UNIVERSITY OF CALIFORNIA, SAN DIEGO

The mutagenic chain reaction: from Evo-Devo to active genetics.

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Biology

by

Valentino M. Gantz

Committee in charge:

Professor Ethan Bier, Chair
Professor William McGinnis
Professor Scott Rifkin
Professor Joseph Vinetz
Professor Steven Wasserman

2015

The Dissertation of Valentino M. Gantz is approved and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A. gam.	*Anopheles gambiae*
A. ste.	*Anopheles stephensi*
ab	*abrupt*
caup	*caupolican*
cDNA	complementary DNA
Ci	*Cubitus interruptus*
D. gri.	*Drosophila grimshawi*
D. mel.	*Drosophila melanogaster*
Dl	*Delta*
en	*engrailed*
indel	insertion/deletion mutation
kni	*knirps*
M. abd.	*Megaselia abdita*
Ma	million years ago
rho	*rhomboid*
ri	*radius incompletus*
salm	*spalt-major*
al	*aristaless*
brk	*brinker*
dpp	*decapentaplegic*
Cas9	CRISPR associated
chiRNA	chimeric RNA (or sgRNA)
CRISPR	clustered regularly interspaced palindromic repeats
CRM	cis-regulatory module
crRNA	CRISPR RNA

tracrRNA	trans-activating CRISPR RNA
GRN	Gene regulatory network
gRNA	guide RNA
knrl	*knirps related*
sd	*scalloped*
ssODN	single strand Oligo Deoxynucleotide
sgRNA	synthetic guide RNA (or chiRNA)

ACKNOWLEDGEMENTS

First, I want to thank my mother and father for cultivating my curiosity and creativity since I was a child, and for supporting my life decisions whether they agreed with them or not. I also need to thank my grandmother for fostering my interest in biology, and my grandfather for structuring my problem-solving mind set. I would like to thank all present and former member of the Bier lab for support, insights, suggestions or reagents that were vital for my thesis project: Orna Cook, Francisco Esteves, Long Do, Claudia Mieko Mizutani, Margery Smelkinson, Tamar Grossmann, Beatriz Moreno Cruz, Annabel Guichard, Mark Sturtevant, Dave Kosman, Shannon (Xiang-Ru) Xu, Matthew Smith, Eric Babajanian, Elisabeth Manao, Sachin Sethi and Stephen Chin. In particular I would like to thank Orna for being an outstanding mentor and teaching me fruit fly genetics and techniques to study gene expression; Francisco for being an amazing friend, as well as a colleague, with whom I shared all the ups and downs of my time at UCSD; Long for providing essential computational skills that where key part of my project (Chapter 2.2); and Adrianne who other than lab mate and friend came to be my best roommate ever. I would like to also acknowledge all my mentors that contributed to my career throughout my studies: Antonio Contestabile, Giovanni Perini, Patricia Dietz, Bjoern Hock,Vicki Lundblad, Jan Karlseder, Amy Pasquinelli, Martin Hetzer, Jessie Vargas, Daniel Anderson, as well as the members of my PhD program committee: William McGinnis, Steven Wasserman, Therese Markow, Scott Rifkin and Joseph Vinetz who all contributed and supported my project throughout its development. I also would

like to thank all the people (fly people and not) on the 4th floor in Bonner Hall which together allow for an extremely collaborative environment; special thanks to Steve Miller, Scott Lindsay, Alexa Clemmons, Bo Zhou, Ella Tour, Dmitry Karachentsev, Sandra Diaz Garcia, Anita Hermann. I would also like to acknowledge the contribution of collaborators and I'll other colleagues I contributed to my project: Marco Gallio, Urs Schmidt-Ott, Johannes Jaeger, Luca Cozzuto, Anthony James, Nijole Jasinskiene, Saurabh Sinha, Md Abul Hassan Samee, Yoshinori Tomoyasu, Maxi Polihronakis Richmond, Richard Kolodner, Brenda Wong and Aristea Lubar. Also I want to thank my friends, scientists and not, that one way or another helped along the way: Andrew Halff, Micah Cranman, Giordano Lippi, Justin Lu, Shigeki Miyake-Stoner, Annie Chou, Gunnar Poplawski, Pedro Resende, Nuno Camboa, Paola Cattaneo, Martin Kolnik, Ivan Razinkov, Catarina Fernandes, Cherine Belal, Jack Nguy, Cameron and Skye Esfahani, Allis Pham, Robert Teachenor, Tasha and Frank Toro, Bridget Baumgartner and Margherita Paschini. I also would like to thank my girlfriend Ellen Sun for helping me with my ramping stress while I was writing my thesis. Last but not least, I would like to thank my PhD advisor Ethan Bier for being an outstanding mentor and supporting my ideas no matter how crazy they were, his wife Kathryn for the amazing culinary experiences, and his bird Sake for training my reaction time.

