect sperm undergo a Wolbachia-mediated modification event that can be conditionally rescued in the egg by a Wolbachia-encoded factor (Beckmann et al. 2019). Paternal chromatin fails to condense properly at the first zygotic mitosis, causing lethal mis-segregation and bridging of paternal DNA at anaphase. As a future direction of our research, we are investigating two potential antidotes: a hairpin found that is invariant across *Medea* individuals and divergent in non-*Medea* individuals, and a transposase that may potentially be active.

Possible bacterial endosymbiont origins of the M^1 DUF1703 gene

The DUF1703 family groups together many hypothetical proteins from various bacterial species (Knizewski et al. 2007, Steczkiewicz et al. 2012). Species distribution for PDDEXK_9 (PF08011) according to the pfam database includes 425 species of bacteria, six archaea, and 20 eukaryotes: 19 fungi and one metazoan. The pfam database does not list *T. castaneum* but sequence analysis using the PD-(D/E)XK recognition web server (accessible at http://www.ibt.lt/bioinformatics/software/pdexk/) gives a consensus classifier of 0.9987788. A consensus classifier probability of 0.8 or higher is considered to indicate a reliable assignment to the PD-(D/E)XK superfamily (Laganeckas et al. 2011). Additionally, searching the pfam database with PfamScan with the deduced ATPSynthC-DUF1703 sequence returns a significant match to

AAA-ATPase_like (PF09820). See Supplementary Material 4.1 for an alignment of conserved motifs between Tribolium and bacteria.

Due to the wealth of genomic sequence data deposited at NCBI, BlastX analysis using the M^1-associated DUF1703 gene as query reveals high scoring matches with sequences from beetle, cockroach, aphid, and tick genomes (Table 4.8), as well as a range of bacteria and fungi. We suspect the hits to insect genomes to actually be hits to their endosymbionts since most sequences are from unassigned scaffolds. Interestingly, the next most high scoring matches are from bacteria genera Rickettsia and Cardinium (Table 4.9), both known to function as reproductive parasites in arthropods (Goodacre et al. 2015). Previous studies have identified DUF1703 genes in the Rickettsia and Cardinium endosymbionts infecting biting midges (Pilgrim et al. 2017, Siozios et al. 2019), though they have not been linked to CI.

Table 4.8

Top BlastX hits of the M^1-associated DUF1703 in insects

BLAST Name	Organism	Score	Hits
beetles	Tribolium castaneum	1320	24
beetles	Asbolus verrucosus	174	1
beetles	Diabrotica virgifera virgifera	138	2
beetles	Ignelater luminosus	170	4
beetles	Photinus pyralis	131	13
roaches	Blattella germanica	136	1
aphids	Diuraphis noxia	140	3
aphids	Acyrthosiphon pisum	136	2
aphids	Myzus persicae	130	1
aphids	Melanaphis sacchari	128	1
aphids	Aphis gossypii	127	1
aphids	Aphis glycines	124	1
whiteflies	Bemisia tabaci	124	2
aphids	Rhopalosiphum maidis	121	1
aphids	Aphis craccivora	119	1

Table 4.9

Top BlastX hits of the M[1]-associated DUF1703 in bacteria

BLAST Name	Organism	Score	Hits
a-proteobacteria	Candidatus Phycorickettsia trachydisci	141	2
bacteria	Marinitoga hydrogenitolerans	135	2
bacteria	Marinitoga hydrogenitolerans DSM 16785	135	2
a-proteobacteria	spotted fever group	133	2
a-proteobacteria	Rickettsia endosymbiont of Ixodes scapularis	133	2
a-proteobacteria	Rickettsia buchneri	133	2
CFB group bacteria	Candidatus Cardinium hertigii	130	11
bacteria	Marinitoga sp. 1154	129	2
CFB group bacteria	Bacteroidia bacterium	127	2
a-proteobacteria	Rickettsia endosymbiont of Culicoides newsteadi	126	4
bacteria	Marinitoga sp. 1155	125	2
a-proteobacteria	Rickettsiales endosymbiont of Peranema trichophorum	121	6
CFB group bacteria	Bacteroidetes bacterium	124	2
verrucomicrobia	Opitutae bacterium SCGC AG-212-L18	123	2
firmicutes	Caldanaerovirga acetigignens	122	2
firmicutes	Halanaerobium sp.	120	2
firmicutes	Clostridium beijerinckii	119	19
firmicutes	Ruminococcus sp.	118	1
firmicutes	Clostridium sp. P21	118	2
bacteria	Marinitoga sp. 38H-ov	117	2

Wolbachia is the most common parasitic bacteria clade associated with arthropods, with incidence generally reported around 20% of species. Other bacterial reproductive parasites in arthropods include Cardinium, Flavobacteria, Rickettsia, Spiroplasma, and Arsenophonus (Zchori-Fein et al. 2004, Werren et al. 1994, Hacket et al. 1986, Hurst et al. 1997, Gherna et al. 1991). However, little is known about Tribolium symbionts. One study surveyed a single strain of *T. castaneum* which they found to be negative for both Wolbachia and Cardinium (Zchori-Fein and Perlman 2004). Another study surveyed 40 strains of *T. castaneum* for Wolbachia, Rickettsia, and Spiroplasma (Goodacre, Fricke, and Martin 2015). Of these 40 strains, none tested positive for Wolbachia, one carried Rickettsia, and three carried Spiroplasma. These studies exemplify the limited data on endosymbionts in Tribolium. Comparatively more is known about harmful organisms that infect Tribolium such as the entomopathogenic protozoans *Adelina tribolii*, *Farinocystis tribolii*, and *Lymphotropha tribolii* (Bhatia 1937, Rabindra et al. 1981, Ashford 1965).

Owing to the lack of knowledge on bacterial endosymbionts in Tribolium, there are no obvious sources of the DUF1703 gene. Given the dearth of information, the most common endosymbionts of insects that also matched our BLASTX, Wolbachia, Rickettsia, and Cardinium, provide a starting point in the search for the origin of Tribolium DUF1703. The BlastX search returned more hits to *Cardinium* than any other bacteria, but an arthropod survey found 0% incidence of *Cardinium* in beetle species (Coleoptera), with the only positive orders being Diptera (flies), Hymenoptera (wasps and ants), and Hemiptera (piercing/sucking bugs) (Weinert et al. 2015). However, horizontal gene transfer has been detected between *Wolbachia*, *Rickettsia,* and *Cardinium* (Zeng et al. 2018), so it is still worthwhile to investigate *Cardinium*.

Cardinium hergetti is known to cause cytoplasmic incompatibility in the wasp *Encarsia pergendiella* (Hunter et al. 2003) and mites *Eotetranychus suginamensis* and *Bryobia sarothamni* (Gotoh et al. 2007, Ros and Breeuwer 2009). It also causes feminization of genetic males in multiple *Brevipalpus* mite species (Weeks et al. 2001, Chigira and Miura 2005), and thelytokous parthenogenesis in multiple *Encarsia* wasps (Zchori-Fein et al. 2001, Zchori-Fein et al. 2004),

multiple *Brevipalpus* species (Groot and Breeuwer 2006), and the scale *Aspidiotus nerii* (Provencher et al. 2005). One strain of *Cardinium* (cCpun) from biting midges contains an expansion of the DUF1703 gene family which contains an N-terminal AAA-ATPase like domain (unlike Wolbachia) and a C-terminal PDDEXK_9 nuclease domain just like M^1 (Siozios et al. 2019). Genome analysis revealed 25 gene paralogs containing both domains, an additional six with only the AAA-ATPase like domain, and an additional two with only the PDDEXK_9 nuclease domain. This is in contrast to the *Wolbachia* gene *CinB* implicated in CI (reported to be a DUF1703), which does not have the characteristic AAA-ATPase domain and instead has two PDDEXK_9 nuclease domains.

Conclusion

Insects vector hundreds of plant and animal diseases by transmitting pathogenic parasites, viruses, and bacteria. One proposed strategy of disease prevention is to replace wild insect populations with genetically modified individuals unable to transmit pathogens. In order to ensure the spread of anti-pathogen transgenes through wild populations, a gene drive mechanism is needed. One potential gene drive system is *Medea*, which naturally occurs in many populations of red flour beetle (*Tribolium castaneum*). We investigated the function of a *Medea1* candidate gene (*ATPsynthC-DUF1703*) by inserting it as a transgene into *Drosophila* and a non-*Medea* strain of *Tribolium*. Based on these transgenic assays *ATPsynthC-DUF1703* appears to be toxic to both Tribolium and Drosophila.

These findings seem to support the toxin-antidote hypothesis (two tightly-linked genes) over our self-rescuing (single-gene) hypothesis. However, a single gene system cannot be completely ruled out since our transgenic assays may not have mimicked expression levels closely enough. Perhaps expression levels resulted in too much maternal inheritance of "incompatible" mitochondria and too little zygotic expression of the antidote. Nonetheless, the revelation that *ATPsynthC-DUF1703* is lethal brings us closer to uncovering how the *Medea* system works, since it strongly supports *ATPsynthC-DUF1703* as part of *Medea*.

REFERENCES

Akbari, O. S., Chen, C. H., Marshall, J. M., Huang, H., Antoshechkin, I., & Hay, B. A. (2014). Novel synthetic Medea selfish genetic elements drive population replacement in Drosophila; a theoretical exploration of Medea-dependent population suppression. *ACS synthetic biology*, *3*(12), 915-928.

Ashford, R. W. (1965). Lymphotropha tribolii gen. nov., sp. nov., Neogregarinida, Schizocystidae, from the Haemocoele of Tribolium castaneum (Herbst). *The Journal of Protozoology*, *12*(4), 609-615.

Beckmann, J. F., Bonneau, M., Chen, H., Hochstrasser, M., Poinsot, D., Merçot, H., Weill M., Sicard M. & Charlat, S. (2019). The toxin–antidote model of cytoplasmic incompatibility: genetics and evolutionary implications. *Trends in Genetics*, *35*(3), 175-185.

Beeman, R. W., Friesen, K. S., & Denell, R. E. (1992). Maternal-effect selfish genes in flour beetles. *Science*, *256*(5053), 89-92.

Beeman, R. W., & Friesen, K. S. (1999). Properties and natural occurrence of maternal-effect selfish genes ('Medea' factors) in the red flour beetle, Tribolium castaneum. *Heredity*, *82*(5), 529-534.

Beeman, R. W. (2003). Distribution of the Medea factor M4 in populations of Tribolium castaneum (Herbst) in the United States. *Journal of Stored Products Research*, *39*(1), 45-51.

Beeman, R. W., Haas, S., & Friesen, K. S. "Beetle Wrangling Tips: An Introduction to the Care and Handling of Tribolium Castaneum." *Tribolium Stock Maintenance*, USDA ARS, 31 Dec. 2019, www.ars.usda.gov/plains-area/mhk/cgahr/spieru/docs/tribolium-stock-maintenance/.

Berghammer, A. J., Weber, M., Trauner, J., & Klingler, M. (2009). Red flour beetle (Tribolium) germline transformation and insertional mutagenesis. *Cold Spring Harbor Protocols*, *2009*(8), pdb-prot5259.

Bhatia, M. L. (1937). On *Adelina tribolii*, a coccidian parasite of *Tribolium ferrugineum* F. *Parasitology*, *29*(2), 239-246.

Bischof J, Maeda RK, Hediger M, Karch F, Basler K (2007) An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. Proc Natl Acad Sci USA 104(9):3312–3317

Buchman, A., Marshall, J. M., Ostrovski, D., Yang, T., & Akbari, O. S. (2018). Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii. *Proceedings of the National Academy of Sciences*, *115*(18), 4725-4730.

Cary, L. C., Goebel, M., Corsaro, B. G., Wang, H. G., Rosen, E., & Fraser, M. J. (1989). Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*, *172*(1), 156-169.

Cash, S. A. (2016). An Experimental and Theoretical Analysis of the Selfish Genetic Element Medea in Red Flour Beetle Populations.

Cash, S. A., Lorenzen, M. D., & Gould, F. (2019). The distribution and spread of naturally occurring Medea selfish genetic elements in the United States. *Ecology and evolution*, *9*(24), 14407-14416.

Cash, S. A., Robert, M. A., Lorenzen, M. D., & Gould, F. (2020). The impact of local population genetic background on the spread of the selfish element Medea-1 in red flour beetles. *Ecology and evolution*, *10*(2), 863-874.

Chen, C. H., Huang, H., Ward, C. M., Su, J. T., Schaeffer, L. V., Guo, M., & Hay, B. A. (2007). A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. *Science*, *316*(5824), 597-600.

Chen, H., Ronau, J. A., Beckmann, J. F., & Hochstrasser, M. (2019). A Wolbachia nuclease and its binding partner provide a distinct mechanism for cytoplasmic incompatibility. *Proceedings of the National Academy of Sciences*, *116*(44), 22314-22321.

Chigira, A., & Miura, K. (2005). Detection of '*Candidatus* Cardinium'bacteria from the haploid host *Brevipalpus californicus* (Acari: Tenuipalpidae) and effect on the host. *Experimental & applied acarology*, *37*(1-2), 107-116.

Concha, C., Belikoff, E. J., Carey, B. L., Li, F., Schiemann, A. H., & Scott, M. J. (2011). Efficient germ-line transformation of the economically important pest species Lucilia cuprina and Lucilia sericata (Diptera, Calliphoridae). *Insect biochemistry and molecular biology*, *41*(1), 70-75.

delToro, D., Ortiz, D., Ordyan, M., Sippy, J., Oh, C., & Keller, N. et al. (2016). Walker-A Motif Acts to Coordinate ATP Hydrolysis with Motor Output in Viral DNA Packaging. *Journal Of Molecular Biology*, *428*(13), 2709-2729. doi: 10.1016/j.jmb.2016.04.029

Gherna, R.L., Werren, J.H., Weisburg, W., Cote, R., Woese, C.R., Mandelco, L. and Brenner, D.J. (1991) *Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. *International Journal of Systematic Bacteriology*, 41, 563–565.

Gilles, A. F., Schinko, J. B., & Averof, M. (2015). Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle Tribolium castaneum. *Development*, *142*(16), 2832-2839.

Goodacre, S. L., Fricke, C., & Martin, O. Y. (2015). A screen for bacterial endosymbionts in the model organisms *Tribolium castaneum*, *T. confusum*, *Callosobruchus maculatus*, and related species. *Insect science*, *22*(2), 165-177.

Gotoh, T., Noda, H., & Ito, S. (2007). *Cardinium* symbionts cause cytoplasmic incompatibility in spider mites. *Heredity*, *98*(1), 13-20.

Groot, T. V., & Breeuwer, J. A. (2006). *Cardinium* symbionts induce haploid thelytoky in most clones of three closely related *Brevipalpus species*. *Experimental & applied acarology*, *39*(3-4), 257-271.

Groth AC, Fish M, Nusse R, Calos MP (2004) Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics 166(4):1775–1782.

Hackett, K.J., Lynn, D.E., Williamson, D.L., Ginsberg, A.S. and Whitcomb, R.F. (1986) Cultivation of the *Drosophila* sex-ratio spiroplasma. *Science*, 232, 1253–1255.

Handler, A. M., & Harrell Ii, R. A. (1999). Germline transformation of Drosophila melanogaster with the piggyBac transposon vector. *Insect molecular biology*, *8*(4), 449-457.

Hay, B. A., Chen, C. H., Ward, C. M., Huang, H., Su, J. T., & Guo, M. (2010). Engineering the genomes of wild insect populations: challenges, and opportunities provided by synthetic Medea selfish genetic elements. *Journal of Insect Physiology*, *56*(10), 1402-1413.

Hunter, M. S., Perlman, S. J., & Kelly, S. E. (2003). A bacterial symbiont in the Bacteroidetes induces cytoplasmic incompatibility in the parasitoid wasp Encarsia pergandiella. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1529), 2185-2190.

Hurst, G.D.D., Hammarton, T.C., Bandi, C., Majerus, T.M.O., Bertrand, D. and Majerus, M.E.N. (1997) The diversity of inherited parasites of insects: the male-killing agent of the ladybird beetle *Coleomegilla maculata* is a member of the Flavobacteria. *Genetical Research*, 70, 1–6.

Hurst, G. D., & Schilthuizen, M. (1998). Selfish genetic elements and speciation. *Heredity*, *80*(1), 2-8.

Knizewski, L., Kinch, L. N., Grishin, N. V., Rychlewski, L., & Ginalski, K. (2007). Realm of PD-(D/E) XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. *BMC structural biology*, *7*(1), 40.

Klobasa, W. A. (2012). Annotation and Functional Analysis of a Unique Family of Bacteria-Like Genes in the Red Flour Beetle, Tribolium castaneum.

Laganeckas M, Margelevičius M, Venclovas Č. (2011) Identification of new homologs of PD-(D/E)XK nucleases by Support Vector Machines trained on data derived from profile-profile alignments. Nucleic Acids Res. 39(4), 1187-96.

Li, X., Harrell, R. A., Handler, A. M., Beam, T., Hennessy, K., & Fraser Jr, M. J. (2005). piggyBac internal sequences are necessary for efficient transformation of target genomes. *Insect molecular biology*, *14*(1), 17-30.

Lindsey, A. R., Rice, D. W., Bordenstein, S. R., Brooks, A. W., Bordenstein, S. R., & Newton, I. L. (2018). Evolutionary genetics of cytoplasmic incompatibility genes cifA and cifB in prophage WO of Wolbachia. *Genome biology and evolution*, *10*(2), 434-451.

Long, D., Lu, W., Zhang, Y., Bi, L., Xiang, Z., & Zhao, A. (2015). An efficient strategy for producing a stable, replaceable, highly efficient transgene expression system in silkworm, Bombyx mori. *Scientific reports*, *5*(1), 1-14.

Lorenzen, M. D., Brown, S. J., Denell, R. E., & Beeman, R. W. (2002). Transgene expression from the Tribolium castaneum Polyubiquitin promoter. *Insect molecular biology*, *11*(5), 399-407.

Lorenzen, M. D., Berghammer, A. J., Brown, S. J., Denell, R. E., Klingler, M., & Beeman, R. W. (2003). piggyBac-mediated germline transformation in the beetle Tribolium castaneum. *Insect molecular biology*, *12*(5), 433-440.

Lorenzen, M. D., Kimzey, T., Shippy, T. D., Brown, S. J., Denell, R. E., & Beeman, R. W. (2007). piggyBac-based insertional mutagenesis in Tribolium castaneum using donor/helper hybrids. *Insect molecular biology*, *16*(3), 265-275.

Lorenzen, M. D., Gnirke, A., Margolis, J., Garnes, J., Campbell, M., Stuart, J. J., Aggarwal, R., Richards, S., Park, Y., & Beeman, R. W. (2008). The maternal-effect, selfish genetic element Medea is associated with a composite Tc1 transposon. *Proceedings of the National Academy of Sciences*, *105*(29), 10085-10089.

Lovero, D., Giordano, L., Marsano, R. M., Sanchez-Martinez, A., Boukhatmi, H., Drechsler, M., ... & Caggese, C. (2018). Characterization of Drosophila ATPsynC mutants as a new model of mitochondrial ATP synthase disorders. *Plos one*, *13*(8), e0201811.

Miller, J. M., & Enemark, E. J. (2016). Fundamental characteristics of AAA+ protein family structure and function. *Archaea*, *2016*.

Pfeiffer, B. D., Ngo, T. T. B., Hibbard, K. L., Murphy, C., Jenett, A., Truman, J. W., & Rubin, G. M. (2010). Refinement of tools for targeted gene expression in Drosophila. *Genetics*, *186*(2), 735-755.

Pilgrim, J., Ander, M., Garros, C., Baylis, M., Hurst, G. D., & Siozios, S. (2017). Torix group Rickettsia are widespread in Culicoides biting midges (Diptera: Ceratopogonidae), reach high frequency and carry unique genomic features. *Environmental microbiology*, *19*(10), 4238-4255.

Provencher, L. M., Morse, G. E., Weeks, A. R., & Normark, B. B. (2005). Parthenogenesis in the Aspidiotus nerii complex (Hemiptera: Diaspididae): a single origin of a worldwide, polyphagous lineage associated with Cardinium bacteria. *Annals of the Entomological Society of America*, *98*(5), 629-635.

Rabindra, R. J., Balasubramanian, M., & Jayaraj, S. (1981). The effects of *Farinocystis tribolii* on the growth and development of the flour beetle *Tribolium castaneum*. *Journal of invertebrate pathology*, *38*(3), 345-351.

Ros, V. I. D., & Breeuwer, J. A. J. (2009). The effects of, and interactions between, *Cardinium* and *Wolbachia* in the doubly infected spider mite *Bryobia sarothamni*. *Heredity*, *102*(4), 413-422.

Rylee, Johnathan C., Dylan J. Siniard, Kaitlin Doucette, Gabriel E. Zentner, and Andrew C. Zelhof. "Expanding the genetic toolkit of Tribolium castaneum." *PloS one* 13, no. 4 (2018): e0195977.

Siozios, S., Pilgrim, J., Darby, A. C., Baylis, M., & Hurst, G. D. (2019). The draft genome of strain cCpun from biting midges confirms insect Cardinium are not a monophyletic group and reveals a novel gene family expansion in a symbiont. *PeerJ*, *7*, e6448.

Steczkiewicz, K., Muszewska, A., Knizewski, L., Rychlewski, L., & Ginalski, K. (2012). Sequence, structure and functional diversity of PD-(D/E) XK phosphodiesterase superfamily. *Nucleic acids research*, *40*(15), 7016-7045.