Chapter 4 and 4.1 contain part of, and Chapter 4.6 is in full a reprint of the material as it appears in Science "*The Mutagenic Chain Reaction: a Method for Converting Heterozygous to Homozygous Mutations*." Gantz, Valentino M, Bier, Ethan

(2015) Science, 348(6233): 442–44. The dissertation author was the primary investigator and author of this paper.

VITA

2006	Biotechnology Intern, MERK KGaA, Darmstadt, Germany
2006	Dottore in Biotecnologie, Università di Bologna, Italy. (Bachelor equivalent)
2007-2008	Cellular and Molecular Biology Intern, Salk Institute, La Jolla, CA
2008	Dottore in Biotecnologie Molecolari ed Industriali, Università di Bologna, Italy. (Masters equivalent)
2008-2015	Doctor of Philosophy in Biology, University of California, San Diego

PUBBLICATIONS

Gantz, Valentino M, and Ethan Bier. (2015). *"The Mutagenic Chain Reaction: a Method for Converting Heterozygous to Homozygous Mutations."* Science, 348(6233): 442–44.

Akbari, Omar S, Bellen, Hugo J, Bier, Ethan, Bullock, Simon L, Burt, Austin, Church, George M, Cook, Kevin R, Duchek, Peter, Edwards, Owain R, Esvelt, Kevin M, **Gantz, Valentino M**, Golic, Kent G, Gratz, Scott J, Harrison, Melissa M, Hayes, Keith R, James, Anthony A, Kaufman, Thomas C, Knoblich, Juergen, Malik, Harmit S, Matthews, Kathy A, O'Connor-Giles, Kate M, Parks, Annette L, Perrimon, Norbert, Port, Fillip, Russell, Steven, Ueda, Ryu and Wildonger, Jill. (2015). *"Safeguarding gene drive experiments in the laboratory."* Science, DOI: 10.1126/science.aac7932.

ABSTRACT OF THE DISSERTATION

The mutagenic chain reaction: from Evo-Devo to active genetics.

by

Valentino M. Gantz

Doctor of Philosophy in Biology

University of California, San Diego, 2015

Professor Ethan Bier, Chair

While there has been a great deal of research trying to understand the link between DNA information the regulatory level in the final morphological outcome of an organism, a deep understanding of this relationship remains elusive. Here I attempted a new integrated approach to problem by: 1) comparing DNA sequences taken from different animals that have been diversified during evolution to generate altered function. 2) exploring a novel approach to dissecting cis-regulatory modules (CRMs) using *in locus* Cas9 genome editing; and 3) developing the mutagenic chain reaction, a novel active genetic element that can be use to easily perform cis-regulatory analysis in organisms with no existing genetic tools.

1. Introduction: understanding the link between DNA and body shape.

One of the most challenging problems in biology is to understand the link between gene expression during development and the final morphological outcome observed in the adult body plan. Such an analysis can be largely reduced in many instances to understanding the linkage between DNA-encoded transcriptional information within Cis-Regulatory Modules (CRMs) and the morphological structure determined by the regulated genes. CRMs, which include enhancers, silencers and insulators, are cis-acting DNA sequences capable of driving gene expression in a specific place and time during development. The term "module" indicates the CRM property to maintain their function independently from their genomic context. This property allows CRMs to be linked to a heterologous transgene reporter, which from a practical experimental standpoint has helped streamline functional analysis. The implicit inference that the behavior of CRMs will be the same in endogenous versus novel genomic contexts, however, has not been rigorously tested in most cases. Since CRMs contain the information for correct gene expression in time and place doing development, they are the primary substrate for morphological evolution (Davidson and Erwin, 2006, Erwin and Davidson, 2009) and thus a CRM Evo-Devo analysis should unveil relevant information.

1.1. The wing development of *Drosophila melanogaster.*

In *Drosophila melanogaster* the cis-regulatory relationships that determine development of the adult wing structure have been extensively studied (reviewed in: Blair, 1995, Bier, 2000, de Celis, 2003). The sum of all such relationships constitutes the wing Gene Regulatory Network (GRN) summarized in Figure 1B. The Drosophila wing disc (or wing primordium) develops form an epithelial monolayer known as the larval wing imaginal disc. The wing disc everts (buds out) during early pupal stages (0-12h after pupation) to form a sleeve shaped bilayer, which will constitute the wing blade (Fig. 1A). The dorsal and ventral surfaces of the wing come together and are joined via integrins during late pupal stages (16-24 h) to form two morphologically distinct structures: (A) the veins which are hollow tubes formed by small pigmented cells that provide structural support for the flight and a conduit for nutrients to living cells in the wing such as sensory organs; and (B) the wing membrane, which is composed of several layers of cuticle in the adult wing and provides a resistance surface during movement of the wing through the air.

Vein patterning in the *Drosophila* wing is initiated by the transcription factor *engrailed* (*en*), which is expressed in the posterior compartment of each body segment in all *arthropoda* (Patel et al., 1989), including the wing primordial structure. Engrailed activates the expression of the *hedgehog* (*hh*) gene which encodes a diffusible signaling molecule. At the same time, Engrailed represses the function of Hh within its own

expression domain, restricting Hh activity to a narrow strip of anterior compartment cells that abut the posterior compartment (Strigini and Cohen, 1997). In this central region, Hh activates the expression of several genes required for development of the L3 and L4 veins and induces expression of *decapentaplegic* (*dpp*), which encodes a long range morphogen responsible for establishing the positions of the L2 and L5 veins (and possibly L1 and L6 according to the model schematized in Fig. 1B). Dpp protein symmetrically diffuses from its central zone of production in both anterior and posterior directions resulting in graded activation of the *Tkv* (*Thickveins*) BMP-receptor, which decreases as a function of distance from the A/P boundary. Activation of the *tkv* receptor leads to the downstream phosphorylation of Mad (Mother against Dpp) which is a transcriptional effector of the Dpp signaling pathway that translocates from the cytoplasm into the nucleus as a consequence of signal-dependent activation. Transcriptional targets of Mad are activated above a specific threshold level of Dpp signaling so that the genes *spalt* and *omb* have a progressively wider expression domain. It has been shown that expression of *knirps* (Lunde et al., 2003) and *abrupt* (Cook et al., 2004) is induced respectively at the edges of the *spalt* and *omb* expression domains. *abrupt* is expressed in L5 and L1 (Fig. 1B), which form respectively along the posterior and anterior edges of the *omb* expression domain (Cook et al., 2004). *knirps* is expressed in L2 primordium, which runs along the anterior border of the *salm* expression domain (Lunde et al., 2003), but only low levels of Kni protein can be detected along the posterior *salm* border (Fig. 1B, "○", data not shown). No vein forms along the posterior *salm* border in *Drosophila* while a vein, termed M4, is

found in this position in more primitive insects (Fig. 1C, "○" vs "●"). This morphological difference suggests that the M4 border might have had an ancestral role in determining the differentiation of a vein that was lost in the *Drosophila* lineage (*Schizophora*).

Several attempts have been made to dissect functions of wing specific CRMs such as *spalt* (de Celis et al., 1999, Barrio and de Celis, 2004), *optomotor-blind* (Sivasankaran et al., 2000), and *knirps* (Lunde et al., 2003). None of these studies, however, have provided a complete roster of the composition of Transcription Factor (TF) binding sites and their functional relation to the transcriptional output of the whole module. The transcriptional readout of a CRM is reflected into a morphological structure, which in case of the Dipteran wing, is under a high selective pressure, being an organ under severe mechanical constraints necessary for the survival of the animal. Such selective pressure acts at the level of CRM DNA sequence (and more specifically TF binding) constraining mutations that such a sequence can withstand.