Tamura, T., Thibert, C., Royer, C., Kanda, T., Eappen, A., Kamba, M., Kômoto, N., Thomas, J.L., Mauchamp, B., Chavancy, G. and Shirk, P.,Fraser, M., Prudhomme, C., & Couble, P. (2000). Germline transformation of the silkworm Bombyx mori L. using a piggyBac transposon-derived vector. *Nature biotechnology*, *18*(1), 81-84.

Taylor, B. L., & Zhulin, I. B. (1999). PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Reviews*, *63*(2), 479-506.

Webster, S. H., Vella, M. R., & Scott, M. J. (2020). Development and testing of a novel killer–rescue self-limiting gene drive system in Drosophila melanogaster. *Proceedings of the Royal Society B*, *287*(1925), 20192994.

Weeks, A. R., Marec, F., & Breeuwer, J. A. (2001). A mite species that consists entirely of haploid females. *Science*, *292*(5526), 2479-2482.

Weinert, L. A., Araujo-Jnr, E. V., Ahmed, M. Z., & Welch, J. J. (2015). The incidence of bacterial endosymbionts in terrestrial arthropods. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1807), 20150249.

Werren, J.H., Hurst, G.D.D., Zhang, W., Breeuwer, J.A., Stouthamer, R. and Majerus, M.E. (1994) Rickettsial relative associated with male killing in the ladybird beetle (*Adalia bipunctata*). *Journal of Bacteriology*, 176, 388–394.

Zchori-Fein, E., Gottlieb, Y., Kelly, S. E., Brown, J. K., Wilson, J. M., Karr, T. L., & Hunter, M. S. (2001). A newly discovered bacterium associated with parthenogenesis and a change in

host selection behavior in parasitoid wasps. *Proceedings of the National Academy of Sciences*, *98*(22), 12555-12560.

Zchori-Fein, E., Perlman, S.J., Kelly, S.E., Katzir, N. and Hunter, M.S. (2004) Characterization of a 'Bacteroidetes' symbiont in *Encarsia* wasps (Hymenoptera: Aphelinidae): Proposal of '*Candidatus Cardinium hertigii*'. *International Journal of Systematic and Evolutionary Microbiology*, 54, 961–968.

Zchori-Fein, E. & Perlman, S. J. (2004). Distribution of the bacterial symbiont *Cardinium* in arthropods. *Molecular ecology*, *13*(7), 2009-2016.

Zeng, Z., Fu, Y., Guo, D., Wu, Y., Ajayi, O. E., & Wu, Q. (2018). Bacterial endosymbiont Cardinium cSfur genome sequence provides insights for understanding the symbiotic relationship in Sogatella furcifera host. *BMC genomics*, *19*(1), 1-16.

Zhao, Y., Schetelig, M. F., & Handler, A. M. (2020). Genetic breakdown of a Tet-off conditional lethality system for insect population control. *Nature communications*, *11*(1), 1-9.

Zhuang, L., Wei, H., Lu, C., & Zhong, B. (2010). The relationship between internal domain sequences of piggyBac and its transposition efficiency in BmN cells and Bombyx mori. *Acta Biochim Biophys Sin*, *42*(6), 426-431.

SUPPLEMENTARY MATERIAL

Supplementary Material 4.1

CLUSTAL O(1.2.4) multiple sequence alignment of M[1] DUF1703 and bacteria DUF1703 from the PD-(D/E)XK nuclease review paper by Knizewski et al. (2007)

```
                                                                    -        -
                                                EEE  HHHHHHHHH      EEEEEE      HHHHH
Tcastaneum               VEDNPANRLTFSLSPDFEDV-LESDTFVDKSGLMLAVFETDKKPMVLCTAP  FG TVNL   59
Ctetani                  -----MKRIPYGIS-NFQVLRKENYLYVDKTSYIEIL-DRY-APYQFFIRP  FG SLFI   52
Fnucleatum               ----MEKSIGIGIE-DFKKVINENCYYIDKTKWIEEI-IKDKSEIKLFTRP  FG TLNM   54
Bthetaiotaomicron        -MEDLQRLYPIGIQ-TFSEIRERNYLYIDKTEYVYRM-THSSSKYMFLSRP  FG SLLT   57
Bfragilis                ----MAKLYPIGIQ-NFEKIRREGYLYIDKTALVCRL-VKT-GSYYFLSRP  FG SLLI   53
Cchlorochromatii         -----MKPLPVGIQ-TFSEIINQDYLYIDKTGLASNL-INK-YKYVFLSRP  FG SLFL   52
Mhungatei                ---MVPLKLPIGIQ-SFSEIRTGGYAYVDKTPYVASL-VQE-GKYYFLSRP  FG SLFL   54
Achroococcum             ---MSRKRLPIGIQ-TFVRIREDNCYYVDKTGIALQL-IEE-GNYYFLSRP  FG SLFL   54
                                    .:.   *   :    .  ::**:     :         :    ******::

                         HHHHHHH                        HHHHHHHHHHHH             EEEEEEEE
Tcastaneum               SMMQRFLELEVDENGKEKTTVLEEKDNYKLFSSEKFKLKVMENKKFVETHFGQYPIININ   119
Ctetani                  STLENYYDINKKDKFQE------------LFGD-----LYIGKNPTE--ERNKFLVWKIS   93
Fnucleatum               STLKYFFDIHNKDENRK------------LFSG-----LDIEKSP-YFSEQGQYPVVFIS   96
Bthetaiotaomicron        STLHSYFS-----GREE------------LFRG-----LAMEKLEKEW---IQYPVLHFD   92
Bfragilis                STLEAYFQ-----GKKD------------LFRG-----LAMEELEKDW---IKYPILHLD   88
Cchlorochromatii         DTLKNIFE-----GKQE------------LFKN-----LLI---YNQWNWDVTYPVIKIS   87
Mhungatei                DTLDCALS-----GKKD------------LFKG-----LFLDSPESSWDFSVTWPVIRIS   92
Achroococcum             DTLAELFE-----GNEP------------LFRG-----LVA---HGQWDWSRRYPVVRIS   89
                         . :     .       .            ** .                  : :  :.

                                  HHHHHHHHHHHHHHHHHHHHHHH
Tcastaneum               FLCDDDVKSKSDAVNVCRSAISEAFRQHSYLYKSKLSPLEDYEKKDCEQWCDEFTYTGIP   179
Ctetani                  FAGIDAGHGEEELRNSFNFKVASAAREFLYKYGDILGGNKL----------------SME   137
Fnucleatum               LKGIKNTSWKD-ALFNIKSLLREIYE----EFSFIKNSLSPSEQDEFD-----KIWFKKD   146
Bthetaiotaomicron        MSMAKHVDKDS-LVNLLGFMLAEYER----TLGIDA-----------------------A   124
Bfragilis                LNTEKYDTPES-LDRILNDTLAKWEM----VYGTAP-----------------------S   120
Cchlorochromatii         FSGGIRDK-ES-LRKNLFYILKDNQK----RLNIICE----------------------E   119
Mhungatei                LGQRINTTSDE-LHEYLTRIISREAE----RFGCAIN----------------------P   125
Achroococcum             FADGVVNSRQA-LDTRINDLLQVNAQ----RLGLSAF----------------------T   122
                         :        .           :

                                                  -        -
                             HHHHHHHHHHHHHH     EEEEEE   HHHHHHHHHHHHHHHHHHHHHHHHHHH
Tcastaneum               DAKVIKGLGRLAAYLNKFHKRMVFLLV  YDSIISNALYKVKNADPETELEELEEAASFP   239
Ctetani                  NMSA-EMVVAHISMLSKINNKPVFVLI  YDNFANELITGGKQSTYES----ILHGEGF-   191
Fnucleatum               DGEYENALKNLTTYLYKYYKKEVILLI  YDSPLISAYKYQY----------YDEASNF-   195
Bthetaiotaomicron        GTDPNLRLTNLIKRSYEQTGRKVVVLI  YDAPLLDVVHEREN---------LGVLRDV-   174
Bfragilis                ETSIPLRFKGIVQRACEQSGQRVVILI  YDKPMLQAIGNEEL---------GEKYRDT-   170
Cchlorochromatii         KEDPNQCFAELIQQAFEKYQKKVVILI  YDKPILDNIEKIPE---------ALIIRDG-   169
Mhungatei                HISPGFQLEDLIRDVNKKTGLQVVLLV  YDKPILDNISTPSQ---------SAVMRDE-   175
Achroococcum             AADIPGRFAELLREAERIHGQRVVVLI  YDKPILDNLGNPEL---------AREMRDG-   172
                           .    .        .     *.:*:****     .                     .

                                                 -        -
                         HHHHHHHH       EEEEEEE   HHH             EEE               HHH
Tcastaneum               IKVIGYALKKF-SEIFKGAFITGILDVGGVSISTLLGNVLRLQFGSKDLYQDFCGFTEDE   298
Ctetani                  VKVFYKALKDATMDNFNRIFMTGVSPIMLDDLTSGFNITRNYT--LDENLNAMMGFTREE   249
Fnucleatum               FKIFYGEALKT-NPYLKMGIMTGIIRVVKAGIFSDLNNLKVYS-VLDDFYSDFFGFTQKE   253
Bthetaiotaomicron        MRNFYSPLKAC-DPYLRYIFLTGITKFSQLSIFSELNNIKNIS--MDEPYAAICGITGNE   231
Bfragilis                LKGFYSVLKTM-DGYIRFALLTGVTKFGKVSVFSDLNNLNDIS--MDEPYVELCGITEKE   227
Cchlorochromatii         MRDFYTKIKES-DEYLRFVFLTGVSKFSKVSLFSGLNNLEDIS--LNPDFGNVCGYTQDD   226
Mhungatei                LKSFYGMIKDI-DHLLKFVFLTGVSKFAKTGIFSGLNNLNDIT--LDSRYSAICGYTHED   232
Achroococcum             LRNLYSVVKGQ-GAHIRFAMLTGVSKFSKASLFSGLNNLNDIT--LDAKYSAICGYTDED   229
                         .: :           :.  ::**:  .    .: : :.          .     . * *  .:
```

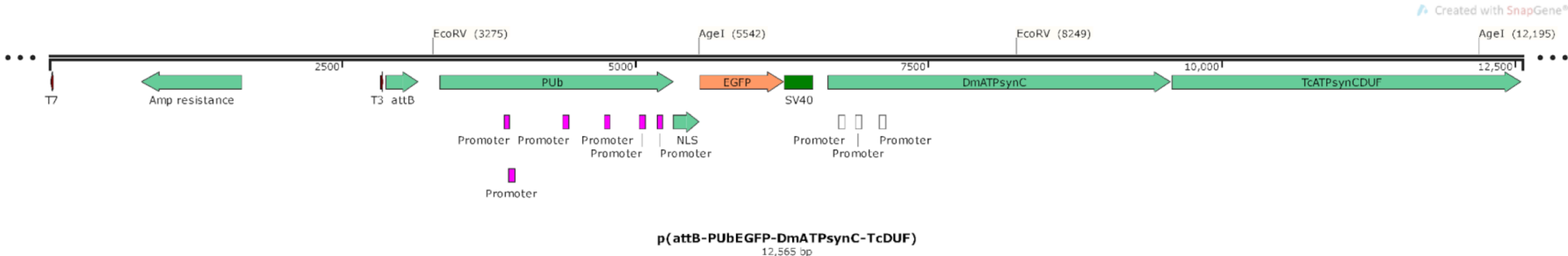
```
                       HHHHHHH  HHHH          HHHHHHH                HHHHHHHHHH
Tcastaneum             VCTLFNRFAVKE---------ELGKKAKDNYNGYLN-GSKQCLYNPQSIVQFLRNLSLGE   348
Ctetani                LSNIMDEVNI----KDTELRVKICTDMTTYYNGYKFNKKADSVFNPDMSMYFLDN----Y   301
Fnucleatum             VEEALKYFGIEYE----------ISQVKTWYDGYKFGNS--DIYNPWSILNFL-S----D   296
Bthetaiotaomicron      ILTQMKKDVDRLAVKLNITAEEVLFKLKENYDGYHFTYPSPDIYNPFSLLNAF-A----D   286
Bfragilis              IHHYLEPEIRQLAKYQKMSYEDACRELKERYDGYHFTENSIGLYNPFSILNTF-Y----K   282
Cchlorochromatii       VDTIFAPYLEGVD----------MAQVKRWYNGYNFLGD--KVYNPFDILLFIKN----Q   270
Mhungatei              LMEVFSDYSKNFL----------ESEIREWYNGYSWTGQ--QVYNPFDILLLF-S----K   275
Achroococcum           VDTVFAPELADLD----------RDEIRQWYNGYNWTGT--SVYNPFDLLLLF-S----K   272
                       :   :                    .    *:**        ::**   :  :


Tcastaneum             GALKNYWQQSGYIHNMYKLFGEGSV---RRLLYELLQGVKEDFQFRLRPLDIRDLKSLHN   405
Ctetani                LAYNEYPEE------MIDNNVKTDYGKVNQLAYN-----------------FNDIEALEE   338
Fnucleatum             RKIRPYWIDTSDNFLIKDILKNVDIDTMDT--------------------------LQK   329
Bthetaiotaomicron      GKFNSYWFGSGTPTYLIKMLDKFDVAPSE----------------------------IGA   318
Bfragilis              MKFGSYWFETGTPSYLVKLLLRDKYDLQQ----------------------------LAH   314
Cchlorochromatii       RMFKNYWFETGTPRFLIELIKKNNYFIPK----------------------------LGK   302
Mhungatei              GMYRPYWFETGTPTFLLKLWQEKPRFPAE----------------------------YDG   307
Achroococcum           REFRNYWFETGTPTFLLDLLTRRQAWLPE----------------------------LGQ   304
                            *         : .   .

Tcastaneum             I-L------------ANQNLTCDQEIVVRFLVSLGYLTYSFDD--------RKLQIPNAE   444
Ctetani                IMTKGETSTMLVDRFNIHTMYSVKENFKSLLFYLGMLTIKGQGPLG-----TVLRVPNYV   393
Fnucleatum             -LFLGENVKVSINGYSDLSSLLEQLDLWEFLLFSGYLTINEKVGE-DYVDIYSLRIPNKE   387
Bthetaiotaomicron      -KTATA--------EDFDAPTETMTNITPLLYQSGYITIKGYD---EELNLYTLDIPNKE   366
Bfragilis              -DEATS--------DMLNCIDSTSKNLLPVIYQSGYLTIKGYD---ERFDIYRLGFPNRE   362
Cchlorochromatii       -IQVNE--------FLVNSFNLENLNLETILFQTGYLTIKQLLS-DVGVSYELGFPNKE   352
Mhungatei              -LVAGE--------ELLGSFDPENIRTETLLFQAGYLTIRSFVSSAEEGTWYTLGFPNRE   358
Achroococcum           -LETDA--------ELLSTFEVDNIATEALMFQAGYLTIDEEINL-DGNWFYRLRYPNRE   354
                                                    .:    * :*                 *  **

Tcastaneum             LIDEITHLLRAYFFERFGINSELAED-CSASFSKLTILCNEGESLAQFKEELAKFQ-NSI   502
Ctetani                IKTIY----WEQYFQRMNLEYNIQPKDVRIAVNEMR-SYGN---IEPLAEIIGEIL-EDL   444
Fnucleatum             VREFF----KKKFIDAKFSESLFRKT------MEAL-VNLK---FNSFEKYLQDILLKST   433
Bthetaiotaomicron      VRIGL----MKSLLPHYVSSPTEKTNTMVAYLSRDI-RNGD---MDAALRRLQEFL-STI   417
Bfragilis              VEEGF----IRYLLPFYANIDKGQTGFHITRFVSEV-EQGD---CDAFFRRLQSFF-ADT   413
Cchlorochromatii       VQMSF----NDYLLHDITTVSEKEPI-R-HELLAII-KAGD---IANLEPIIKRLF-ASI   401
Mhungatei              VRESF----NRQILTVLQDVQNLVPL-P--DIRSVL-ESGN---SQDLHGIFHTFL-SSI   406
Achroococcum           VRQSL----HGSLLKAWTPEGQPVTR-QRMSLYRLL-LAND---FGGLQKLFSAFF-AGI   404
                       :            :                          .          :  :

                                  HHHHHHHHHHHHH          EEEE         EEEEEEE
Tcastaneum             TCLLQRSTLINPLN ATLQSIITL--TCSRAGFLFGNEIIIPKHRTKV VVLFND-----   555
Ctetani                SNRD----L-        -KMMFLTLLGIDSTYFIQSEAENN--KGYV IMLKRKIQYKD   496
Fnucleatum             SYND-------TKN DFYHGLVLGMMFYLDNQYYVKSNEESG--LGRY VIVEPKNK---   481
Bthetaiotaomicron      PQCD------NTKY GHYQQIFYI--IFSLLGFYVDVEVRTP--RGRV MVLRTK-----   462
Bfragilis              PYEL------VRDL LHYQNVLFI--VYKLLGFYVQAEYHTS--NGRI MVLKTD-----   458
Cchlorochromatii       AYNNFTNNY-IESY GFYASVLYA--YFASLGFDIIAEDITN--KGRI LTLKTF-----   451
Mhungatei              PHDNYRNNL-ISRF GYYASVVYA--YLASLGYEIIPEDTTN--KGRI LTVKTR-----   456
Achroococcum           PHDWYRNNP-IAQY GYYASVFYS--HFAALGLDVRVEDATS--HGRI MAVLFN-----   454
                                     *     :.             .  :          *:  :

                          EEEEEEEE           HHHHHHHHH                             EEE
Tcastaneum             --KSGIIM L CDAN-----TDVQDAFE IFQKRYYNVFTN--KKYFDERKKFPSLSFVM   606
Ctetani                ITKFQWII L YIKE------SERNTLEKVKEEGLKQLQGYVESKIVKEQLGEEGLKKAL   550
Fnucleatum             -NNRGFIL F IIDNEKDLEKTSQKAIE IIEKKYD------------TTLKERGIKDIT   528
Bthetaiotaomicron      --TTLYTF L LDK-------NATVAME INLKNYP------------ERFALCGL-PIV   500
Bfragilis              --RYIYVM F FDG-------SAEEALK IEEKGYA------------APFANDPR-QLL   496
Cchlorochromatii       --DKTYIF F VIA---------EEPLE IKRMKYY------------EKYDGER----Y   484
Mhungatei              --SYIWIF F VKGIDH---SGDKEPLK IKERGYA------------EKYAAESR-PVI   498
Achroococcum           --SQVYLF F VVELVP---E--GKALQ IRERGYA------------EKYRTRGE-TIH   494
                              :*:*                  ::::
```

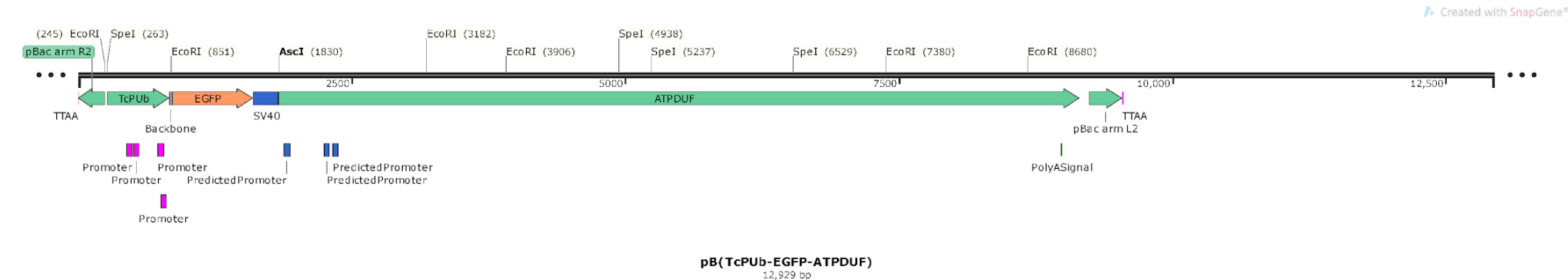
```
                         EEEE
Tcastaneum               AMHVNSDNKASVMC-LRREDVIHVLDDLQLDARKSSLEDRIEIAVNSMGERFINSCREKA   665
Ctetani                  II--------------VVGKKDIYTEQL-------------------------------   564
Fnucleatum               FVGIAFFGKLVKIC-YKN-----------------------------------------   545
Bthetaiotaomicron        KVAVNFDSESCTLGDWKIEKV--------------------------------------   521
Bfragilis                KAGVNFSSKTRNIDCWVVD----------------------------------------   515
Cchlorochromatii         IIGIVFDPKERNVSRFAWERV--------------------------------------   505
Mhungatei                EVGIVFHPGTRNIESWEVGKPDF------------------------------------   521
Achroococcum             LIGVEFSRESRSVVGFEVERDG-------------------------------------   516
```

Note. Adapted from (Knizewski et al. 2007) to also include the M[1]-associated DUF1703. All organisms are bacteria besides *T. castanuem* and *M. hungatei* (archaeon). Predicted secondary structural elements: α-helix (Blue H), β-strand (Green E). AAA+ motifs: Walker A, Walker B, Sensor I, and Sensor II. Core AAA+ helices and strands (blue and green), inserted helices (orange). Residue conservation indicated by highlights: invariant active site resides (red), active site residues in DUF1703 AAA+ (pink and blue), uncharged (yellow), charged or polar (gray), small (red letters). PD-(D/E)XK nuclease fold core (αβββαβ) with an additional β-strand at C-terminus (αβββαββ). Active site resides in DUF1703 PD-(D/E)XK nuclease (black) and alternative active positions with conserved polar/charged residues (blue). Specifically, a charged aspartate (E) in the first α-helix of nuclease (presumably for metal ion binding) and a glutamine (Q) in the second α-helix of nuclease (may take part in binding of nucleic acid substrate).

Supplementary Material 4.2

SnapGene feature maps





Supplementary Material 4.3

> *pB(TcPubEGFP- ATPsynthCDUF1703)*
ttaaccctagaaagataatcatattgtgacgtacgttaaagataatcatgcgtaaaattgacgcatgtgttttatcggtctgtatatcgaggt

ttatttattaatttgaatagatattaagttttattatatttacacttacatactaataataaattcaacaaacaatttatttatgtttatttatttattaaa

aaaaaacaaaaactcaaaatttcttctataaagtaacaaaacttttatcgaattcctgcagcccgggactagttattgtcgtccgtatttac

atgataaacaagactttagaacaacaacgataaggataatttttgtaaacaaataaattcctttcaaaattgaacagacggaaatttgg

caatttcatgcaggtattaatgggtggtttatttgaaacgtttgaaaagcataaattggaaaataaattaaaatttgaaaaataaaaaagg

gcgtaattccaaaacaaagaaaacagctgatcgataaagctgacgacgtagataaaaatcagctgacatgtcaccaacaaatact

ggccaagtccctagaatcttgttcaattaaaagcaattttcttaatacctgaaacattgttttgaccacgattaaccgaattggcattaaact

gatagcaactaaagtttatgtataaatttttgattaaagaggacgtttttggataattagtcgaaaattctagaaagtgtactttgggttcga

gagttaaaaaggagtgggggaagccgtccgtataaaagcggttccacagctgcattcgccatattgtctgagcaaaagtggtgagac

aggtcgtaccatggcggccgcgggaattcgattatggatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggt

cgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccct

gaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccg

ctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaagga

cgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgactt

caaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcaga

agaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaac

acccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgag

aagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtaaagcggccg

cgactctagatcataatcagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacata

aaatgaatgcaattgttgttgttaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaaatttcacaaataaagcat

ttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcgatggcgcgccatcaagctccaatttagtctaatttggtatttttttcag

cctagctgtttaaagaaaagtttaaaaaaaaacacccaaatttttttacacaatttagtgcatttccaagttgtttttaccaaaattttgcatttct

gtttgtaaaacgtgccaaaacaatcgaaaaaacacaaaaataatgccattttcaatctaatctagtatttccagccagtgaaaaacaa

gttactaacactccgccattaaagtagtaccggaccaaaatttagtttttcgggtaaaacggtttagttaattaattaattataattaattaaa

aatttgtttgaatttcccaattaggagtatgattagggcttaagggttgatcaatttgtttaatttgagaaccaatccgtgattatcgatccctgt
aaattcaaaatcgaaaaatagccgcgcaaactctcgcctgcgattggcccaccccgatgtctcacctgtcgccacctgtgagccgcg
ccctaaactccccagctcatcggcgggcggtccggagtcaacccagcacgcacctgcagccattacgattgcgccgttgcattcctgc
ccggtctgtcactccgaagtacacattccaagtgctgggtggctgggtcacctctctcttattctatgggtcacctcctaccgacactgtgc
cgtggtgagtcgccgttaagtgagtacttttcccgcaattttcctcccacaaaacccaactcaagccccttcactgcaccgcttgctgtga
ctattttgctcttttgtttcgtacttgccgggcaacggcgacttgatcgcccctagagttcctttcagttcctttattgggagagttgggaccac
gttgggacactgaacgtcaaaccgagccaattttatattatgtggataaggtatatatttctatcctttttattaaaatttttgttagtaacttatg
tcagaggtttatgtcgaaaataatttcaaggttatagtttttgtcaaaggcataaaatgacattattttacgtcaaatttaaccaggggcgta
gaaaacgtagaaattttgggattactatttaaaaaaagaaaataattgaatacctactaatattaatgttcacatttctgaaatttttcgatat
ttattttattattaaatacataacttaaaacatacttcaatgctaaaaatgcttttctttccgtcacaatttttagttaaacatctcgtcctaaaaat
tgatcatcaagaacccttcacatttctcgtcgtctttttaaatttcccaaaaattaatattaaaggtcgtctcgtgtggaaaacgtaacaaaa
tcgtgaattcccaataactttataatttgaaacagaatctaaagctcacgtaaatgctaaaggcggccaagcagcgaaggttcaacttt
cctaaaaattagaaccaattttgggttcaagtacctacgtattttataaatgcatagacaaattcccaaaatctcatagattttttacataac
attatttagacaatccacttttccaaatgcttgttaataataaacgtaaaacttggcaggaaaggttcaatttctctctctttcatcgctgatta
ccatttcttgaagttaaggaaaaaatgtttgtgatgaagagtttagacgtaaaattcttgtacttacaaaaaaatctcttttaaaatttggagt
tcatataaaaaaaaaacatttctaaacggagttagtcgacaaaaaatggaaagatttttgacaaatttatttgactttttgccaaattactg
atttctgtcccaatttgttttttacacactccaaccgacgtatatgaaacaaaatttcgaaaattgagattcagtgtcccaactctcccaata
taggaactctaatcgcccctaataccgacatcttgcgccattattgcctccaagtgtcacttcagcccggattttcattattgttccgggatttt
cggtgcacgagcgggattctggggcggacgagcaaatcgaggccgctggtcatacgtttctcatatgatttgcgcgaattcggcccga
gttgctgttacgtaacccgtgctggtaggctggggctacctgtgatttgctcacaactaacggacttgcccttgtttcaggttgttaagtgtatt
ccaaccgactaaatttaattaaataaaatgttcgcctgcgctatattgatcgcccccgccgccaggagcgccgtaagtaccacccttttt
atttctactaacaatttgcaaatttttcccaaaactagctgggatttgttgggttatgtaattttagcccaaacgcggttatgtaatgttttcctcg
caatgcttctaatctgcaccaacgtcacgcatttttattccaatttttaaactttgtgtatcaagtccagtaaactttgaacaacatatttttaac
cgagtttcatcatttgatcgatgctttattggttcatggaaatttttatcgcatttcgttcatacagtgtgtcccaaaatcattctataacctttaac
ccaaaatgcacacaagtgtgttacataacacctgacatttcaaatgcccagccaactaattttgaccactcgtgtttttaattatttttacag
tgttatcagttgaggatttcaacagcagtttaattgctcaattgttttacacattgctatggcaaccacttttggggtgtttaaatttcagctttagt

tacatttttaatttgtgacataatagattagataaattaaaatcaaagggcacactgggtgtgtcaccatcactatgaaatagtgataatat

tgtgtgcgattttcctaaataatgttagctgttaggtgaccttgtcacttgtcattttgtccatggcaatgttgattttcaattggcgataaaaa

ataactatcagaaataacctaattttagttatttaggtaaaaaaaagcttgtgaagcagataaaatatcaacattttagctttatttttgtgaa

cttgttttaaagagttcttgaaaaataaataagtttctggttcaagaacgaaccaaatctagtaaactagtaaacttccatattatttgtcact

caattttagcaattttttttacttaagttagaaaattttcggtttatttttagcgaaatattattaaattatagctcttctggctgttagatgtactaaa

acgttcaagaaacgttcaaaaaaaaaatacgtattttttaacattatgtggtcatttttcagtttgttttagccaaaatgtcacaaaataccga

aaatttagttgaaaaataagcttaaaattttgtaaaacgtgaaataacataacttttgattaaattattttcaaaaactagttaaatcgagtg

aaacaatcaattttagcaattttttttgaaaaaaaaatttacattttacttttgagtctgacataaccgtatacttacaagacagatttaaattta

acgttttaaacacatttttaaataaaattttgaaaaataattgcttttctgaaatgtttttttatttttttaacggatttcttcacgttcttaagccaaa

aactcagctcgaacttgaaattttgttgaaactgcactgaaaaatcgtagattaagtcgaatttattattttttttgagcattttaagtcaagtttt

aagttgtaagtttagttttctcgtcaaatttcactaaactggatcagttttgaattggaaatgactttattttgcctcaaaattaaaaaaaaaatt

ttgggcttcgggtaataaattttcagatacttttttttgtactgttggaaacagttggaataaaaatgttaaagtatacaaagtgatttcggctt

tttctctgctaaatctataagaatttttaattttttctcattttactactatttctttcattttaaaaaagagcgtgtcatcaataaaggaaacaga

agtttgtaaaaaataatttattttcataattattacaccatatttgtgataatttagtttaaaaaagtagtgttatttaatatgatgacaatatacat

tcactgaccaccgtgacgaacgtaacttccttggatgtctgaagtggccaaggcttgcgattaacgtttgattggtgattgtggtttgattgg

tgcgttaattttcagcttatttcaaactctaaagtatacttaaggccattgagcaccgccctcagccaaaacccatcattagtccagagtcc

tgtcgttcagcaacacaaacaagccacacttctcccagcagtgaggagtttccagaccaccccagtgtccagggacatcgattcagc

tgctaaatttattggtgcaggcgctgccacagtaggagtagctggatcaggtttgttgcgattgtttttcttttaagattgtttggtgttttgtggc

actaatagattccgagcggtttgttttgtgttatgtttgatgaaaatcatttttttccactctgtagaatatcattaatatattcaaggaatcaaat

ttattatatcctcttggagaaacatccgtttttataagcagactcgtataatttctttgattttaagaatcttcagaatcactcatatcaaaatggt

tgcttttcgtgactgattcgtaaaaaatactagtatgcaatacgttgcaaaatgcctttgtttttttgcagcttattgcataaagagcttagagct

attacaaattctattttgaattttgtcattaatccattgacttacaatgtccgtaaaaggcgcatattgtacatactcaaacttacgtaattattg

cataaattataagttaaagggttaattcaaaattcaacgtttgttaaactcactcaaggtttgtctatcatgcttcccatttctttgtagtatttttc

gatttattgtgcaaacattaattcgtagtttttattcattgttacaggcagtttgtttattgtggaggataatccggccaatcgtttaacatttagc

ctttcaccggattttgaagatgtcttggaatccgatacttttgttgataaaagtgggttgatgttagcagtttttgaaacggacaaaaaacct

atggtattgtgtaccgcaccgcgccgatttggcaaaactgtaaacttatccatgatgcaacgattcttggagttggaagttgatgaaaatg

gaaaggagaaaacaacggttttggaagaaaaagacaattacaaactgttcagttcagaaaagttcaagttaaaagtcatggagaat

aagaagtttgtagaaactcatttcggacaatatccgataatcaatataaattttctctgtgatgatgacgtgaaaagcaaatctgatgcag

tgaatgtgtgtagaagcgctatatcggaagcttttaggcaacacagttacctttacaaatccaaattgtcaccgttggaagactatgaaa

aaaaggattgcgaacaatggtgtgacgaatttacctacacaggaattccagatgccaaagtaatcaaggggttgggacgcttggcag

cttatctgaataagtttcataaaagaatggtttttctactcgttgacgagtacgactcgattatttcaaatgcgttgtataaagtgaagaacg

ctgatccagagactgagcttgaagagctagaggaagctgcctcattcccaattaaagtaattggatatgctctaaaaaaattttctgaga

tctttaaaggagcgtttattaccggaattttggatgtgggtggtgtatcgatttccaccctattgggtaatgtattgcgactgcaattcggaag

caaagatctgtaccaagatttctgtggattcactgaagatgaagtttgtactctatttaatcggtttgcagtaaaggaggaattgggaaaa

aaagcgaaggataactataatgggtatttaaatggctcgaaacaatgcctatataatcctcaatcgattgtgcagtttttgagaaatttatc

tctgggagaaggtgctttaaaaaattattggcaacaatctggttacattcacaatatgtataaactatttggagagggtagcgtgagaag

gttgttgtatgaacttctacaaggtgttaaagaagattttcaattcaggttaagaccgttagacataagagatctgaaatctttgcataacat

attagctaatcaaaatttgacttgtgaccaagaaatagttgttcgatttcttgtatctctaggatacttaacatattcatttgatgatcgcaaact

ccaaatacctaacgctgagctaattgatgaaattacgcacctgctacgagcttattttttgaaagatttggtattaacagtgaattggctga

agattgttcagcatcattttctaagctaaccattctctgtaatgaaggtgaatcattagcacagttcaaagaagaattagcaaaatttcaaa

atagtataacctgtctacttcaacgttcaacacttatcaatccattgaatgaggcaactttacagagtataattacattgacatgtagtcga

gcaggatttctcttcggaaacgaaataataattccaaagcatagaaccaaagttgatgtagtactgtttaatgataaatctgggataata

atggaattaaaatgtgatgcaaatactgatgtacaagacgcatttgaacaaatctttcaaaagcggtactacaatgttttcacaaataag

aagtattttgatgaacgtaagaaatttccttctctttcttttgttatggcaatgcatgtgaattctgacaacaaagcatctgtgatgtgcctgcg

acgcgaggatgttattcacgttctggatgatctgcaacttgacgctagaaagtcatcgctagaagatcgcattgagatagctgttaattct

atgggtgaacgtttcattaacagttgtcgcgaaaaagcaatttttgtctcttaacataccagagaaaccggttcagtttttgtttgaatttaaa

gtagctatttttaagttcaacgcggcatgttaatttcagtttcgatcaactaaatctgttccatttcatatttatatatagttataataaaacatga

attacaaagtttatattgttttagtaggaatgcaaagactccttttgatattggactttaataattggaagccggtctgatgggttttaacggtt

cagttacccaaaccagcttgatcaggaaatgaaatccacaatatttcggggtattcaggattgggccggccagatcttaatacgactca

ctatagggcgaattgggtaccgggccccccctcgaggtcgacggtatcgataagcttgatatctataacaagaaaatatatatataata

agttatcacgtaagtagaacatgaaataacaatataattatcgtatgagttaaatcttaaaagtcacgtaaaagataatcatgcgtcatttt

gactcacgcggtcgttatagttcaaaatcagtgacacttaccgcattgacaagcacgcctcacgggagctccaagcggcgactgaga

tgtcctaaatgcacagcgacggattcgcgctatttagaaagagagagcaatatttcaagaatgcatgcgtcaattttacgcagactatct

ttctagggttaatctagctgcatcaggatcatatcgtcgggtctttttccggctcagtcatcgcccaagctggcgctatctgggcatcggg

gaggaagaagcccgtgccttttcccgcgaggttgaagcggcatggaaagagtttgccgaggatgactgctgctgcattgacgttgag

cgaaaacgcacgtttaccatgatgattcgggaaggtgtggccatgcacgcctttaacggtgaactgttcgttcaggccacctgggatac

cagttcgtcgcggcttttccggacacagttccggatggtcagcccgaagcgcatcagcaacccgaacaataccggcgacagccgg

aactgccgtgccggtgtgcagattaatgacagcggtgcggcgctgggatattacgtcagcgaggacgggtatcctggctggatgccg

cagaaatggacatggataccccgtgagttacccggcgggcgcgcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatcc

gctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgc

gttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttg

cgtattgggcgctcttccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcg

gtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaa

aaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaa

cccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacct

gtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctggg

ctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttat

cgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaact

acggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggca

aacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatc

ttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatcc

ttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcaccta

tctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggcccc

agtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgc

agaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcg

caacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgag

ttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactc

atggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgag

aatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcat

cattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactg

atcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgac

acggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatt

tagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccacctaaattgtaagcgttaatattttgttaaaattcgcgtt

aaatttttgttaaatcagctcattttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttg

agtgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgat

ggcccactacgtgaaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccc

cgatttagagcttgacggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcg

ctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattc

aggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcg

attaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgagcgcgcctcgttcattcacgtttttgaacc

cgtggaggacgggcagactcgcggtgcaaatgtgttttacagcgtgatggagcagatgaagatgctcgacacgctgcagaacacg

cagctaga

Supplementary Material 4.4

> *pB(attB-PubEGFP- DmATPsynthCDUF1703)*

ggcccggtacgtacccaattcgccctatagtgagtcgtattacaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctgg

cgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaa

cagttgcgcagcctgaatggcgaatggaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcattttttaac

caataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtc

cactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaa

tcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccg

gcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgc

gtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcaggtggcacttttcggggaaatgtgcgcggaacccctatt

tgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagtatg

agtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagat

gctgaagatcagttgggtgcacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaac

gttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcat

acactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtg

ctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaa

catgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgc

ctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggag

gcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtct

cgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatg

aacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattga

tttaaaacttcatttttaatttaaaaggatctaggtgaagatcctttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactg

agcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccacc

gctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaata

ctgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtg

gctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacg

ggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgcc

acgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccag

ggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagc

ctatggaaaaacgccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattct

gtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgagg

aagcggaagagcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact

ggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgt

atgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagctcgaaattaaccctcacta

aagggaacaaaagctggctagaactagtgtcgacatgcccgccgtgaccgtcgagaacccgctgacgctgccccgcgtatccgca

cccgccgacgccgtcgcacgtcccgtgctcaccgtgaccaccgcgcccagcggtttcgagggcgagggcttcccggtgcgccgcg

cgttcgccgggatcaactaccgccacctcgacccgttcatcatgatggaccagatgggtgaggtggagtacgcgcccggggagccc

aagggcacgccctggcacccgcaccgcggcttcgagaccgtgacctacatcgtcgacactagtggatccagcggccgcacctgca

ggccggccgttaacacgcgtccgcggtctagactcgaggattccagatctggtaccgctagcatgcgcaaatttaaagcgctgatatc

gatcgcgcgcagatctgtcatgatgatcattgcaattctgcagtcgacggtacccgatcttgtcgccggaacgcagcgacagagattcc

aatgtgtccgtatctttcaggcttttgcccttcagttccagacgaagcgactggcgattcgcgtgtggggtctgcttcagggtcttgtgaatta

gggcgcgcagatcgccgatgggcgtggcgccggagggcaccttcaccttgccgtacggcttgctgttcttcgcgttcaaaatctccag

ctccattttgctttcggtgcgcttgcaatcagtactgtccaaaatcgaaaatcgccgaaccgtagtgtgaccgtgcggggctctgcgaaa

ataaactttttttaggtatatggccacacacggggaaagcacagtggattatatgttttaatattataatatgcaggttttcattacttatccag

atgtaagcccacttaaagcgatttaacaattatttgccgaaagagtaaaaacaaatttcacttaaaaatggattaagaaaagcttgtgta

agattatgcgcagcgttgccagatagctccatttaaaacacttcaaaaacaataagttttgaaaatatatacataaatagcagtcgttgc

cgcaacgctcaacacatcacacttttaaaacaccctttacctacacagaattactttttaaatttccagtcaagctgcgagtttcaaaattat

agccggtagagaagacagtgctatttcaaaagcaaactaaataaacaccaatcctaacaagccttggacttttgtaagtttagatcaa

aggtggcattgcattcaatgtcatggtaagaagtaggtcgtctaggtagaaatcctcattcagccggtcaagtcagtacgagaaaggtc

tcaatttgaaattgtcttaaaaatattttattgttttgtactgtggtgagtttaaacgaaaaacacaaaaaaaaagtgatacacagaaatca

taaaaaattttaatacaaggtattcgtacgtatcaaaaacatttcggcacaatttttttctctgtactaaagtgttacgaacactacggtatttt

ttagtgattttcaacggacaccgaaggtatataaacagcgttcgcgaacggtcgccttcaaaaccaattgacatttgcagcagcaagta

caagcagaaagtaaagcgcaatcagcgaaaaatttatacttaattgttggtgattaaagtacaattaaaagaacattctcgaaagtca

caagaaacgtaagtttttaactcgctgttaccaattagtaataagagcaacaagacgttgagtaatttcaagaaaaactgcatttcaag

gtctttgttcggccatttttttttattcaacgctctacgtaattacaaaataagaaattggcagccacgcatcttgttttcccaatcaattggcat

caaaacgcaaacaaatctataaataaaacttgcgtgttgattttcgccaagatttattggcaaattgtgaaattcgcagtgacgcatttga

aaattcgagaaatcacgaacgcactcgagcatttgtgtgcatgttattagttagttagttctttgcttaattgaagtattttaccaacgaaatc

cacttatttttagctgaaatagagtaggttgcttgaaacgaaagccacgtctggaaaatttcttattgcttagtagttgtgacgtcaccatata

cacacaaaataatgtgtatgcatgcgtttcagctgtgtatatatacatgcacacactcgcattatgaaaacgatgacgagcaacggaa

caggtttctcaactacctttgttcctgtttcttcgctttcctttgttccaatattcgtagagggttaataggggtttctcaacaaagttggcgtcgat

aaataagtttcccatttttattccccagccaggaagttagtttcaatagttttgtaatttcaacgaaactcatttgatttcgtactaattttccaca

tctctattttgacccgcagaataatccaaaatgcagatcggggatcccaccccacccaagaagaagcgcaaggtggaggacgatcc

cgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaat

agcgaagaggcccgcaccgatcgcccttcccaacagttgcggtcgactctagaggatccccgggatccaccggtcgccaccatggt

gagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtg

tccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggc

ccaccctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccg

ccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcga

gggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagta

caactacaacagccacaacgtctatatcatggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatc

gaggacggcagcgtgcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaacca

ctacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgcc

gggatcactctcggcatggacgagctgtacaagtaaagcggccgcgactctagatcataatcagccataccacatttgtagaggtttta

cttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattgttgttgttaacttgtttattgcagcttataatg

gttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatc

ttaaagcttatcgatacgcgtacggcactagagcggccgccaccgcggtggagctccagcttttgttccctttagtgagggttaattagat

cggccggccttggcgcgcccgtaccagggctgcactgcattctcgggctcttagcgagtgcaccctcctaactaatttaagtctagcag