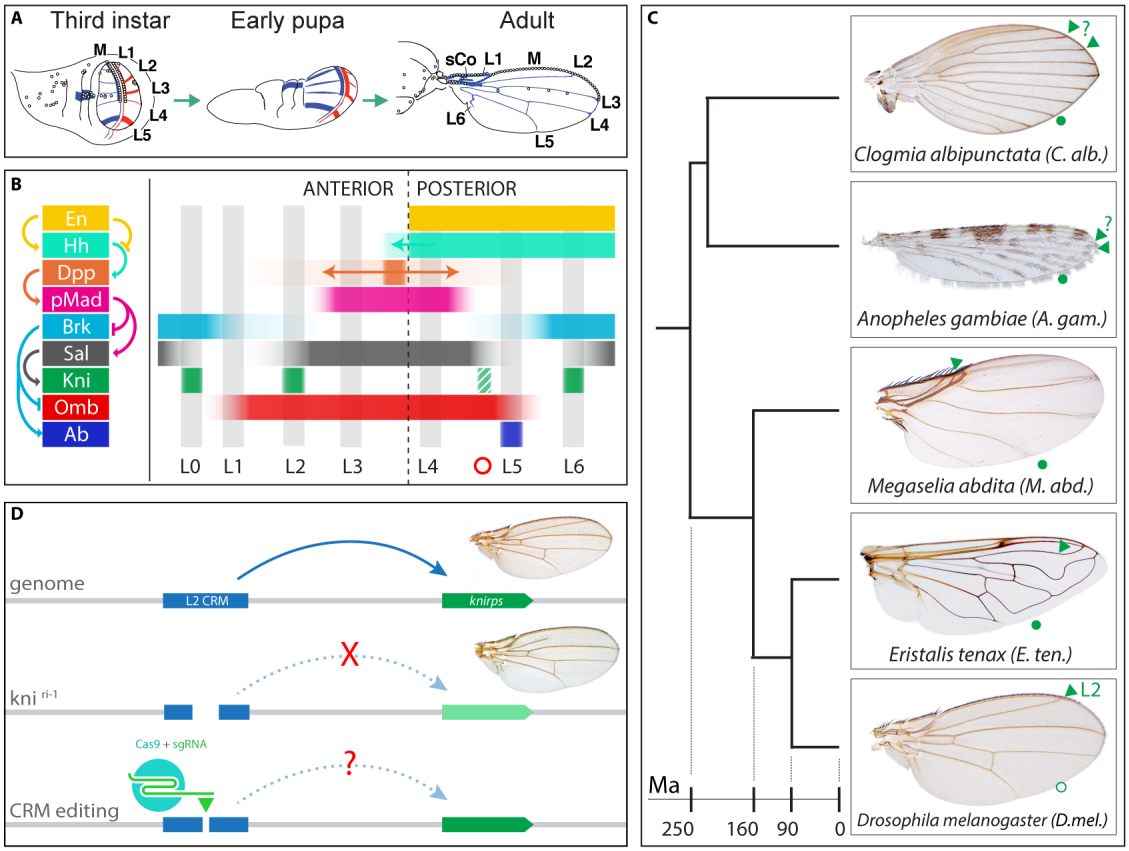

Figure 1 - The wing of *Drosophila melanogaster*.

The developmental steps that lead to the formation of the adult wing structure are depicted in (A): during the third larva instar veins are determined by specific transcription factors expressed in longitudinal lines drawn on the cell monolayer that constitutes the wing disc. During early pupal development the disc everts, folding along the mid line, and brings together the dorsal (blue) and a ventral (red) surfaces, therefore generating the bilayer cellular structure that forms the adult wing blade. Panel (B) summarizes, on the left, known regulatory relationships that constitute the wing GRN and, on the right, a graphical representation of the gene expression patterns along the anterior/posterior (A/P) axis. (C) shows a sample of wing venation patterns along the *diptera* evolutionary tree including the one *D. mel.* and *M. abd.* (studied in chapter 2). The green triangles (▼) identify the wing vein L2, determined by *knirps* expression, and its structural homologs (or putative homologs,▼?). The green circles (●), instead, identify the vein M4 which associated with the posterior *spalt* border in *M. abd.* and hypothesized to have been lost in *D. mel.*(○). Panel (D) outlines the activity of the L2 CRM on the gene *knirps (kni)* which if impaired, such as in the kni^{ri-1} mutant, results in defects in L2; the bottom section of panel (D) schematizes a CRISPR-based approach aiming to create new regulatory alleles of the L2 CRM (see chapter 3).