atacaccttcctatagaacattaaatattttgagaatacaaacttatatttcgatgaaaatatgccggccggctctagttatttactaaggca

ttattccataagaaattattacaagaaataaataaaaaattttatattgtttattgcgtatagaattatgaaaagatgccagagggctctagt

cgtttccttttgtatatatagttccgtaaaaattaatgaaaacaataaataataattaattatatattatattatatattatatataatataaaatta

cacttttatttcaatttcgggataggcatacttgaagcttagctaaggtttattaaaatgttaccatatatataatttttacatatgaatttaatattt

tccaccccatcaaatatattttaaaaagacagggaaacttagggcacgatgatgaatattattacaactgaagtctggttatattttatata

cataatttcatcaactatattgatgataaaaatatatttcacaatacatatgtatgtagttaaataaaagaatagattatgtgcatgtaaaac

acaaacgattatacatacttatttggccctacacttgccaaattcattgaaaacttatctaatttatctaattatctaattatctaatttatctattg

catttgctggttgaaaaattttcagaatgacccgtttttatggaccgttcgccaaagattttattttattaatttgtagtataaataaataaacaa

aattagaaagacgcagtttcgtgggaaagggctatttaattgattctacgtatccactactgttatggtcacactgcgcctattgtgcaagg

gcaacctcctattctagcaaagactcgctcctggggcgcgtctctgctgaaaatgtgtagccaggacttgcaccgtggattgggttctca

accaaaatggtatttctggtactttccagtattaaatcgacctatcgtcctttttcccagctcccgctgacggccacacttagcacccccca

aaagtcagcgaaattgtcaaattcgttttactgttgtgaagggttcgttcgaagtgagttgctttttccgggtttttccgtccgctggcggagc

accccccttatcagcacccctctgttgggtaacagccgtgcaaacgcctcagatagccaaaaaacgcggaattagtgaagaatcgtg

tgaaatttcgccaattaaagtgcgaaacgtgtttgataatgcgccgccatcttgcggcacactcgcacacacacacgtgtgccattctgg

ttttgcagtggtgtgcgtgtgtgcctcgagtcgttgcgtcactggaaaagttagaaaatcggcaaattcgcagaggaaaagccagcaa

atcggtttcaattgatgactgagcacagatcgagctgagagtcgagtgaaaatccggtgcgagggcagattttccgcatgttacgtaac

ttttctgaggcggcgtgggacttgatacaatgtattcttgggacagaaagtcagcccagaactaatcgatatctccgtctatatctttgcag

agccgtgcgctagcttcaattttgtcgtaagcaaaacaaaaaggaaattcaaccgaaaatgttcgtgtcgacagtctctcgcattgccc

cagttgccaggagcgccgtaagtaccccttccccatttttatgtccgaattgtgcgagataaagttaaaatgtgcaggaaattacataatt

atcggaaaagtgcattgattttcccgatctaaccatttgaatgccgccctggcgtggtgaacttccaatggccgcggtgagaaaacacg

gacggcacacagttgcataactttgaggttatgtgtgcgcccagtggaatctactaattaaatgcgagaaattgtgaattatcgctcagc

atctgtgcgtagaatttagtgagttcttttatttgcagtttcaaaggctatcccttcattgtataacacctgctttcaggtctgtggtgtgtgtctttg

aggttagaaccggcgaaagctttccagtagggtgttgaaaaatgagaggggtgcggggtaatacaaattgacaataattgacattgttt

ataaaactatagttgttaatatcgggccaccaacaactatgctagtagcctctgaagcacaacaaacacagcagccaaagtgccaat

cgagggaaaattacttaacgattgcataacttggcattacagcagagcaatgaacttggtcggcactcgagggttaagcagcagtatt

agtgtatttgccaagatccggtcatgtgtgatctaaatgcccaagcactaagtttctatttcacacacaaatagttcttacttgtagtacgat

gtgaaggaatttcgggtatttatggaacattttccatagcattcctatacttataaattgaattggtaaacatgataaaccatgaaaagatta

tatgaatcaattctcaatttttgttgtctataagctattagtagtagattagtagtggtaatgggggttaggtaaaatattatgaattttatgcata
atatatctcttatcacggccagaaactttatctcaagcaatgtacacacggcgtggaataccgcttatataatcaacgaacgctcatcta
cctttctagttcctcgccaactccaagcagtacttgcgaccattgagcagcgccatcatcagccagagccagactttggccgctcagaa
cacaacccccgttgcattgctgccacagatcaggtcattccagacctcgccagtcacgcgtgacattgactcggccgccaaattcattg
gtgctggtgccgcaacagtcggtgtcgctggatccggtttgttgcgattgtttttcttttaagattgtttggtgttttgtggcactaatagattccg
agcggtttgttttgtgttatgtttgatgaaaatcatttttttccactctgtagaatatcattaatatattcaaggaatcaaatttattatatcctcttg
gagaaacatccgtttttataagcagactcgtataatttctttgattttaagaatcttcagaatcactcatatcaaaatggttgcttttcgtgactg
attcgtaaaaaatactagtatgcaatacgttgcaaaatgcctttgtttttttgcagcttattgcataaagagcttagagctattacaaattctatt
ttgaattttgtcattaatccattgacttacaatgtccgtaaaaggcgcatattgtacatactcaaacttacgtaattattgcataaattataagtt
aaagggttaattcaaaattcaacgtttgttaaactcactcaaggtttgtctatcatgcttcccatttctttgtagtatttttcgatttattgtgcaaa
cattaattcgtagtttttattcattgttacaggcagtttgtttattgtggaggataatccggccaatcgtttaacatttagcctttcaccggattttg
aagatgtcttggaatccgatacttttgttgataaaagtgggttgatgttagcagtttttgaaacggacaaaaaacctatggtattgtgtaccg
caccgcgccgatttggcaaaactgtaaacttatccatgatgcaacgattcttggagttggaagttgatgaaaatggaaaggagaaaac
aacggttttggaagaaaaagacaattacaaactgttcagttcagaaaagttcaagttaaaagtcatggagaataagaagtttgtagaa
actcatttcggacaatatccgataatcaatataaattttctctgtgatgatgacgtgaaaagcaaatctgatgcagtgaatgtgtgtagaa
gcgctatatcggaagcttttaggcaacacagttacctttacaaatccaaattgtcaccgttggaagactatgaaaaaaaggattgcgaa
caatggtgtgacgaatttacctacacaggaattccagatgccaaagtaatcaaggggttgggacgcttggcagcttatctgaataagttt
cataaaagaatggtttttctactcgttgacgagtacgactcgattatttcaaatgcgttgtataaagtgaagaacgctgatccagagactg
agcttgaagagctagaggaagctgcctcattcccaattaaagtaattggatatgctctaaaaaaattttctgagatctttaaaggagcgtt
tattaccggaattttggatgtgggtggtgtatcgatttccaccctattgggtaatgtattgcgactgcaattcggaagcaaagatctgtacca
agatttctgtggattcactgaagatgaagtttgtactctatttaatcggtttgcagtaaaggaggaattgggaaaaaaagcgaaggataa
ctataatgggtatttaaatggctcgaaacaatgcctatataatcctcaatcgattgtgcagtttttgagaaatttatctctgggagaaggtgc
tttaaaaaattattggcaacaatctggttacattcacaatatgtataaactatttggagagggtagcgtgagaaggttgttgtatgaacttct
acaaggtgttaaagaagattttcaattcaggttaagaccgttagacataagagatctgaaatctttgcataacatattagctaatcaaaat
ttgacttgtgaccaagaaatagttgttcgatttcttgtatctctaggatacttaacatattcatttgatgatcgcaaactccaaatacctaacg
ctgagctaattgatgaaattacgcacctgctacgagcttattttttgaaagatttggtattaacagtgaattggctgaagattgttcagcatc

attttctaagctaaccattctctgtaatgaaggtgaatcattagcacagttcaaagaagaattagcaaaatttcaaaatagtataacctgtc
tacttcaacgttcaacacttatcaatccattgaatgaggcaactttacagagtataattacattgacatgtagtcgagcaggatttctcttcg
gaaacgaaataataattccaaagcatagaaccaaagttgatgtagtactgtttaatgataaatctgggataataatggaattaaaatgt
gatgcaaatactgatgtacaagacgcatttgaacaaatctttcaaaagcggtactacaatgttttcacaaataagaagtattttgatgaa
cgtaagaaatttccttctctttcttttgttatggcaatgcatgtgaattctgacaacaaagcatctgtgatgtgcctgcgacgcgaggatgtta
ttcacgttctggatgatctgcaacttgacgctagaaagtcatcgctagaagatcgcattgagatagctgttaattctatgggtgaacgtttc
attaacagttgtcgcgaaaaagcaatttttgtctcttaacataccagagaaaccggttcagtttttgtttgaatttaaagtagctatttttaagtt
caacgcggcatgttaatttcagtttcgatcaactaaatctgttccatttcatatttatatatagttataataaaacatgaattacaaagtttatat
tgttttagtaggaatgcaaagactcctttttgatattggactttaataattggaagccggtctgatgggttttaacggttcagttacccaaacc
agcttgatcaggaaatgaaatccacaatatttcggggtattcaggattgatggtcgtgccgatgattctcctcgatagacacatacctata
aacggtataagctgtcgtgaaagaggagatgctgtaagaatagag

CHAPTER 5

Responsible Innovation in Genetic Engineering Research: What Does it Mean to Early Career Researchers?

Introduction

The concepts of risk and responsibility are significant features of discourse on biotechnology and have historically focused on health, safety, and the environment. Over the last decade this focus has broadened to address governance and societal issues through public engagement and an emerging paradigm in the literature, responsible innovation (RI). Community and stakeholder engagement address some challenges of biotechnologies, such as uncertainty in their potential risks and benefits and decision-making processes which are unavoidably value laden. Building on engagement, responsible innovation goes further in giving various publics voice in the research and innovation process, responding to their concerns, including them in the anticipation of risks, and asking researchers to reflect alongside various and diverse publics on the underlying purposes, motivations, and goals of innovation (Stilgoe, Owen & Macnaghten 2013). Previous studies have analyzed the perspectives on RI of biotechnologies held by stakeholders and established researchers, yet early career researchers have been largely overlooked. Therefore, we sought to assess and describe what early career researchers think about RI and the feasibility of incorporating it into their research and work. To our knowledge, this is the first account of early career biotech researchers in the US and the second globally, following a study that analyzed early career scientists' concerns about synthetic biology in the UK (McLeod, de Saille & Nerlich 2018).

Responsible Innovation

Responsible innovation (RI) is defined by Stilgoe, Owen & Macnaghten (2013) as 'taking care of the future through collective stewardship of science and innovation in the present.' The goal is to foster inclusive and sustainable research and innovation through anticipation and assessment of potential impacts and societal expectations. RI does not seek to replace codes, standards, certifications, or other mechanisms of research integrity, but rather to enhance the value, benefit and impact that can be generated from science for society. Of multiple existing RI

frameworks, the most cited (Stilgoe, Owen & Macnaghten 2013) proposes four integrated dimensions:

- Anticipation — exploring possible desired futures as well as intended and unintended consequences before technologies are fully developed
- Reflexivity — considering motivations, assumptions, limits of knowledge, biases, worldviews, and alternative framings of problems
- Inclusion — advancing deliberation among citizens, publics, and stakeholders to include alternative perspectives, values, and preferences in governance of research and innovation
- Responsiveness — to react and answer to the other three dimensions; translating the above practices into action as new knowledge, perspectives, views, and norms are identified

The RI framework is transferable across sectors, with potential applications in government, industry, and academia. RI has gained traction in national policies in the EU and UK (Macnaghten, Owen & Jackson 2016), but has yet to have a strong presence in US policy, with the exception of the US approach to nanotechnology governance (Roco et al. 2011). Funding is similar, some major EU and UK research funders have built RI into their funding programs, which is not the case in the US. Despite lack of traction in US policy and funding, RI has had a growing presence in the academic literature over the last ten years. Most publications have come from the Science and Technology Studies (STS) and Science, Technology and Policy (STP) literatures and have focused on emerging technologies like geoengineering, nanotechnology, robotics, and biotechnology (i.e. Stilgoe 2015, Pandza & Ellwood 2013, Stahl & Coeckelbergh 2016, Macnaghten, Owen & Jackson 2016).

Responsible innovation in biotechnology

Many areas of biology have recently had an intense focus on gene editing, which some have named the 'CRISPR-craze' (Pennisi 2013). CRISPR has also sparked a renewed focus on gene drive, a specific type of gene editing meant to spread new genes in the wild (Esvelt et al. 2014). Developers, scholars, regulators, and other stakeholders are actively discussing gene drives in academic literature, policy forums, and the news. These conversations often include questions about ethics and responsibility, and developers have been proactive in discussing how to better engage the public in their research to avoid the mistakes and controversy associated with the 1st generation of Genetically Modified Organisms (GMOs) (Jordan et al. 2017; Kaebnick et al. 2016; Kuzma et al. 2016; Kuzma et al. 2018a, Baltzegar et al. 2018). At the same time, the idea of responsible innovation has been emerging in the EU and in social science literature. Over the past five years multiple groups have presented the RI framework as a tool to foster inclusive and sustainable biotechnology research and innovation. We build on the present literature by describing the previously unheard perspective of US early career biotech researchers. Our analysis additionally focuses on RI within the scope of gene editing and gene drives as well as their use in pest management (known as "genetic pest management" or GPM).

The 1st generation of GMOs (particularly GM crops) were globally controversial, though the US policy approach was and still is relatively pro GM technology. In the EU, societal resistance to GM crops was one impetus for the development of RI. Recently, RI has been suggested as a way to reframe the debate over GM crops in the US (Biddle 2017). Biddle criticizes the current distinction between GM and non-GM crops, which is misleading and unproductive for risk assessment and stakeholder engagement. Instead, they propose using RI to reframe the debate in terms of responsible design of agricultural biotechnologies (regardless if they are GM). Engagement would benefit from RI by helping stakeholders find common ground as it acknowledges potential benefits while also taking seriously concerns about potential social,

ethical, economic, cultural, or political implications. They call for the creation of new spaces and incorporation of RI into institutions, proposals, and experiments.

Extending RI to the commercialization of GM livestock has been proposed as a way to engage a wider-range of stakeholders and publics to inform decision making (Bruce and Bruce 2019). The authors posit that the speed and breadth of genome editing developments are overwhelming to the point that it makes public understanding and ethical reflection difficult. Therefore, they suggest the commercialization stage may be a good point to use RI since it is a slower process that is more conducive to reflection than the research and development stage. This differs from most RI papers which emphasize upstream engagement and anticipation, before products are fully developed. Nonetheless, they identify several challenges like engagement exercises take time and only reach a small number of people, and the knowledge gap of publics on current practices in livestock agriculture could lead to unexpected outcomes from public consultations. Therefore, RI has an important role in considering an innovation in the context of current agricultural processes.

In a few cases, the RI framework has been investigated with respect to GM insects. One paper applied the RI framework to the release of GM mosquitoes in Brazil (De Campos et al. 2017). The GM mosquito project was funded and largely conducted by the government, which may have contributed to the research and development trajectory the project ultimately took. From the perspective of international projects, the authors conclude "responsible innovation needs to be more closely interlinked with a strong notion of political accountability." RI has also been applied to spotted wing drosophila, an invasive pest of berries, for which pest management efforts in the US have focused on GM sterile males, gene drive, and RNAi yeast (Kokotovich et al. 2020). The authors use RI to outline questions for community and stakeholder engagement across decision phases (problem framing, research and development, risk assessment, regulatory review, and deployment, management, and monitoring) and identify potential engagement goals across sectors (Government, Industry, NGOs, and Academia).

In contrast to GM crops, livestock, and insects, organisms with a gene drive are still in the beginning phases of research and have not been approved or deployed anywhere in the world. To date, most gene drive research has been in insects. Gene drive insects differ from regular GM insects in their ability to spread engineered traits through a wild population at a greater rate than would normally be expected. Much of the gene drive literature calls for upstream engagement and inclusive decision-making frameworks. One workshop, Roadmap to Gene Drives, asked participants to make sense of the ways social, policy, economic, and ecological variables inform research agendas, regulatory oversight, and the pursuit of RI around gene drives (Kuzma et al. 2018). The authors call for collaborative policy design and responsiveness to the concerns of publics and stakeholders early in gene drive research and deployment. This aligns with the 2016 report titled 'Gene Drives on the Horizon' by the US National Academy of Sciences, Engineering, and Medicine (NASEM) which supports the use of public and stakeholder engagement to address risk and human values, especially in light of current knowledge limitations.

Another paper suggests incorporating RI principles into risk assessment of GM insects and gene drive organisms (Kuzma 2019). RI has also been used to identify economic issues developers and regulators should consider for gene drives (Mitchell, Brown, and McRoberts 2018). Their takeaway was individual-level benefits may be partly or wholly offset by unintended, aggregate and long-term effects that play out through complex feedback loops. RI has many potential applications but is not without challenges. In the context of global governance of gene drives, one paper points out the limitations of current possible governance structures (RI among others) to safeguard its use (Kofler et al. 2018). They highlight the importance of community inclusion and propose the integrated deliberation framework to more formally involve local communities in the gene drive innovation process.

Stakeholder and researcher perspectives

There are many ongoing conversations about the responsible application of GM insects and gene drive. Of the papers that have analyzed stakeholder perceptions of RI, almost all have

focused on established researchers and professionals. Thus, in this paper we investigate what early-career biotechnology scholars and researchers think about RI in the context of biotechnology innovation, particularly gene editing and gene drives for GPM, and the challenges of incorporating RI into their research in this early-stage of their careers.

Several studies have considered RI within academia, though only a few in the US. One UK study examined the implementation of RI practices within a synthetic biology multi-disciplinary research center (Pansera et al. 2020). One key insight is RI should be enacted as a creative process (as opposed to strict and bureaucratic) so that it can overcome barriers to RI such as disciplinary norms, incentive structures, institutional logics, lack of resources, leadership and agency. One stakeholder study found RI perceived as mostly irrelevant to nanotechnology and synthetic biology research managers, PIs and group leaders they interviewed from UK, Denmark, and US (Glerup, Davies, and Horst 2017). However, researchers also articulated and practiced a range of 'bottom-up' responsibilities, some which resonate with RI yet are shaped by academic capitalism. They call for RI scholarship to more actively address the political context of contemporary scientific research. In contrast, another study assessed US biotechnology stakeholder attitudes towards RI and found at the research stage both public and private sectors see the social science conception of RI as infeasible and potentially hazardous to careers, technologies, and ultimately society (Kuzma and Roberts 2018). They suggest RI scholars lower their expectations so collective efforts to help translate RI into practice are not immediately dismissed by innovators and regulators. Instead, they suggest refocusing the RI agenda on "continuing to develop effective tools, concepts and approaches that are tailored to the specific innovation system contexts – including levels, stages, vested interests, and institutional dynamics." The same group also assessed stakeholder attitudes and implications for research policy (Roberts, Herkert, & Kuzma 2020). They found industry, trade organizations, and academics had more negative reactions than government and advocacy groups to social science definitions of RI and to RI practices that relinquish control to people outside of technology

development pipelines. In short, all the examples above found differences among stakeholders in their attitudes and perspectives on RI.

Early career researcher perspectives

There have been many calls in academic literature for integration of RI into biotechnologies. Comparatively, analysis remains limited on how it might be articulated by early-career scientists and incorporated into their research and work. Our study aims to address questions about adoption and implementation of RI by graduate researchers, with an added focus on genetic engineering and GPM. This is the first study to ask US early-career grad students their "bottom-up" meanings of RI and use their responses to identify opportunities and challenges to achieving it.

Although the perspective of early career researchers has not been the focus of many studies, there are several examples where this demographic is included within the broader scope of a paper, or the importance of this demographic is highlighted. For example, Lukovics et al. (2019) consider the education of Generation Z university students and conclude that integration of RI concepts and practices at a pre-career stage, before researchers fully develop their daily routines, can strengthen the assimilation and long-term sustainability of RI principles. The importance of early-career exposure to RI and generational differences were also pointed out by Wickson and Carew (2014) who said a 'generational mix' was crucial to their workshop which asked researchers to consider what exemplary RI practice entails and criteria and parameters for evaluating it. Approximately one third of workshop participants were graduate students or early career researchers, though in an attempt to disrupt typical power dynamics of academic discourse they were not asked to reveal their position. Thus, the authors do not draw any conclusions that are specific to the viewpoints of early career researchers.

A few studies have made observations of early career researchers including one by Lacey, Coates, and Herington (2020) which assessed a variety of demographic factors and their association to the perception of transparency and openness with respect to RI and open science

in Australia. They found career stage to be significantly associated with two of twelve variables. First, more early-career stage participants (including students, young professors and professionals) disagreed that 'institutional arrangements in research delivery agencies/research funding agencies encourage openness and transparency from our scientists' than did mid or established career participants. Second, slightly more early- to mid-career participants disagreed that 'government funders effectively communicate decisions' than established career participants. Kuzma and Roberts (2018) conducted biotechnology stakeholder focus groups to probe attitudes towards RI and suggested a barrier to implementing RI in academe --graduate students who rely on certain streams of research cannot suddenly change course in a socially responsive manner without peril to their progress and perhaps their careers.

Thus far, McLeod, de Saille & Nerlich (2018) is the only publication to specifically assess early career researcher perspectives. The authors conducted workshops with PhD students, postdoctoral researchers, and technicians from a UK synthetic biology research center and found their interpretation of responsibility to be a difficult path to navigate between economic expectations, work-life realities, and difficulties of cutting-edge science. The majority of participants focused on "personal risk to their own mental health or career, societal risk in relying upon a technological "fix" or more ephemeral risks to science as the pursuit of knowledge." Perhaps the most troubling finding, as reported by the authors, was early career researchers questioning what responsibility they should, or even could, exert over the field given their lack of control over incentive systems, which in turn undermined their confidence and passion for science. They found multiple means of mitigation for these risks: collaboration and teamwork, communication, dissemination, public education and being open and honest about risks.

Recommendations given by many of these papers have a common thread in their recognition of barriers present in academic institutions and scientific disciplines, the need to adjust RI demands and expectations to fit the context, and the importance of flexibility, curiosity, and creativity. McLeod, de Saille & Nerlich (2018) call for more attention to "what science policy can

reasonably demand of the research workforce without crushing the curiosity and vitality of its postgraduate and postdoctoral participants, who are still in the apprenticeship phases of their education." This is similar to the insight from Pansera et al. (2020) that "for RI to overcome these barriers and succeed it must be enacted as a creative process that invites, creates space and adds value in a flexible way, rather than being perceived as a process of strict ethical oversight and a coercive, bureaucratic burden." These link back to the suggestions to address the political context of contemporary scientific research (Glerup, Davies, and Horst 2017) and lower RI expectations so collective efforts to help translate RI into practice are not immediately dismissed by innovators and regulators (Kuzma and Roberts 2018).

In this article, we explore what early career researchers, namely graduate students, think about RI and the feasibility of incorporating it into their research. Graduate researchers may have unique insights as frontline laboratory workers in biotechnology innovation. We study what graduate students in biotechnology think about responsible innovation (RI) as a framework, identify what grad students think is important, probe their attitudes about the appropriateness and feasibility of RI within systems of biotechnology research and innovation, and document the previously unheard student perspective in the US. We collect qualitative data on what graduate students think about RI in biotechnology innovation through ten semi-structured interviews.

Methods

Research approach and sampling

We use 10 semi-structured interviews and iterative thematic coding to see what bottom-up themes emerge from the data that best describe the perspectives of the graduate students towards applying RI to research in GPM. Ten graduate students from North Carolina State University were recruited to interviews from a concurrent research project that also explored meanings of responsible innovation: an NSF Cultivating Cultures of Ethics in Science, Technology, Engineering, and Math (CCESTEM) grant entitled "Comparing Meanings of Responsible Innovation across Bioengineering Communities." Interviewees played the role of

focus group moderators in the CCESTEM project (Roberts et al. 2020). As part of their roles as moderators, students completed a three-day training in focus group moderation, ethics, and responsible innovation (Herkert et al. 2017). Students were interviewed after the training workshop, before the focus groups took place.

Qualitative inquiry greatly differs from quantitative methods in sampling size, strategy, and purpose (Patton 2002). Where the purpose of quantitative inquiry is to yield empirical generalizations, the purpose of qualitative inquiry is to yield insights and in-depth understanding. Quantitative methods generally require random sampling, which is better achieved through high sample numbers. For quantitative inquiry, non-random sampling introduces bias and decreases the statistical significance of findings. This is not the case for qualitative research, which typically focuses on a relatively small number of samples that have been purposefully selected. Purposeful selection of samples allows for intended focus and becomes a strength of the study. It involves selecting information-rich cases from which the researcher can learn a great deal about the topic of inquiry (Patton 2002). There are multiple types of purposeful sampling. For these interviews we used a combination of intensity sampling and criterion sampling. The purpose of intensity sampling is to select information-rich cases that provide in-depth knowledge about the phenomenon of interest.

In this case, our phenomenon of interest is the process of graduate research, and the potential implementation of RI in the research process. Components of RI arguably require either researchers or research teams to be interdisciplinary. Therefore, graduate students who are already active within interdisciplinary programs are an information-rich resource. They have experiences and deep understanding of interdisciplinary processes that enable them to provide a great deal of information towards answering our questions on RI. See the following section ‘Student backgrounds’ for a breakdown of the interviewees.

Towards this same goal, we also used criterion sampling. The purpose of criterion sampling is to include cases that meet a predetermined criterion of importance, and this can be a

specific quality or event. In this study, our criterion was interdisciplinary students who completed a three-day training workshop on topics that included ethics and RI. The purpose of the workshop was to train students to be focus group moderators for a concurrent study assessing the meaning of responsible innovation across stakeholders involved in biotechnology research and innovation. Since this focus group study was happening concurrently to our interview research, it provided a mechanism for us to recruit students to interviews. During this workshop, students were assigned Stilgoe Owen & Macnaghten's paper (2013) on the RI framework, and the paper was the topic for a round table discussion.

The logic and power behind purposefully selecting students that had completed this workshop is threefold: First, students would have an at least a semi-detailed understanding of RI before the start of the interview and be more equipped to provide in depth, detailed answers to questions concerning the multiple components of RI and specifics on how the framework can be applied in the graduate research process. Second, students in this workshop are not RI subject matter experts. They have not researched RI for their own dissertation and therefore would have more raw answers to questions, as opposed to repeating what they learned from others during their own course of study. In other words, their answers will be generated from their own thoughts and experiences and may better represent the student perspective on the case of genetic pest management (GPM). During interviews, we use GPM as a case study to focus the conversation and illustrate a particular field of research for which RI might be applied. Third, half of the students in this workshop are subject matter experts on GPM. Including a mixture of experts and non-experts will ensure breadth and depth of the data. We make the assumption that GPM experts can give more detailed answers related to the technical aspects of technology, while non-experts are less hindered by the norms of GPM and can therefore give a greater and perhaps unique variety of answers.

Student backgrounds

Interviewed students were participants of interdisciplinary programs outside their home discipline, with five students being from the Initiative for Maximizing Student Development (IMSD) program and five from Integrative Graduate Education and Research Traineeship (IGERT) program. Students in the IGERT program and GES Center have been introduced to GPM through their course or laboratory work, whereas all IMSD students stated the training workshop and interviews were their first introduction to GPM. Though not used as a sampling criterion, students came from different academic disciplines. Disciplines included: genetics, forestry and environmental resources, entomology, communication, rhetoric and digital media, public administration, psychology, electrical engineering, chemical and biomolecular engineering, and computer science.

Data collection

Ten semi-structured interviews were conducted to probe graduate student attitudes and meanings of responsible innovation in the context of genetic pest management. Interviews used a list of predetermined but open-ended questions and retained some flexibility by following the semi-structured format (Table 5.1). Using this format, questions were read in order but the conversation was allowed to follow topical trajectories. Questions were guided by previous literature on the RI framework. We were interested in exploring the varying roles of researchers and public(s) and identifying congruence or differences in meanings of responsibility. Interview questions were designed to explore students' conceptions of each of the four components of responsible innovation: anticipation, inclusion, responsiveness, and reflexivity. Specifically, who and what should be included in considering impacts of genetic engineering and if, how, and when researchers should include the public. There were a total of 12 questions, with approximately five minutes allotted per question. These one-on-one interviews were each one hour long, except for two interviews that went three minutes over time.

The purpose of using open-ended questions is to capture and understand the points of view of the interviewees without predetermining what those might be before the start of the interview. This allows for a rawer response that better captures the deep thoughts, experiences, perceptions, and ways they have organized the world. Open-ended interviews, as opposed to closed-ended, can therefore allow interviewees to give a more thorough and accurate response. This ensures a higher level of data quality by protecting against preconceptions of the interviewer.

Table 5.1

Semi-structured interview questions

#	Interview Question
1	From your past experiences and anything you learned at the workshop, can you describe in as much detail as possible what RI means to you?
2	In the context of Genetic Pest management (GPM), what does responsible innovation mean to you?
3	What do you think the role of scientists is in the process of responsible innovation?
4	What do you think the role of 'the public' is in the development of GPM technologies? When do you think the public should be involved?
5	Do you consider yourself to be part of the public? If so how and if not, please explain.
6	Would you consider a GPM researcher to be a part of 'the public'? If so how? If not, please explain?
7	Please describe any ways in which your understanding of RI in the context of GPM has been shaped by any specific experiences you have in your academic background as a social/natural scientist?
8	How do you think this understanding differs or overlaps from the four dimensions of responsible innovation (anticipation, reflexivity, inclusion, responsiveness) that we covered in the workshop?
9	What are some ways that graduate students researching GPM can conduct their research responsibly, in the context of responsible innovation?
10	How do you think responsible innovation overlaps and differs between graduate students and professors?
11	GPM is a very broad category. In the workshop I described two GPM strategies: population suppression and population replacement. Can you describe a definition for RI in these contexts? Would two different definitions be required? If so, state two.
12	In addition to multiple strategies like suppression and replacement, GPM technologies can be used for a variety of different purposes. For example population replacement can be used to make pesticide resistant populations susceptible again, or to make plant disease vectors unable to carry disease. How do you think the purpose of a GPM program changes the definition of responsible innovation?

Data analysis

Interviews were audio-recorded, transcribed, and imported into NVivo 12 Pro (QSR 2018) software for analysis. NVivo was used for analyzing the interview transcriptions for themes and assigning codes. A codebook was established based on existing literature of RI, and was iteratively updated with new themes emerging from the data. Students spent more time and explained in more detail the challenges and opportunities to doing RI, rather than defining their ideas of what RI is. Thus, the codebook was adjusted to include these and organize them into micro, meso, and macro levels, as this classification schema was found to be a useful structure of analysis for previously classifying barriers of RI in biotechnology (Kuzma and Roberts 2018). We adapted definitions of these levels from Kuzma and Roberts to be challenges or opportunities that exist at the:

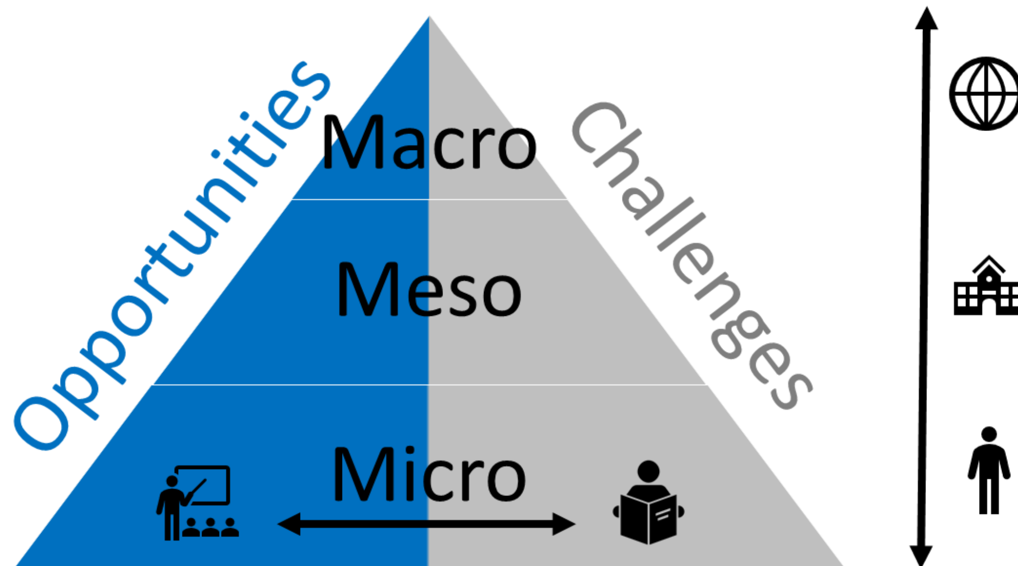
- Micro-level: individual researcher; graduate students or professors
- Meso-level: within organizational structures; department, center, college, university, etc.
- Macro-level: relating to the larger political, economic, cultural, social contexts

Once the final version of the codebook was established, the interview transcripts were assessed and coded to completion. Each observation (or “reference”) was one area of text, usually 2-3 sentences long that matched one theme in the codebook. Each theme was a particular issue at the micro, meso, or macro level, also deemed either a challenge or opportunity (Figure 5.1). To give an example, in one interview a student talked about their desire to conduct a focus group with local farmers to shape the direction of their research, but this idea was shot down by their advisor so this project never happened. According to the codebook (Supplementary Material 5.1), this would be coded as a micro-level issue under “authority” and “challenge.”

The micro-level is concerned with themes that exist at the level of individuals and each reference was categorized as professor or student, depending on the context. For example, if the context of a reference was “As a graduate student I experience [insert challenge],” then it would be categorized student, while a reference that was “I’ve noticed my advisor experiences [insert challenge],” then it would be categorized professor.

Figure 5.1

Schematic for the organization of themes and observations



Note. Each observation was coded as a particular theme (i.e. authority, funding, publication) within a level (micro, meso, macro). Each observation was also coded as either an opportunity or a challenge. The micro-level is the only one that was further divided into two parts: statements referring to professors or the graduate students themselves. See Table 5.3 for a full list of themes with abbreviated definitions. See the codebook (Supplementary Material 5.1) for complete definitions.

New themes identified by participants mostly characterized academic practices, culture, infrastructure, and funding that promote or inhibit responsible innovation. Students generally had positive views of RI as a possible way to bridge disciplines and explore societal and technical impacts. Thus, we focused our analysis on what impacts the feasibility of RI in academic genetic engineering research. We identify perceived challenges and opportunities of RI in academic research, at what level they exist, and propose how they can be overcome.

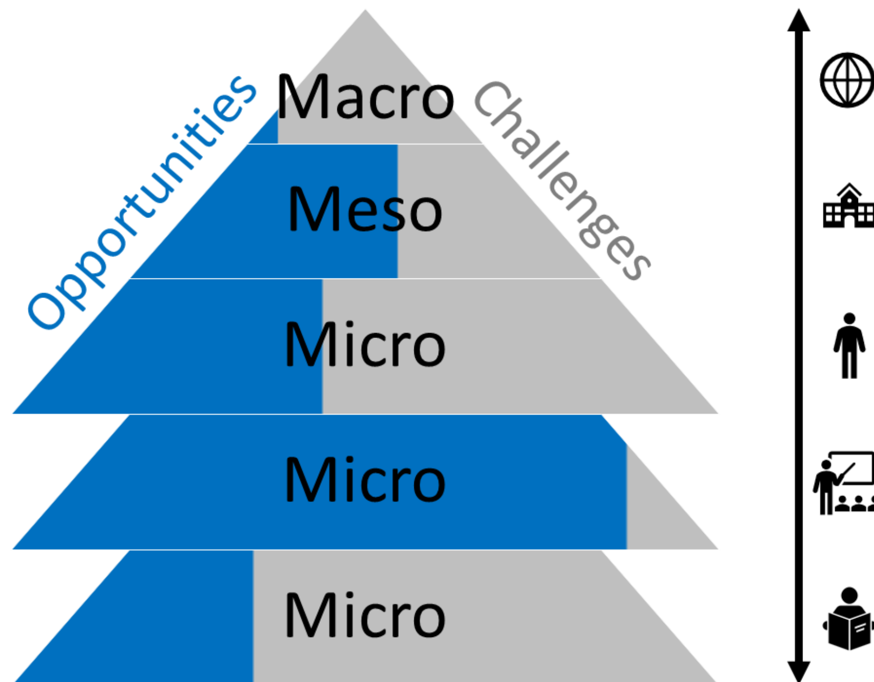
Results

Responsible innovation is a dynamic framework that can be used to guide discussions and decisions at multiple levels (Stilgoe, Owen, and Macnaghten 2013). Our graduate student interviews suggest RI still has untapped potential to shape graduate research practices. Students viewed RI positively with regards to the need for such a framework and its potential impacts, but also viewed the current framework as “highbrow theory” and forward-thinking. In order for RI principles to be performed in practice, students identified changes that need to be made to mitigate challenges and strengthen opportunities. For ease of analysis and discussion, themes (topics) were organized into micro, meso, and macro levels (from individual, to organizational, to larger political, economic, cultural, social contexts).

Of all the themes discussed by students, most occurred at the micro level (109), then meso (90), then macro (38) (Table 5.2). Each observation (or reference) was further categorized as either a challenge or opportunity. The ratio of challenges to opportunities varied between levels (Figure 5.2). The meso-level had the lowest ratio of challenges (43% of meso-level observations) compared to opportunities (57% of meso-level observations). The macro-level had the highest ratio of challenges (87%) and the micro-level the second highest (56%). When the micro-level was divided into statements that graduate students made primarily regarding barriers professors face and statements they made referring barriers to themselves graduate students. The interviewee comments suggested professors had a much lower ratio of challenges (13%) than students (67%) (Figure 5.2, Table 5.4). As a reminder, these are perceived differences identified and discussed by graduate students in our interviews. Though outside the scope of this chapter, it would be interesting to follow up with professors to see if they agree with the themes and differences in opportunities.

Figure 5.2

Proportion of opportunities versus challenges of RI at multiple levels, as perceived by early-career researchers



Note. Within each level, the percent of coding references that were opportunities (blue) and challenges (grey) are demonstrated. Macro-level: relating to the larger political, economic, cultural, social contexts, meso-level: within organizational structures (department, center, college, university, etc.), and micro-level: individual researchers. Micro-level is the only level that was further divided into two parts: professors and graduate students. The top micro level is the total of the two underneath. From bottom to top: micro-level students, micro-level professors, micro-level total.

Table 5.2

Number of observations at the macro, meso, and micro levels

Level	Opportunities	Challenges
Macro	5	33
Meso	51	39
Micro	48	61

Note. Comparison of the number of opportunities and challenges at each level. Macro-level: relating to the larger political, economic, cultural, social contexts, meso-level: within organizational structures (department, center, college, university, etc.), and micro-level: individual researchers.

Before going through each level, we would like to point out a few assumptions made by multiple students that may have guided their interpretation of the interview questions and the concept of RI as a whole. The first assumption, which could also be an opinion, is genetic engineering research currently has an insufficient level of emphasis on social impacts. This assumption likely contributed to their perception that RI, or similar approaches, is needed to facilitate socially conscious research practices. On a smaller scale, the perception of the immediate usefulness of RI was linked to the ability and agency of a student to incorporate RI into their research process. Most students spoke to the need for interdisciplinary training and collaborations, with the assumption that these kinds of knowledge and interactions are required for RI. If these challenges were addressed, most students held an optimistic view of the possibilities of RI to contribute to an individual's research program. Nevertheless, one student felt that RI was not a good fit for genetic engineering, as too many factors would need to be changed to make it work. The other nine students tended to focus on challenges when answering the questions, yet still came back to the assumption that RI is needed. See Table 5.3 for a summary of all the themes and the number of students that talked about each theme.

Table 5.3

Summary table of themes at all levels (micro, meso, macro)

Level	Theme	Definition	# of Students	# of References
Micro		Individual researcher: graduate students or professors.	10	109
Student		Micro-level themes experienced by graduate researchers.	10	86
	Authority	The power to give orders and make decisions.	9	20
	Independent mentor	A mentor that is partially or completely independent of the committee.	2	5
	Duty	The idea that students have the ‘onus’ or ‘responsibility’ to do RI.	7	17
	Skills or knowledge	Having/not having the skills or knowledge needed to do RRI.	8	37
	Time	Having/not having enough time in the workweek to dedicate to RI.	4	7
Professor		Micro-level themes experienced by professors.	9	23
	Access	Having the means to do RI. i.e. time, funding, network, collaborators	1	2
	Authority	The power to give orders and make decisions.	6	13
	Mentorship	Professional relationship where one is guided by a more experienced scholar.	6	8

Note. For each theme: an abbreviated definition, the number of students that mentioned it (# students), and the number of observations or times each theme was mentioned in all interviews (# references). See Supplementary Table 5.1 for complete definitions.

Table 5.3 *(Continued)*

Level	Theme	Definition	# of Students	# of References
Meso		Within organizational structures: department, center, college, university, etc.	10	90
	Collaboration	Working together on ideas, projects, or otherwise forming a community	6	12
	Extension programs	Public service of land grant university.	2	3
	Funding	Monetary resources for: research, tuition, training workshops, public engagement, etc.	6	11
	Lab focus and inheritance	Existing grants and therefore vision and accountability	2	4
	Public engagement	Event(s) with bidirectional information flow between researchers and public.	9	28
	Public outreach	Event(s) where scientists share their research or general science knowledge with the public.	6	19
	Training opportunities	Professional development or educational courses, workshops, or events.	6	13
Macro		Relating to the larger political, economic, cultural, social contexts.	10	38
	Mechanics or practicality	Difficulties and lack of a clear framework to follow for the inclusion component.	9	16
	Publication	A published journal article or being an author on a paper.	5	6
	Risk and values	Standard narrow framing of risk does not typically include human values or societal issues.	6	16

Note. For each theme: an abbreviated definition, the number of students that mentioned it (# students), and the number of observations or times each theme was mentioned in all interviews (# references). See Supplementary Table 5.1 for complete definitions

In the following sections we move through each level (micro, meso, macro) providing an overview of interview responses, as well as detailed examples of the challenges and opportunities. From hundreds of units of observation, we give direct quotations of representative quotes from the themes. The first quote selection goal was to relay common thoughts that occurred across multiple interviews. The second goal was to give voice to unique points that may be thought provoking for the reader, introduce a new perspective, or otherwise add variety to the discussion.

As argued by multiple students, if RI is to be incorporated by early-career researchers, existing opportunities should be fortified and challenges addressed. In some references, students talk about more than one theme and draw connections between multiple challenges, or between challenges and opportunities. Some challenges were seen by the students as immovable, like a major advisor discouraging activities that are not directly related to their dissertation or research. Other challenges are perhaps more flexible, with multiple possible strategies at different levels to overcome them.

Micro-level challenges and opportunities to RI

Of all the levels, the micro-level had the highest number of references (Table 5.2, n=109) and the second highest challenge to opportunity ratio, with 56% of references categorized as a challenge (Table 5.4). Micro-level challenges mentioned by students were skills or knowledge, authority, and time, while opportunities included duty, authority, mentorship, independent mentor, skills or knowledge, access, and time (Table 5.4).

The delineation made by student interviewees between students and professors in their ability to pursue RI was interesting to find during our analysis because our interview questions did not prompt this discussion, with the exception of question 10 (of 12) (Table 5.1). However, what quickly became apparent from the beginning of the interview is that students perceive differences between professors and students in terms of their roles and abilities with regard to responsible innovation. For example, Table 5.4 shows students mentioned authority more times as an

opportunity for professors (10 references) to implement RI than for students for whom authority (or lack thereof) is more often mentioned as a challenge to implementing RI (17 references). Overall, mentions of professors had a much lower ratio of challenges to opportunities (13% challenges) than students referring to themselves (67% challenges). The content of these will be discussed in more detail in the subsections below.

Table 5.4

Comparison of micro-level opportunities and challenges to RI

Theme	# of Students	# of References	Opportunity	Challenge
Authority	9	20	3	17
Independent mentor	2	5	5	0
Duty	7	17	17	0
Skills or knowledge	8	37	2	35
Time	4	7	1	6
Total Student	10	86	28	58
Access	1	2	2	0
Authority	6	13	10	3
Mentorship	6	8	8	0
Total Professor	9	23	20	3

Note. For each theme: the number of students that mentioned it (# students), the number of observations or times it was mentioned in all interviews (# references), and the number of references that were categorized as either an opportunity or challenge. See Table 5.3 for theme definitions. Under # of students, total student and total professor shows the total number of students that mentioned these issues. For the remaining columns, the totals are simply the sums in that category.

Student: Skills or Knowledge

Overall, when asked about their conception of each component of RI, students spent the most time talking about challenges at the micro level. Out of all micro level challenges, the theme skills and knowledge was the most referenced (n=37, Table 5.4). Eight students talked about various skills or knowledge needed to use RI and there were 37 total mentions, with 35 of those related to challenges to RI and 2 to opportunities. For example, skills in science communication and working in interdisciplinary research groups, and knowledge in ethics and methods of public engagement were mentioned as themes in this area. In most cases, students spoke about their lack of skills or knowledge as a challenge to incorporating RI into the research process. For

disciplines like genetic engineering that do not have a code of ethics or code of best practices, the concepts of social impacts and accountability may not be familiar. As a first step, they suggested that there needs to be increased awareness and understanding of responsible innovation, even as the definitions of the term continue to evolve. Of greatest concern to students was their lack of knowledge of ethics and social science, and the accompanying lack of skills to carry out necessary components of RI. One student expressed:

> "I think one of the biggest ones is probably reflexivity –thinking about your own assumptions and where you are coming from and taking the time to worry about that. Some of it is learning the terminology, or even learning more about social science to be able to reflect on your own assumptions."

One student expressed the need for education in things like environmental justice, to increase awareness and ability to identify and express concerns they would have been blind to otherwise. Without this knowledge, they felt they could not anticipate social types of risk, or challenge their own assumptions or framing of the issue as required to be 'reflexive'. Another student similarly expressed the need for a deeper understanding of multiple types of ethics, more than could be gained by the two-week online training module required by their program. Many students felt they had trouble contextualizing their research into society, especially for basic research projects as opposed to applied research. One student linked their inability to use RI to their inability to answer the questions: innovation for whom, and responsible to whom? If social impacts are going to be explored as part of the research process, then they need to understand and be able to predict potential cost-benefit structures of the technology.

Skills sought after by the interviewed students include both technical and professional soft skills that would enable them to do specific tasks required by some of the four RI components: anticipation, reflexivity, inclusion, and responsiveness. Students described the need for formal

training that would increase professional soft skills like communication and working on teams. Many felt they would benefit from training in communication, because it would increase their ability to engage with the public(s) and work on interdisciplinary teams. One student expressed difficulty over the idea of applying 'inclusion' to their research, because they struggle to effectively explain their research to a broad audience. Another student expressed their desire to learn how to have respectful, meaningful deliberation with people with different views or values. Communication skills may therefore be critical to enabling RI for graduate students researching contentious topics like genetic engineering. One student views interdisciplinary collaboration to be important for RI, but as a bench scientist thinks they haven't been trained for this collaboration. This was reflected by a social science student who thinks that natural scientists are expected to be experts on how they feel about the technology, but in many ways are not currently trained or capable of doing so. On the other hand, this student, as a social scientist, thinks that natural scientists are in a unique position to contribute to the conversation because of their deep technical knowledge, sometimes undervalued by social scientists. In fact, this unique ability or position of genetic engineers is the biggest perceived opportunity of RI for genetic engineers.

Student: Duty

Duty was the most coded opportunity at the micro level for students (n=17, Table 5.4), and was talked about by seven students. We defined duty as the onus or responsibility of researchers to responsibly innovate, especially as required by one's unique knowledge or position. The frequency of duty in responses is perhaps explained by the common view across interviews that genetic engineers should integrate responsible innovation into their research and not leave it to an ethicist or others to work on separately.

As identified by interviewees, a common reason for this belief is because genetic engineers' unique knowledge put them in a position to contribute vital information to responsible innovation analysis. One social science student described the drawbacks to simply having an 'ethics consultant' on a project: Because the consultant's expertise is different, they will inevitably

miss details pertaining to technical parts of the technology, which could contribute to a social issue down the road. Thus, this student suggests there is a somewhat misplaced authority given to social scientists by genetic engineers. Instead, they explained RI would need to be a group effort to have the most thorough analysis. They expressed confidence in the ability of genetic engineers to contribute:

> “I also think it's easy being in graduate school to feel like you have no power at all because you are so low on the totem pole, but the expertise that we're developing, and especially for my end, the expertise that people in say genetic pest management are developing - that's really powerful knowledge. Knowledge that has a tremendous amount of influence and not very many people in the whole world know what this is, and how it works, and how to do it, and what if something goes wrong, and what might go wrong, at least from the genetics perspective. That's no small thing.”

In other words, with their expert knowledge on the subject, genetic engineers can greatly contribute to the ‘anticipation’ of social impacts. The four elements of RI (anticipation, reflexivity, inclusion, and responsiveness) relate to duty as an opportunity in different ways and strengths. For example, the previous two quotes showcase how reflexivity and anticipation relate to duty, through a researcher's role in framing the research problem and analyzing potential impacts. Responsiveness relates to duty through the idea of stewardship.

Possibly one of the most important factors contributing to the optimistic view of the RI framework, was the view that the onus is on genetic engineers, and students more specifically, to be stewards of the technology. Students feel their unique position as the labor force gives them power to influence the trajectory of the technology, and therefore it is their duty or obligation to take an active role in responsible innovation. Many expressed the view that because students are

the ones on the ground doing the work, they are uniquely positioned to be actively thinking about possible futures and consequences. As one student describes it:

> "I think [graduate students] have at least more of an ability to say I disagree or I think this is wrong. Or they're actually doing more of the hands-on work. How much time does an actual PI get to actually spend on what's going on, or actually see what could be an issue or could be a problem? While they're doing the best they can, they're looking at many projects…So while [PI's] may have power in agency I don't know if they actually are able to see what needs to be seen."

Many pointed to this as being one area where students are better positioned to responsibly innovate than professors. Multiple students echoed that although professors also have a deep understanding of the technology, students spend more time "in the trenches" or "actually doing the work" and so can give richer, deeper thoughts on the intricacies of the technology. Proximity to a fewer number of projects also means a keener focus. In summary, hands-on experience and focused dedication most describe graduate students' ability and duty to responsibly innovate.

Student: Authority

The second largest micro-level challenge, authority, (n=17, Table 3) was present across many themes and answers. We defined authority as the power to give orders and make decisions. Authority, or lack thereof, was tightly linked across themes. Power dynamics were prevalent in many answers, as evidenced by the previous two quotes in the duty section. The influence major advisors have on shaping their graduate students' careers was especially highlighted.

Nine students expressed the opinion that without their major advisors' openness to societal implications of genetic engineering and broader types of concerns, they would not be able to incorporate responsible innovation into their graduate career. For one student, lack of authority was an express concern that they felt made RI a non-starter:

> "If you had a PI who is totally not interested in anything to do with what the public had to say, could you initiate public engagement on your own without undermining your relationship with your PI?"

A second student described the importance of having an open and supportive relationship with their mentor, but also describes how this support is still not enough:

> "For me I have a very good relationship with my PI, that I have no problems saying, hey, I think this is the problem, we should be really concerned… the problem is he has so many issues going on at the same time that he needs to prioritize whether my concern is one of his top concerns."

This student said their research practice has in many ways followed the principles of responsible innovation and explains the difficulties in doing so. Prelims were an opportunity to express concerns and ideas about the trajectory of their project to their committee. On one hand they said the prelim process gives weight to students' concerns, and special attention to their ideas for future direction. On the other hand, this was extremely stressful and they were worried they might not pass as a result of their objection to move forward on a certain aspect of the project. With a bit of courage and confidence in their skills and mentor relationship, this student was able to make responsible innovation a priority. A third student who followed multiple principles of RI during their academic career also cited support from their mentor as a reason for success.

In addition to authority being talked about in the context of PI and graduate student relationships, it was also talked about in the context of graduate school. On a broad level, rigidity of the dissertation process may stifle the ability of students to be responsive, especially once classes and dissertation chapters are agreed upon by the committee. One student pointed out graduate researchers experience varying levels of control over their projects. They may have a

lot or very little control on aspects ranging from decisions on research topics to the minute details in design.

The previous two quotes point to relationships, support, and a fear of negative consequences related to finishing their dissertation. These results are interesting because again, the questions originally sought to explore what students thought the RI framework would look like for genetic engineering, and the role of scientists and the public(s). Issues of power and authority were not specifically targeted by the questions, besides implied perhaps in question 9 (Table 5.1). In relation to lack of authority, students also pointed to a lack of resources, namely time and money (see below). Funding was talked about as it pertains to students, professors, centers, agencies, and research in general, and can be found later in the Meso section.

Student: Time

Related to authority was the issue of time (n=6, 3rd most cited challenge in Table 5.4). In this case, we defined time as participants' ability to dedicate some time to RI versus being unable to do so. Of the four students who talked about time, one described it as a challenge of their own choosing. The technology they work on is so new that it likely won’t have applications for ten or twenty years, and they feel “from a practical standpoint” that their time is better spent completing research and getting their degree. They also feel overwhelmed when trying to anticipate implications of the technology that far in the future. A second student also brings up time in the context of a technology development timeline, and says that at any stage most of a scientist's time goes into developing the application. However, this second student also talks about how the possible challenge of time can be overcome through collaboration. They describe collaboration as a way to share responsibilities and therefore decrease time required for RI.

As described in the interviews, students are accountable to their advisors, committee, and funding agency. In many cases there was an overlap between time and funding. Depending on the agreement between the professor and funding agency, there may be little room for flexibility if concepts of responsible innovation were not already written into the grant proposal. This

constraint on time is a concern that may be more noticeable in the context of inheritance of a research project or narrow lab focus, as noted by one student:

> "If [a graduate student] were to say 'Hey I want to go talk to stakeholders in the field this summer instead of doing my own replicated research,' that wouldn't go over so well. Or they might get told 'That's a really nice thought but I don't have money to do that.'"

This sentiment was echoed in other responses that also describe the obligation and "percent effort" of a researcher's obligations.

In summary, time is mostly viewed as a challenge due to the nature of graduate school to focus on research products more than process. It is also often linked to the meso level issues of funding and inheritance. Students operate on a shorter time scale than professors and may only be active for one period of a project. These conditions unavoidably create a lack of continuity between researchers and also mean that some students start their graduate career by inheriting an ongoing project. In some cases, this short time scale may be coupled with basic research, many years from any application. In this case, students may feel overwhelmed and paralyzed when trying to anticipate implications of their work. Thus, changing the incentive structures within which professors and students operate is one way to promote the incorporation of RI into research agendas and will be discussed in the "Funding" section.

Student: Independent Mentor

Besides duty, independent mentor was the only other theme at the micro level that was exclusively discussed as an opportunity (n=5, Table 5.4). This theme defined an independent mentor as an adviser that is partially or completely independent of a student's committee. Two students talked about multiple experiences where having an independent mentor increased their authority to innovate responsibly. One student talked about the confidence and power they get from having someone they can turn to whenever a problem arises within their committee. They

also talked about their positive experience working in a research community where a well trusted person in a leadership position had taken an active role in being a “watchdog” that could be alerted to help solve any issues. A second student had a suggestion for new students to select at least one committee member that is interested in RI activities or questions. They alternatively suggested creating a mentorship relationship with someone outside the committee to provide additional support. In summary, having an independent mentor can increase a students’ authority to do RI. It is therefore a way to overcome the second most referenced challenge for students.

Professor: Authority

When the micro-level was divided into professors and graduate students, professors had a much lower ratio of challenges (13%) than students (67%) (Figure 5.2, Table 5.4). While authority was the second most coded student challenge, it was the most coded opportunity when students referred to the authority of professors (n=10, Table 5.4). Six students identified the authority of professors as an opportunity to responsibly innovate. Most references talk about the authority professors have to incorporate RI components into the research agenda of their lab. Since they are “in charge,” it is “ultimately their responsibility” to make sure the components of RI are explored and addressed. Their viewpoint is represented by the following quote:

> “Professors have more autonomy, and they're more decision makers, and have more influence on the direction of the project, I would say. So, I think that those kinds of decisions — the decisions made about how exactly research is conducted, and what research is conducted, is more squarely on their shoulders.”

In addition to this power of professors to make decisions about a research project, two students also pointed out the authority of professors to encourage or delegate RI tasks to their students. Multiple students listed having a mentor open to inclusion and responsiveness as being a major opportunity to carry out RI. Support from the major advisor was described in different

ways: positivity or buy-in to the concepts, acceptance of time and energy spent pursuing questions outside of the lab, funding for interdisciplinary competitions, and aid with professional networking that inspired and increased access to potential collaborators.

One student introduced a different point that highlights the complexity of the systems in which research takes place. As they point out, professors operate under an incentive structure that guides their research agenda, which can in turn greatly affect their students' ability to responsibly innovate. They state:

> "I mean there's a power differential, but the funny thing is like most of the professors I talk to, a lot of them also feel they don't like all of a sudden you get a professor position and then suddenly feel like really empowered, right. They're feeling a lot of pressure to get grants and to get tenure and that's going to encourage them to do certain kinds of work and spend their time in certain kinds of ways. I think sometimes the people who make the incentives that we're putting experts under like graduate students and professors, they have unintended consequences too, right?"

In addition to the above quote, there were two other references that pointed out possible challenges of authority. One explained that on different projects, professors may have widely varying degrees of control. So they may have the authority to include RI in some cases and not in others. The other explained that professors may have less control in applied research projects where the university patent office is involved. Or, even if they have control in the research phase, they might not have control over how the technology is deployed.

In summary, the perceived level of authority varied greatly between students and professors. The interviewed students presented authority as a challenge to a student's ability to use the RI framework eighty five percent of the time. For professors, authority was only presented as a challenge twenty three percent of the time (Table 5.4). However, student

authority was talked about by more students and for a greater length of time than that of professors. For student authority there were 20 references from nine interviews, and for professor authority there were 13 references from six interviews (see Table 5.4). Overall, students reported that professors have more of a responsibility to make sure RI is happening in their labs, because of their authority to write RI into grants, and encourage participation in outside training opportunities and public engagement type activities.

Professor: Mentorship

The second opportunity for RI that students mentioned which focused on professors was mentorship (n=8, Table 5.4). Mentorship is a professional relationship where one is guided by a more experienced scholar, in this case between a professor and graduate student. Six students talked about mentorship as a possible opportunity for researchers to do RI. As evidenced by the interviews, professors can be a mentor by doing the following: use RI themselves, encourage communication on relevant topics, set laboratory standards, and provide feedback on the RI process. By doing these things, professors are in the words of one interviewee “training their grad students to be responsible innovators in the future.” Students talked about mentorship as an opportunity to start using RI processes, and felt RI would help them be successful in their careers after graduate school. One student talked about how their advisor is on an ethics review board which has positively impacted the knowledge and expectations of their lab. The advisor uses lab meetings as a platform to discuss ethics journal articles and recent experiences from sitting on a research ethics review panel. Due to their professor’s experiences, the student says they are held to higher standards than some other academic labs. On one hand, their everyday processes take more time, but on the other hand, they take pride in the ethical level of their work and also feel more prepared for a job in industry.

As another attribute of good mentorship, the importance of open communication was explained by multiple participants. They said researchers should practice reflexivity and be conscientious in their research process, then communicate any thoughts or concerns they have.

In this way, they can make sure "nothing slips through the cracks." On the other side of that communication line, professors need to be open to the conversation and invested in addressing any issues. One student gave an example of speaking with a PI about a RI related question:

> "I think anyone working, at least people [in the GES Center], have the kind of PIs that if the grad student came to them and said 'how is this going to land in the field?' [their PI] would be open [to that question]."

As discussed previously, lack of authority is a major challenge to students, so students are enabled by mentors who are open to discussing and possibly researching RI types of questions. This quote also points to the theme collaboration, which will be discussed in the meso section. Collaboration and interdisciplinary spaces like the GES Center can be used by students to increase their authority.

Overall, multiple students posited that dynamics of a mentorship and level of communication "depend on the relationship" between a student and their PI. Mentorship dynamics and a PI's level of buy-in to RI concepts greatly affect a student's ability to innovate responsibly. Most students discussed this in a positive way by referencing experiences where they felt encouraged by the support they were receiving from their mentor.

Professor: Access

Related to authority, access is another theme identified by one student as a possible opportunity for professors to do RI (n=2, Table 5.4). Here we have defined access as having the means (i.e. time, funding, network, collaborators) to do RI. It is interesting that students saw access as an opportunity for professors since two of the components, time and funding, were identified as challenges for students. Though they used the term access, they mostly focused on networking or collaboration aspects. They also mention how the culture, or perhaps values, of a university can affect a professor's ability to collaborate:

> “Especially somewhere like here at [North Carolina] State [University] where there's already a big push for interdisciplinarity… there's a lot of professors that I think would be willing to work with them… so I think professors need to be asking these questions more and not just relying on grad students or outside reviewers to push them into asking those questions.”

This student goes on to talk about the greater access professors have to collaborators as compared to students. They can use their network to find grant collaborators from across departments. On a less formal level, they could ask for feedback on project design when trying to incorporate RI aspects into a new grant application.

To conclude, students perceived more opportunities for professors than for themselves as students at the micro-level. For students, the micro-level challenges far outweighed opportunities. In the next section we turn to the meso-level opportunities and challenges. Potential solutions to the already-discussed micro-level challenges may exist at the meso and macro-level.

Meso-level challenges and opportunities to RI

The meso-level is concerned with themes that exist within organizational structures (department, center, college, university, etc.). Of all the levels, the meso-level had the second highest number of references (Table 5.2, n=90) and the highest opportunity to challenge ratio, with 57% of references categorized as an opportunity (Table 5.5). Meso-level challenges mentioned by students were public engagement, funding, lab focus and inheritance, training opportunities, public outreach, and collaboration, while opportunities included Public outreach, collaboration, training opportunities, public engagement, and extension programs (Table 5.5). The most predominant opportunity for RI that was mentioned was “public outreach” (n=18) whereas the most predominant challenge mentioned was “public engagement” (n=19). We discuss these and other predominant meso-level themes below in each subsection.

Table 5.5

Comparison of meso-level opportunities and challenges to RI

Theme	# of Students	# of References	Opportunity	Challenge
Collaboration	6	12	11	1
Extension programs	2	3	3	0
Funding	6	11	0	11
Lab focus and inheritance	2	4	0	4
Public engagement	9	28	9	19
Public outreach	6	19	18	1
Training opportunities	6	13	10	3
Total	10	90	51	39

Collaboration

Collaboration was the second most prevalent theme with regard to opportunities for RI (n=11, 6 students mentioned, Table 5.5). We categorize collaboration as a meso-level issue as it is highly influenced by the organizational context in which individuals find themselves (norms and training with respect to collaboration that exist within and between institutions, as well as incentives and disincentives for collaboration that occur within and between institutions) and some communities are more likely to collaborate than others. Collaboration as a theme in our interviews reflects a transition from the micro-level to the meso-level as individual abilities and incentives for collaboration are affected by the climate and structure of their institutions. One student viewed collaboration as a vital part of responsible innovation. They reassert the need for professors to take an active role in collaborations:

> "I think professors need to be actively pursuing these opportunities, and not just waiting for them to fall into their laps."

Another student describes the importance of collaborations, specifically between natural and social scientists. They point out the existence of "ethical issues even in the technical work"

and explain that social scientists who "don't have deep, deep, deep understandings of genetics wouldn't even know about [these ethical issues]." They reflect on the expertise gaps of individual researchers from their perspective as a social scientist:

> "My ability to answer those questions [as a social scientist] pales in comparison to yours [as a geneticist], I think. I think it's valuable to have both, but there are things that can happen at the technical level that I think are a really big deal."

For example, they describe some of the limitations or pitfalls of using mice as models for human medicine and point out that they would not have a deep enough understanding of these intricacies to be able to analyze the accompanying ethical issues on their own. This is another good example of the duty of graduate students to responsibly innovate, covered in the micro section.

This student also highlights how collaborations can be used by graduate students to increase authority. They talk about an experience of their fellow graduate student collaborators who met with government regulators to slow the process of some field trials related to one of their labs. Without any prompting from their advisors or departments, a team of two geneticists, a biomathematical modeler, a communication scholar, and a wildlife management specialist developed a list of concerns to discuss with regulators. The team didn't want to stop the trials, but had some information they felt necessitated slowing down the process. Originally the meeting was only going to be with the geneticists who were more directly involved in the field trials, but they pursued a collaborative effort to bring up viewpoints within other fields of study. Specifically, a modeler and environmental specialist were invited to contribute their expertise. The interviewee explains that through their collaborative efforts, they "knew some things to think about that the other people in that room hadn't been." Despite the advantages of working as a team, it was still difficult to create these talking points "without sounding like you're throwing your PI under the bus [*sic*] or talking as if the government people don't know what they're talking about." The interviewee

said this is a communication issue of how to "thread the needle" between sharing technical information to make an argument, holding your ground to state what you believe, and keeping the conversation respectful and understandable to a lay person. There was the added challenge of making arguments in this context where some of the team members and their PIs were going to be involved with the field trials. Despite this challenge, one of the collaborators told the interviewee they were deliberate with "not wanting to put things so delicately that she isn't true to herself." Following the meeting the regulators decided to slow down the field trials, and the team attributed their success to their collaboration and communication.

This importance of interdisciplinary collaborations in thinking about problems from multiple perspectives and increasing authority was echoed by another interviewee. In their example, the student talks about the Genetic Engineering and Society (GES) Center at NCSU, which serves as a hub of interdisciplinary research and inclusive dialogue around genetic engineering. Speaking in the context of graduate students who don't have "a ton of autonomy or power," they state:

> "The GES center... understands the importance of interdisciplinarity and multiple values being held at the same time. So I think there is space for that to exist… raising questions and figuring out ways to check [and] understand what your assumptions are, and then... [comparing them] against what is going on and where those particular technologies might be applied later."

While most interviewees talked about collaborations between academics, one student talked about collaborations between academic researchers and nongovernmental organizations (NGOs). For applied researchers, they say that it may be beneficial to collaborate with NGOs since they spend more time "consider[ing] these wider impacts." This is beneficial to a natural

scientist since most of their “time, energy, and effort go into [the] actual application.” They further describe how collaboration can be an opportunity:

> “So, it's not always the easiest thing to do, to think big and think about responsible innovation, which is a hard term to come to grips with. One, obviously, is for the scientist to be cognizant that there are bigger issues, and two, that they don't have to go in it alone. That they can collaborate to consider wider impacts and actually integrate them into the scientific processes.”

The sentiment of integrating RI type questions into the research process was present across all six interviews that mentioned collaboration. The students seemed to challenge the idea that responsible innovation involves an ethicist or social scientist researching seperatly in parallel. The type of collaboration described by all the students seemed to be a deeper collaboration, with research being more integrated and interactions being frequent and sustained over time. This matches well with the RI vision suggested in the most-cited paper on RI, which portrays an iterative, upstream collaboration among disciplines (Stilgoe et al. 2013). This integration occurs at the meso or institutional level, or even across institutions. A tall order, indeed, but the importance of which was highlighted through the success of the collaborative graduate students who met with government regulators and successfully had influence over field trials.

Lab focus and inheritance

The issue of lab research focus and project inheritance was mentioned by two students (n=4, Table 5.5). In the natural sciences, a professor is expected to develop their own research program, complete with vision and funding. This means that the current and future research of graduate students in their lab has to some degree been designed to fit into this research program, and ideally, funding has been secured for future projects. A major part of job talks during the interview process for a professorship is a candidate’s research program for the next five years

and existing or plans for funding. Funding can therefore exist beyond the graduation of a student, and sometimes students inherit research projects that are a continuation of research in the lab. Students identified these inherited projects as a potential challenge to RI since “most graduate students are constrained by what their lab is doing.” One student describes the limitations stemming from inheritance of a research project:

> “I feel like as a graduate student you can only go so far in terms of innovating responsibly when you don’t have the flexibility to do so. It depends on how much control you have of designing your own research. A lot of grad students I think of that work in GPM for pests… can’t necessarily do that on their own. Maybe in designing their own question, they are coming to a project that was already determined by their PI.”

Inheritance could therefore be a contributing factor to the different levels of authority between professors and students covered in the micro section. Lab focus and inheritance connects to issues of authority previously discussed, such as the feeling of being locked in to a research trajectory because of the structure and timeline of graduate school.

Funding

As a related theme, six students talked about funding as a potential challenge to RI (n=11, Table 5.5). One interviewee stated that students are constrained by the source and type of funding, and went on to say that professors are also “constrained, even with tenure, to some extent by funding.” They explain that funding agencies like the NSF and NIH ask for the significance and broader or overall impact of research, but impact is not defined using components of the RI framework (Anticipation, Reflexivity, Inclusion, Responsiveness). The NSF and NIH are looking more for the technical impact the research will have on a scientific field, or the impact of potential products or interventions in society.

> “A lot of the funding doesn't directly go into doing that social science research. So the professor isn't really thinking about this, but from a real standpoint they also need to find ways that they can do it within their own funding constraints.”

On the one hand, this research was funded by an NSF grant. One the other hand, the students' perspective points to the general dearth of funding for STEM-related social science projects. As another student points out:

> “The problem is public engagement can be very expensive, at least in the sort of dialogue or engagement processes we've seen happen. Those were pretty expensive in the context of social science at least to put on.”

Another student brought up funding and related it back to incentive structures:

> “I'm really interested in the incentive systems in which people are doing their work, and those are big right? That's not just your lab but your university, and the governance system, and the economics of the United States… I think it's really easy to be like well you're irresponsible, but I also think it's important to ask ‘What are we asking of people with technical expertise?’ ‘What are we encouraging them to spend their time doing, and what are we *not* encouraging them to spend their time doing?’”

In addition to federal funding agencies, two students spoke to some potential issues with academic labs obtaining funding from private industry. Their sentiment is embodied in this quote:

> “I see that there could be a conflict of interest and issues might arise out of signing a confidentiality agreement… because you are kind of working under a public house that

> uses public funds, even if that isn't necessarily the bulk of what pays for the research. So, any work you do should be made public at some point."

Overall, the students we interviewed pointed out how funding can change the focus of researchers and influence how much time they spend collaborating or doing RI related activities like public engagement. They also pointed out possible constraints of funding based on different types of funding organizations.

Training Opportunities

Six students talked about training opportunities as an opportunity to doing RI, and in three cases they additionally described how some training opportunities can act as a challenge (n=11, Table 5.5). One student described the inadequacy of the online ethics training module required by their department. They say that researchers treat it "like a box to check off" rather than a learning opportunity, and there are "no real consequences" attached. Another student said there is not a required ethics course for their program, except for a two-hour lecture when they first enter the department. They know other departments that offer ethics classes, but since theirs did not offer one they took a course in the philosophy department. This interviewee's experience relates to some thoughts shared by another student on the possible ways training opportunities incentivize RI. They explain that while some students might have trouble finding RI related training opportunities, students involved with interdisciplinary programs or communities might have more access or be more aware of opportunities. They explain:

> "For example, with the IGERT [interdisciplinary training fellowship] I am able to [have training opportunities], we are doing the CCE training [three-day RI workshop] but that is not an opportunity that I created for myself. I am taking advantage of what's there."

This example mirrors a point made in the collaboration section that interdisciplinary centers or communities provide researchers with increased access to collaborations, and as pointed out now, training opportunities. Another student described the importance of an interdisciplinary training program:

> "…where you start very early on, not just think about the science, but about wider societal impacts, cultural impacts, other environmental impacts, I think that's the start of [anticipation]… even at an easy level, like a colloquium or seminar series, or grad student group, where you just talk with other people about what the implications could be."

In other words, having some sort of structural space where researchers can meet and discuss RI topics is a key part of actually doing RI. This process of deliberation aides in the anticipation process, and the structure of the group can take many forms. Another student mentioned a program through which they took an ethics course:

> "I think that to do things responsibly everyone should have an ethics course. Everyone should be required, and not this little module that you take online for a Federal requirement but an actual ethics course. The bio-ethics course I took with the BIT program was really phenomenal and it really forced me to reflect on what I was doing in a different way."

Though most students viewed ethics courses as being a mechanism to learn skills relevant to RI, one student disagreed, saying ethics is too theoretical and therefore not very practical. They explain their point further:

> "I also have a lot of issues with classical ethics in the sense of it was all established by white men in the 1800s… I mean I think there's a lot of normativity in ethics that gets ignored, because ethics is seen as being pure and non-biased a lot."

In summary, multiple students said they have looked for training opportunities such as deliberative seminars, workshops, or ethics classes to develop knowledge that would enable them to responsibly innovate. Half of the students reported either a lack of course offerings or inadequacy of current ones. The perception of the usefulness of ethics courses varied between students, with one student arguing that ethics itself is not an appropriate tool for RI. However, all six students saw training opportunities as an opportunity to RI.

Public Outreach

Six students talked about public outreach as an opportunity or in other words a venue to execute RI at the meso-level, occurring across institutions and expert and non-expert communities. Incentives or disincentives for public outreach would exist at organizational levels (n=18, Table 5.5). Public outreach is when scientists share their research or general science knowledge with the public for the purpose of educating the public. This can take many forms, with the following being mentioned in the interviews: events at the science museum, materials for K-12 classes, open house at the college, citizen science, public lectures, Twitter lab account, Reddit AMA, and popular science books. Public outreach is typically used as a tool for education so is not usually included in RI literature. However, many students described how public outreach is an academic norm they could repurpose to meet some goals of RI like transparency and strengthen RI related skills like science communication.

An additional goal students listed that is not typical for public outreach is transparency. Multiple students said it was the responsibility of researchers to maintain transparency between themselves and the members of the public. Transparency came up multiple times and was talked about in a broad context, with students saying that researchers should convey "what a technology is doing or the potentials for this technology" and to "not do anything hidden." One student gave a specific example; They stressed that researchers have a responsibility to make sure information published online for a general audience remains accurate, and avoids falling into the trap of hyping up the findings or technology. Multiple students also touched on different points to look out for

when doing science communication. For example, none overtly stated that researchers struggle with science communication, but their tone and emphasis on how to go about it suggest it may be an issue. Many students added qualifiers like "in a way that is understandable" when talking about science communication. This suggests that scientists usually communicate in a way that is not understandable. Here it is again:

> "That is an interesting and important part of the roles of the innovators as well. They have to make this information accessible to the public…in a way that they can digest it."

The importance of transparency and science communication also connects to the skills and knowledge code at the micro level.

One student brought up the power of doing public outreach in person. They said they like being able to tailor the information in real time depending on their audience. They gave an example of two different interactions they had while working a booth at a science fair, with a teenager and an elderly. Both times they were sending the same message about their science, but they were able to tailor the opening statements about the big picture of their science to appeal to her different audiences to convey their point.

> "I then use another tactic and still at the end send the same message of how important it is what I'm studying, but tackle it differently."

They said this was enough to get the attention of the 15-year-old girl who had previously been playing on her phone while half-listening to her. Another student brought up another point to keep in mind about the way the science message is conveyed:

> "I think it's really important to realize that when you're talking with the public, you're also really talking about yourself…we're also part of the public."

One person suggested citizen science as a potential avenue to assess "how people feel" about the technology they are developing. Citizen science is more of a public outreach activity than engagement because it is not usually meant for this purpose of collecting feedback from the public. However, it can be thought of as a precursor to RI, in building a baseline of knowledge among communities so they can better engage or be included in the future. Inclusion is a key principle in the RI framework (Stilgoe et al. 2013). This can be thought of as a step they can take towards RI when they otherwise thought it unattainable. There are steps that labs that don't have resources, or graduate students who don't have the authority, can take that align more directly with things they might already be doing, and then make a conscious effort to tailor these activities with the goal of RI.

> "[Public outreach] gives teachers a way to integrate new technologies and new innovations into their curriculum. It gives kids knowledge that they probably wouldn't get from a science textbook. They're getting exposure to new research that is happening now and I think that is really cool, and it gives them something to aspire to. It gives them an interest in science that they might not have had before."

One student said public outreach doesn't always need to be an event. They said that even something as simple as a good elevator pitch can have an impact. They encouragingly stated "You know, there's always an opportunity."

The overall message from these discussions on public outreach is, though outreach is not as ideal as engagement, it is much more feasible. Depending on how the public outreach is done, it could still be considered RI. If researchers learn about the knowledge deficit model (Sturgis and

Allum 2004) and learn to tailor their conversations in a more productive way, it is a low resource, low risk way to collect some quick feedback on their project. This may be ideal for graduate students who are operating on a short timeline and may not have time to wait for focus groups or a consensus conference. It also offers a mechanism of touching base more frequently, and this sustained relationship may have some advantages over a conference that meets once and never happens again.

Public engagement

Public engagement is multifunctional and has a bidirectional flow of information between researchers and the public—cross institutional contexts are important and this was coded at the meso-level. This is a divergence from public outreach, where the end goal is education and information flows in one direction from researcher to audience. Public engagement builds on outreach to use feedback to shape policy or research and development. In social science literature, public engagement is viewed as a more thorough way of including the public in science research. It closely aligns with "inclusion" as a principle of RI (Stilgoe et al. 2013). However, in the natural sciences, public outreach is the norm. In the interviews, public engagement was the most cited challenge at the meso-level. Interestingly, it was also the fourth highest opportunity. Nine students covered public engagement: three students talked exclusively about challenges, another three about opportunities, and the final three covered both opportunities and challenges (n=28, Table 5.5).

Six students talked about public engagement as a mechanism of inclusion that could facilitate the other three elements of RI: anticipation, reflexivity, and responsiveness. Two students said public engagement can increase transparency and accountability. One student said public engagement offers a mechanism for the public to have a "gate-keeper function" to ensure all the risks and benefits are considered, and hold scientists accountable. Another student said public engagement is a way to build upon mandatory "informed consent of field trials." Six students talked about reflexivity in relation to public engagement. Reflexivity is one of the four elements of

RI and is a process where scientists reflect on their assumptions and framing of the problem. These six students expressed concern over the current lack of reflexivity in academic norms. There was a distinct overlap in this particular concern, which is demonstrated in the following quotes from three students:

> "There's so much history of starting with the technology, taking it to communities, and having it fall flat."

> "Is this what the community really wants? ...Or are the scientists just creating a need to fill it?"

> "This is an opportunity to try and understand the social problem...To a man with a hammer, everything looks like a nail."

Thus, they present engagement as a mechanism to confirm the problem with stakeholders within the community, and ensure the possible research solutions align with the real world. As one of the students put it, scientists need to realize that innovation does not occur in a vacuum. They suggest applied researchers should also focus on problems presented by the end user and not just problems presented in the literature.

One student gave a clear example of the role the public can play in analyzing the desirability of potential solutions. This example demonstrates how proposed solutions that are inappropriate and inadequate for solving the problem can be selected when decisions are made without consulting the community. As part of an urban ecology course, the student visited an impoverished town with a high foreclosure rate and problems with squatters living in the empty houses. The local government's initial decision was to buy these properties and convert them to public parks. An assumption had been made that parks would solve the squatting problem and

create a positive space for children and families. In reality, the parks made the affected neighborhoods even less safe because they became gang activity centers and resulted in higher drug traffic. Upon meeting with the communities, the government was asked to tear down the parks and install community gardens. This opportunity to grow flowers and vegetables would increase food security, public safety, and property value. This experience left a strong impression on the student, since they had shared the same assumptions as the local government. They now recognize the public's vital role in assessing "what is actually wanted on the ground."

While six students talked about ways public engagement can incentivize RI, the same number of students talked about how it can act as a challenge. Challenges were mainly logistical such as the distance from current academic norms. One student said engagement is a structural problem, since this type of inclusion is not incentivized for natural scientists. Another student said they "find research to be very insular" and "lacking" with regards to public inclusion. Another student said their broader impact sections in grant applications have not been based on public input, and before learning about RI, they had never even considered that as an option. All of these accounts describe how different public engagement is from the normal level of inclusion.

In addition to academic norms, many logistical difficulties to public engagement were presented, including cost, choosing between different types of engagement, construction of the public, moderating polarized conversations, and mechanisms for incorporating feedback. A point made by two students encapsulates these difficulties; Engagement is a "democratic process," so it is important to keep a high level of transparency and include multiple groups when making decisions, but there are "a lot of politics that go into public and stakeholder engagement." Six students talked about the difficulty in defining the public or community. Two students who are experts on GPM gave some examples that have been paraphrased: If a genetically modified fruit fly was created to decrease pest populations for raspberry farmers, who should be included in the definition of the public, such as raspberry farmers, other farmers, non-farmers living in farming communities, pest control companies, venture capitalists, or raspberry consumers? Even with the

best intentions, construction of the public is difficult. This construction of the public, community, or stakeholders will directly affect the outcome of engagement. All six students who talked about public engagement as a challenge cited this as a major barrier.

One student said it is idealistic to include everyone possibly affected by the technology but recognizes this is impossible. They go on to describe the challenge of determining whose opinion to include and how those opinions get weighed. They also said not everyone will have a valid opinion. Dealing with non-valid opinions was also brought up by several other students. One student said it would be dangerous to make decisions based on all opinions, and cited anti-vaxxers as an example. Another student gave a GPM example:

> "If the community says 'We think that these insects are going to give us Chikungunya because they are genetically modified.' ...that's obviously not true, and that shouldn't stop [the research]."

In some cases it is obvious when feedback can be ignored, like in the quote above, when an opinion is based on objectively incorrect data. However, this student also pointed out this selection process becomes increasingly difficult when the situation is more nuanced. Multiple students presented challenges with interpreting feedback from engagement events, and "incorporating stakeholder feedback" into the innovation process.

A related point is the added challenge of interpreting feedback when the technology being discussed is controversial. Being able to respond to feedback is especially challenging in this case, since much of the feedback may be "black or white" when the reality is gray. When responses are in direct opposition to each other, it can be paralyzing trying to choose which concerns or risks to address. Dealing with a controversial topic also makes the upstream process of having engagement events harder, since they are discourse based. One student suggested graduate students doing public engagement on controversial topics, such as GPM, might want to

seek training in conflict resolution and group moderation. Another student said controversy may make public outreach a better option than engagement, since it is less "politically challenging" with potentially less opposition from their PI.

Another logistically complex question is when to include the public. It is expensive and time consuming to have constant communication or multiple events. One student suggested it would be ideal to have feedback when coming to a forked path in the research, where a major decision is made to commit to a path which would be difficult to change later. A second student had a similar thought that feedback is ideal earlier in the innovation process in order to think of all possible consequences from each trajectory. Then decisions could then be made by taking the path with the least possible risks.

In summary, nine students talked about public engagement, with six listing it as an opportunity and six as a challenge. Public engagement is a mechanism of inclusion that facilitates the other three elements of RI: anticipation, reflexivity, and responsiveness. Most opportunities focused on anticipation, reflexivity, while most challenges focused on inclusion and responsiveness. This is consistent with literature showing that biotechnology developers in industry and academe show most resistance to putting inclusion and responsiveness into practice (in comparison to anticipation and reflexivity) (Roberts, Herkert, & Kuzma 2020). Public engagement events can be expensive and hard to find funding for, and decisions on who to include and when are challenging. This may be especially true for controversial research that may result in polarized viewpoints on the technology.

Extension Programs

As a related topic to collaboration, two students talked about extension programs as an opportunity for RI (n=2, Table 5.5). Land-grant universities were established in the late 1800s with the mission of providing education and research in agriculture, and delivering this knowledge to farmers. Extension services are a big part of land-grant universities that have retained their agricultural roots. Extension programs research agricultural problems and possible solutions, and

aid in the education of farmers and adoption of innovations. Aspects of extension services are related to RI, especially inclusion and reflexivity, as evidenced by the students' responses to the interview. Students' thoughts on extension programs are captured by this quote:

> "Being at an extension school I think there is some ability to do outreach work that can be helpful, and I think that could count as responsibly innovating although it's not directly related to the research project if that makes sense. I think taking advantage of what's there and being reflexive."

Reflexivity involves assessing motivations and assumptions. These two students point out that extension programs are a structural part of universities that would aid in the process of reflexivity and inclusion through establishing connections between researchers and the end user. Opportunities to work with extension programs may especially help basic researchers, in helping them to connect to real-world applications and end users. As stated, even if a researcher's involvement in an extension program is not directly related to their research, it can still aid in RI. Perhaps this exposure to potential end users gives a researcher a better ability to assess research problems and understand the broader impacts.

Macro-level challenges and opportunities to RI

The macro-level is concerned with themes that relate to the larger political, economic, cultural, social contexts. Of all the levels, the macro-level had the lowest number of references (Table 5.2, n=38) and the highest challenge to opportunity ratio, with 87% of references categorized as a challenge (Table 5.6). Macro-level challenges mentioned by students were risk and values, mechanics or practicality, and publication, while opportunities included publication and mechanics or practicality (Table 5.6). The most predominant opportunity for RI that was mentioned was "publication" (n=3) whereas the most predominant challenge mentioned was

"risks and values" (n=16). We discuss these and other predominant macro-level themes below in each subsection.

Table 5.6

Comparison of macro-level opportunities and challenges to RI

Theme	# of Students	# of References	Opportunity	Challenge
Mechanics or practicality	9	16	2	14
Publication	5	6	3	3
Risk and values	6	16	0	16
Total	10	38	5	33

Mechanics or Practicality

Of the macro level themes, most students focused on the mechanics, or practicality of RI. We have defined this theme as logistical challenges of the four components of RI, including the lack of a clear framework to follow for inclusion (i.e. methods of democracy, how to include the public in decisions, and how to do public engagement). This theme has a lot of overlap with public engagement at the meso level, but is broader in scope. Nine students talked about this as a challenge (n=14, Table 5.6), and two of these also talked about the mechanics as an opportunity (n=2, Table 5.6). Many viewpoints on the challenges and opportunities are represented by this quote:

> "It's very high brow theory and it doesn't necessarily give practical ways of implementing it. So I think for a lot of natural scientists it's really good because it pushes them outside the boundaries they would normally use but in my work with social scientists I don't think it goes nearly far enough of questioning motivation and integrating viewpoints that aren't directly from the scientists."

One student took issue with the terminology:

> “First of all, I don't like the word responsibility that much because I think it sounds like you're either a jerk or you're not a jerk …I think life is more complex than that.”

This student suggests the complexity of life is mirrored in the innovation process. This viewpoint connects with the other challenge references for this theme, as they all outline complexities in the mechanics or implementation of the RI framework.

One student described the difficulty in choosing an adequate mechanism of inclusion, since all of them have major drawbacks. They also describe potential limitations of the types of people who might be available to be involved in the public engagement process. They give an example from their personal life, that they try to stay involved in their community by going to town hall meetings. However, information on these meetings is hard to find, and meetings are often at times they are unavailable. They consider their schedule to be fairly flexible as a student with no children and can imagine how difficult it would be for people with families to participate. Another student questioned if the public should even have a role in decisions about science. They state that the media plays a big role in changing people’s opinions about a topic, even if the media articles are not factual. Some people have poor critical thinking skills, and will believe anything the news tells them. Therefore, they would not want decisions about their research to be influenced by these types of people. This is in contrast to another student’s point that public feedback is important for developing a list of risks and benefits, but difficult to obtain since some feedback may be highly polarized and not constructive.

Incorporation of feedback is a challenge identified by multiple students, and one student identifies responsiveness as the biggest hurdle to RI. They said all steps present their own challenges, but if responsiveness falls through, then the entire concept of RI is undermined. A researcher could do the other three steps (anticipation, reflexivity, and inclusion) without responsiveness and pay “lip service” to innovating responsibly. Only if research is actually affected by findings from these processes, would it be considered innovating responsibly under

the RI framework. This makes it tricky since findings could also be interpreted in ways that fit a researcher's needs. Perhaps less defeatist, another student makes a statement about responsiveness:

> "Responsiveness…is the one I especially have a hard time seeing exactly what it is, or what it should be because, …there are always going to be people pushing back no matter what. So, how far do you go in terms of responding?"

This question of how to weigh and include feedback was also covered in the public engagement theme at the meso level.

Another student addresses all four elements of RI, and said they would have trouble applying RI to their own process because they are working on basic research. In their case, their work could support the development of multiple technologies with multiple potential end users. They state that these "context dependencies" prevent them from using RI on this early part of the technology innovation pipeline. The same person goes on to say:

> "If you bring in a bunch of people early on and don't have any answers to anything, I don't know if it's a really good use of time. I don't know if that really gets at responsible innovation because you are kind of floating with a lot of questions and you can't give any direction."

This quote connects back to a point made in the extension theme, that basic researchers may have more trouble doing public engagement. The suggested solution for agricultural researchers was to be involved with extension services in order to have frequent conversations with potential end users, even if the conversations are not within the context of their own research.

While nine students talked about the mechanics of RI as a challenge, two students gave overt endorsements of the RI framework. One said they "really like the idea of the framework" and have relied heavily on it in their own research. They describe their viewpoint:

> "Having tenets, even if they're idealized, to hold onto is something we've used in every [grant] application and often in conference presentations."

Another student also said they have found the RI framework to be useful in thinking about innovating responsibly. They especially like how it calls for inclusion of the public and think it should be something to work towards. Both of these endorsements seem to acknowledge the framework as being idealistic, yet useful nonetheless.

Risks and Values

The second most coded theme at the macro level is risks and values (n=16, Table 5.6). This theme deals with academic norms which transcend institutions and are therefore coded as a macro-level challenge. In natural science research, ecological risk and human safety are usually the only types of risk considered. In other words, risk assessment does not typically include risks stemming directly from human values. These include social justice issues like who gets to make decisions about a project and where field trials will take place. It also includes how environmental risks get weighed during evaluation, since diverse people will view the same risks differently. For example, genetically modified corn has not been adopted in Mexico because corn is a deep cultural symbol intrinsic to daily life. They view corn as a gift from God and a symbol of Mexico itself. Corn is said to spiritually, physically, and economically sustain the indigenous peoples, which means developing genetically modified corn would jeopardize their way of life. In this context, there are risks associated with the research that are not captured by the usual scope of risk assessments. For another example, people may have varying views on the intrinsic value of

nature, which will influence their perception of environmental risks in ways that might not align with regulation. All six students who talked about this theme presented it as a challenge.

Publication

Of the macro level themes, publication was coded the least. For this theme we included both the verb and noun sense of publication—that is ability to publish as well as the content of publications. Six students talked about publication with half talking about it as a possible opportunity (n=3, Table 5.6), and the other half framing it as a challenge (n=3, Table 5.6). One student presented publication as a possible way to influence academic norms. By publishing methods that include RI, a researcher can bring awareness to the topic and influence adoption. As the methods are cited in other papers, there is an increased possibility of adoption and shift in academic norms. They connect publication back to the mentorship theme, and say this expands on the ability of professors to set new standards across universities. In direct contrast to this, another student brought up the point that RI is not typically discussed in the natural science literature. Some researchers are doing things to responsibly innovate but don't include it in publications. This lack of literature makes it hard for graduate students trying to incorporate RI into their own methods or discuss RI as part of a journal club. A possible solution to this problem is the following point made by another student. They suggest the Journal of Responsible Innovation as a venue where technical experts can "feel comfortable to raise their voice" and "share knowledge that enriches the discussion." Having a venue for this type of publication may be a "reward or incentive" to researchers, since publishing is a common goal of academics and criteria for promotion. This also presents a solution to a challenge listed by another student, that RI might not fit under the scope of science journals.

Another student said science communication is an important part of RI, and publications can act as a challenge by making researchers feel like they are having an open dialogue. They argue writing a letter to the editor or blogging about a topic of concern are one-sided processes and not an open forum for communication. This view is contrasted by a point made by another

student that being a co-author with a social scientist or ethicist could be an enriching learning opportunity, as opposed to having an external reviewer. This connects to several points made in the collaboration theme at the meso-level, that collaboration results in richer conversations about the topic.

Discussion

In the US, funding for public institutions historically came from the government but increasingly comes from private funding agencies, industry, and interest groups. Lack of funding for STEM related social science is a potential challenge to interdisciplinary collaborations. While researchers can obtain their own funding and still collaborate, we think grants funding researchers from multiple disciplines would simplify the RI process and promote a deeper integration between researchers. For example, granting agencies could call for applications to fulfill their impact section through RI processes such as public engagement. If funding agencies gave grants with RI components priority, more researchers would be incentivized to collaborate for RI. For example, De Campos et al. (2017) wrote in the context of GPM in Brazil that responsible innovation needs to be linked to political accountability. Extending their research into a US context, we argue accountability be required of both research groups through incentives from funding agencies.

Another opportunity is to create an academic environment conducive to the study of RI. When proposing to create shifts in research culture, academia presents a unique opportunity. Students are undergoing active training and mentoring, and therefore may be more malleable to a shift in research inclusive to RI. Through classes and professional development opportunities, graduate school also presents an opportunity as a format for educating researchers on this complex topic. Therefore, graduate education may be an ideal environment to train genetic engineering researchers in RI. Cross-disciplinary training may help natural scientists better contextualize their research, and give them necessary knowledge and tools. This can be accomplished through classes, training opportunities, and participating in interdisciplinary

communities. Interdisciplinary programs encourage collaboration and mentorship, which can increase the authority of students to pursue RI. For campuses that do not already have ethics courses and sustained interdisciplinary communities, these will need to be better promoted from at high organizational and institutional levels. One of the interviewees talked about collaborative centers as a space and tool for students to increase their authority. Interdisciplinary collaborations can make the multiple steps of RI (anticipation, reflexivity, inclusion, responsiveness) stronger, and universities can facilitate these relationships by providing a space for these interactions to occur.

Conclusion

Some researchers claim responsible innovation can ensure a broad range of stakeholders and publics inform decision making based on their values and preferences. RI also encourages scientists to anticipate, discuss, and reflect on risks, goals, and motivations with these diverse publics. Despite its promise and popularity within the social science community, in biotechnology research RI principles have yet to be transformed into practice. Where previous studies analyzed perspectives held by stakeholders and established researchers on RI of biotechnologies, early-career researchers have been largely overlooked. Thus, we used semi-structured interviews with ten graduate students on RI of biotechnology to ask about RI in respect to gene editing, gene drive, and genetic pest management. One limitation of this study is that the early-career students interviewed had some prior exposure and training in RI, and demonstrated interest in RI by signing up for the NSF-CCE grant program. Thus, the results are not generalizable both due to the small sample size (n=10) and the special nature of the NSF-CCE program and GES Center at NC State University. Interviewees were chosen for their interest in RI, so other samples of graduate students and early-career researchers may differ in this regard.

In the interviews, students were generally enthusiastic about the need for RI in biotechnology innovation systems and spent the most time talking about the feasibility of incorporating RI into their work. They viewed feasibility as facilitated in three ways: duty, public

outreach, and collaboration. Possibly the most important factor contributing to the students' optimistic view of the RI framework was duty. Students feel their unique position as the labor force gives them power to influence the trajectory of the technology. Instead of referring to a social scientist or ethics consultant to check the metaphorical RI box, students posited that genetic engineers' unique knowledge puts them in a position to contribute vital information to the RI process and thereby influence the technology's trajectory. Yet they also feel "powerless" in some ways by needing the support of their adviser to engage deeply with RI activities. To have the time and funding to pursue RI training and activities, students would need buy-in from their advisor. Asking questions about societal implications of genetic engineering could undermine their relationship with their advisor. Demonstrating responsiveness in their project could undermine the progress in their dissertation. However, in some ways students feel better positioned than professors to address RI issues due to their close proximity to the research and focus on a fewer number of projects. There is a tension between the duty they feel and the challenges that exist at the micro and meso-levels.

The most talked about opportunities for overcoming these challenges were collaboration and public outreach. Students viewed collaborations as especially useful in cases where they do not have the authority to conduct RI on their own. A collaborator, independent mentor, or outside funder could support students in pursuing RI, especially in situations where major advisors view the topic as irrelevant. Students see public outreach as an opportunity since it is a current norm in graduate education that can be repurposed for RI. Students cited outreach as a way to increase transparency and build science communication skills important for RI practices like public engagement and interdisciplinary collaboration.

Challenges to implementing RI were more often talked about in the context of the student (rather than professor) experience. This delineation was interesting because, with the exception of one question towards the end of the interview, our questions did not prompt the discussion. From the beginning of individual interviews, students expressed differences between professors

and students in respect to RI roles and abilities. Authority proved the second most coded theme across all interviews, mentioned mostly as an opportunity for professors and as a challenge for students. Overall, students perceived professors as having a much lower ratio of challenges to opportunities (13% challenges) than students (67% challenges).

Beyond authority, the challenges interviewees most commonly cited to implementing RI were skills, knowledge, and public engagement. As mentioned earlier, the genetic engineering community has demonstrated a desire to use the RI framework. However, students pointed to the lack of training in skills and knowledge that would enable them to do so. Genetic engineering curriculum often does not include ethics, social accountability or societal impacts. Students are trained in public outreach, not public engagement. As a first step, students suggested increased awareness and understanding of RI principles could be promoted through graduate education.

Public engagement was the third most referenced challenge and was viewed as a difficult yet essential part of RI. Many students discussed the importance of public engagement to test the assumptions and framing of a problem. They stated that biotechnology research tends to lack reflexivity, as one student quoted "To a man with a hammer, everything looks like a nail." However, public engagement is not the academic norm so there is no training or funding for it. Additionally, incorporating feedback to shape research (responsiveness) may be difficult for students. Rigidity of the dissertation process may stifle students' ability to be responsive. Graduate researchers experience varying levels of control over their projects and some feel locked into a trajectory. Overall, students' greatest concern was their lack of knowledge of ethics and social science and the accompanying lack of skills to carry out necessary components of RI. Despite these challenges, students viewed RI positively. They saw the need for such a framework and its potential to enhance the benefit that can be generated from science for society.

REFERENCES

Baltzegar, J., Cavin Barnes, J., Elsensohn, J. E., Gutzmann, N., Jones, M. S., King, S., & Sudweeks, J. (2018). Anticipating complexity in the deployment of gene drive insects in agriculture. *Journal of Responsible Innovation*, *5*(sup1), S81-S97.

Biddle, J. B. (2017). Genetically engineered crops and responsible innovation. *Journal of Responsible Innovation*, *4*(1), 24-42.

Bruce, A., & Bruce, D. (2019). Genome editing and responsible innovation, can they be reconciled?. *Journal of Agricultural and Environmental Ethics*, *32*(5-6), 769-788.

De Campos, A. S., Hartley, S., de Koning, C., Lezaun, J., & Velho, L. (2017). Responsible Innovation and political accountability: genetically modified mosquitoes in Brazil. *Journal of Responsible Innovation*, *4*(1), 5-23.

Esvelt, K. M., Smidler, A. L., Catteruccia, F., & Church, G. M. (2014). Emerging technology: concerning RNA-guided gene drives for the alteration of wild populations. *elife*, *3*, e03401.

Glerup, C., Davies, S. R., & Horst, M. (2017). 'Nothing really responsible goes on here': scientists' experience and practice of responsibility. *Journal of Responsible Innovation*, *4*(3), 319-336.

Herkert, J. R., Kuzma, J., Roberts, J. P., Banks, E., & Stauffer, S. A. (2017). Ethics and responsible innovation in biotechnology communities: A pedagogy of engaged scholarship. In *Proceedings of the American Society for Engineering Education*.

Jordan, N., Dorn, K., Smith, T., Wolf, K., Ewing, P., Fernandez, A., Runck, B., Williams, A., Lu, Y. and Kuzma, J., 2017. A cooperative governance network for crop genome editing. *EMBO reports*, 18(10), pp.1683-1687.

Kaebnick, G., Heitman, E., Collins, J., Delborne, J., Landis, W., Sawyer, K., Taneyhill, L. and Winickoff, D., 2016. Precaution and governance of emerging technologies. *Science*, 354(6313), pp.710-711.

Kofler, N., Collins, J., Kuzma, J., Marris, E., Esvelt, K., Nelson, M., Newhouse, A., Rothschild, L., Vigliotti, V., Semenov, M., Jacobsen, R., Dahlman, J., Prince, S., Caccone, A., Brown, T. and Schmitz, O., 2018. Editing nature: Local roots of global governance. *Science*, 362(6414), pp.527-529.

Kokotovich, A. E., Delborne, J. A., Elsensohn, J., & Burrack, H. (2020). Emerging technologies for invasive insects: the role of engagement. *Annals of the Entomological Society of America*.

Kuzma, J., Kokotovich, A., & Kuzhabekova, A. (2016). Attitudes towards governance of gene editing. *Asian Biotechnology and Development Review*, *18*(1), 69-92.

Kuzma, J., Gould, F., Brown, Z., Collins, J., Delborne, J., Frow, E., Esvelt, K., Guston, D., Leitschuh, C., Oye, K. and Stauffer, S., 2017. A roadmap for gene drives: using institutional analysis and development to frame research needs and governance in a systems context. *Journal of Responsible Innovation*, 5(sup1), pp.S13-S39.

Kuzma, J., & Roberts, P. (2018b). Cataloguing the barriers facing RRI in innovation pathways: a response to the dilemma of societal alignment. *Journal of Responsible Innovation*, *5*(3), 338-346.

Kuzma, J. (2019). Procedurally robust risk assessment framework for novel genetically engineered organisms and gene drives. *Regulation & Governance*. 1–22. https://doi.org/10.1111/rego.12245

Kuzma, J., & Cummings, C. L. (in review). The influence of cultural beliefs and stakeholder affiliation on attitudes towards responsible research and innovation in biotechnology.

Lacey, J., Coates, R., & Herington, M. (2020). Open science for responsible innovation in Australia: understanding the expectations and priorities of scientists and researchers. *Journal of Responsible Innovation*, *7*(3), 427-449.

Lukovics, M., Udvari, B., Nádas, N., & Fisher, E. (2019). Raising awareness of researchers-in-the-making toward responsible research and innovation. *Journal of the Knowledge Economy*, *10*(4), 1558-1577.

Macnaghten, P., Owen, R., & Jackson, R. (2016). Synthetic biology and the prospects for responsible innovation. *Essays in biochemistry*, *60*(4), 347-355.

McLeod, C., de Saille, S., & Nerlich, B. (2018). Risk in synthetic biology—views from the lab: Early career scientists' concerns about synthetic biology open up new perspectives on risk and responsibility in research. *EMBO reports*, *19*(7), e45958.

Mitchell, P. D., Brown, Z., & McRoberts, N. (2018). Economic issues to consider for gene drives. *Journal of Responsible Innovation*, *5*(sup1), S180-S202.

National Academies of Sciences, Engineering, and Medicine. (2016). *Gene drives on the horizon: advancing science, navigating uncertainty, and aligning research with public values*. National Academies Press.

Pandza, K., & Ellwood, P. (2013). Strategic and ethical foundations for responsible innovation. *Research Policy*, *42*(5), 1112-1125.

Pansera, M., Owen, R., Meacham, D., & Kuh, V. (2020). Embedding responsible innovation within synthetic biology research and innovation: insights from a UK multi-disciplinary research centre. *Journal of Responsible Innovation*, *7*(3), 384-409

Patton, M. Q. (2002). Qualitative research and evaluation methods. Thousand Oaks. *Cal.: Sage Publications*.

Pennisi, E. (2013). The CRISPR craze. *Science*, 341(6148), 833-836.

QSR International Pty Ltd. (2018) NVivo (Version 12), https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home

Roberts, P., Herkert, J., & Kuzma, J. (2020). Responsible innovation in biotechnology: Stakeholder attitudes and implications for research policy. *Elementa: Science of the Anthropocene*, *8*.

Roco, M. C., Harthorn, B., Guston, D., & Shapira, P. (2011). Innovative and responsible governance of nanotechnology for societal development. In *Nanotechnology research directions for societal needs in 2020* (pp. 561-617). Springer, Dordrecht.

Stahl, B. C., & Coeckelbergh, M. (2016). Ethics of healthcare robotics: Towards responsible research and innovation. *Robotics and Autonomous Systems*, *86*, 152-161.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research policy*, *42*(9), 1568-1580.

Stilgoe, J. (2015). *Experiment earth: responsible innovation in geoengineering*. Routledge.

Sturgis, P., & Allum, N. (2004). Science in society: re-evaluating the deficit model of public attitudes. *Public understanding of science*, *13*(1), 55-74.

Wickson, F., & Carew, A. L. (2014). Quality criteria and indicators for responsible research and innovation: Learning from transdisciplinarity. *Journal of Responsible Innovation*, *1*(3), 254-273.

SUPPLEMENTARY MATERIAL

Supplementary Material 5.1

Codebook

Name	Description
Challenges	An obstacle or challenge that prevents access or otherwise decreases the ability of a researcher to do any portion of the RRI framework (Anticipation, Reflexivity, Inclusion, Responsiveness).
Opportunities	A stimulus or motivator that facilitates or otherwise increases the ability of a researcher to do any portion of the RRI framework (Anticipation, Reflexivity, Inclusion, Responsiveness).
Micro-level	Individual researcher: graduate students or professors.
Professor	Micro-level themes experienced by professors (as perceived by students).
Access	Having the means to do RRI. i.e. time, funding, network, collaborators
Authority	The power to give orders and make decisions. This code can also be for the lack of authority.
Mentorship	Professional relationship where one is guided by a more experienced scholar. (In this case between a professor and graduate student)
Student	Micro-level themes experienced by graduate researchers.
Authority	The power to give orders and make decisions. This code can also be for the lack of authority.
Independent mentor	A mentor that is partially or completely independent of the committee. Can be used for support.

Duty	The idea that students have the ‘onus’ or ‘responsibility’ to do RRI.
Skills or knowledge	Having/not having the skills or knowledge needed to do RRI. i.e. communication, ethics, interdisciplinary
Time	Having/not having enough time. i.e. Having time dedicated to RRI vs. being unable to do RRI because other obligations take up all the hours in the work week.
Meso-level	Within organizational structures: department, center, college, university, etc.
Collaboration	Working together on ideas, projects, or otherwise forming a community
Extension programs	Public service of land grant university. Extension programs research agricultural problems and possible solutions, and aid in the education of farmers and adoption of innovations.
Funding	Monetary resources for: research, tuition, training workshops, public engagement, etc.
Lab focus and inheritance	Each academic lab may already have grants and therefore already have a vision and accountability to funders. Can limit the ability to change research or add new direction.
Public engagement	Event(s) with a bidirectional flow of information between researchers and public. Multifunctional: increase transparency through science communication, respond to social needs of research by talking to target communities, and build trust and understanding.
Public outreach	When scientists share their research or general science knowledge with the public for the purpose of education.
Training opportunities	Professional development or educational courses, workshops, or events
Macro-level	Relating to the larger political, economic, cultural, social contexts.

Mechanics or practicality	Lack of a clear framework to follow for the inclusion component of RRI, and difficulties in inclusion. i.e. methods of democracy, how to include the public in decisions and how to do public engagement.
Publication	In both the verb and noun sense. Verb: Being an author on a paper accepted in a scientific journal. Noun: A published journal article.
Risk and values	In natural science research, ecological risk or human safety are usually the types of risk considered (does not typically include human values or societal issues).