1.2. Evolution of the wing Gene Regulatory Network.

As mentioned in the previous section, much of the known transcriptional regulation of *knirps* and *abrupt* in L2 and L5 respectively is summarized in Fig. 1B. One might expect that such a hierarchical Gene Regulatory Network (GRN) would be strongly conserved in closely related species and would serve as a substrate for evolutionary change as wing patterns diversified. It has been hypothesized that genes acting at higher levels in the transcriptional hierarchy ("kernels", Davidson and Erwin, 2006) would be most resistant to change since alterations in those patterns would have global consequences, while genes acting lower in the hierarchy might be less constrained, as changes in their expression would modify more flexible features of the pattern. Thus, it should be possible to identify common signatures in distantly related species form *Drosophila*. This hypothesis has received support from a recent analysis of gene expression patterns in the wing and elitron of *Tribolium castaneum* (Tomoyasu et al., 2005, Tomoyasu et al., 2009), which revealed that several key components of the *Drosophila* wing GRN are expressed in comparable patterns and presumably exert an analogous function in *T. castaneum*. Knock down, via dsRNA injection, of genes acting at lower levels of the GRN such as *knirps* and *abrupt* lead to alteration in the venation pattern of the beetle hind wing (Tomoyasu et al., 2005, Tomoyasu et al., 2009, and personal communication) suggesting that these transcription factors were integrated over 500 million years ago into the wing GRN as vein determining genes. Unfortunately, such results are ambiguous with regard to vein patterning as *knirps* RNAi lead to extra vein

tissue instead of expected vein loss phenotype and because the vein pattern in *T. castaneum* has been greatly simplified. This problem is further compounded by the fact that *Tribolium* wing is analogous structure of the fly *halterae* and not the wing. Nonetheless, the findings in *T. castaneum* suggest that it would be fruitful to examine this problem in insects more closely related to *Drosophila* that have different venation patterns since it should be possible to account for these morphological changes by identifying causative alterations in the wing GRN. Since most of the genes represented in Fig. 1B seem to have at least a partially conserved function in *T. castaneum,* it is highly likely that it would be the similarly maintained among Dipterans. This consideration led me to embark on a comparative analysis of the wing GRN by first analyzing gene expression patterns in other Dipteran species with divergent venation patterns (Chapter 2.1). The goal of such an analysis was to identify differentially expressed genes (and their associated CRMs) responsible for the development of analogous, yet morphological distinct structures. A comparison at the regulatory level of such genes could lead to deeper understanding of mechanisms of CRM evolution.

1.3. Study development using the CRISPR/Cas9 tools.

CRISPR, which stands for "clustered, regularly interspaced, short palindromic repeats" is a bacterial adaptable immune response system that protects the cell from invading nucleic acids such as those carried by bacteriophages or plasmids (Barrangou et al., 2007, Wiedenheft et al., 2012). This bacterial immune response system works by integrating fragments of alien DNA in between CRISPR repeats, which are organized in tandem arrays in the bacterial genome. Such regions are then transcribed to generate CRISPR RNAs (crRNAs), which carry fragments of the exogenous sequence (protospacers), fused to a CRISPR repeat. Each one of these protospacer+repeat RNA segments can hybridize with a second trans-activating CRISPR RNA (tracrRNA, Deltcheva et al., 2011) forming and RNA structure that can bind the nuclease Cas9. The resulting complex is a sequence-specific nuclease capable of targeting any DNA sequence complementary to the "protospacer" followed by a protospacer adjecent motif (PAM, which in the case of the bacterium *Sreptococcus pyogenes* is NGG, see Fig. 2) present only in the invading DNA sequence.

This bacterial immune system has been crafted into a genome editing technology that has been developed and adopted at an unprecedented rate due to its efficiency and universal applicability. So far, it has been successfully applied in a wide variety of organisms including diverse animals, plants, and yeast (Sander and Jung, 2014, Hsu et al., 2014, Zhang et al, 2014). Furthermore, mutant versions of the Cas9 protein that lack

nuclease activity have been proposed for other uses such as a programmable sequence specific transcription factors (Mali et al,. 2013).

Applied to genome editing, the CRISPR/Cas9 system is used to generate double-stranded DNA breaks (or nicks) at chosen target sites in the genome. Following cleavage of a genomic sequence by a Cas9/gRNA complex, two main DNA repair mechanisms can restore chromosomal integrity: 1) non-homologous end joining (NHEJ), which typically generates insertions/deletion of a few base pairs at the gRNA cut site, or 2) homology directed repair (HDR), which can correct the generated lesion using a template with homology to the sequences neighboring the break, that are usually provided by the homologous chromosome. It is possible to highjack the HDR repair mechanism by providing an exogenous DNA that would be used by the cell to repair the cut. For this purpose it is possible to use both single-strand oligodeoxynucleotide (ssODN) with 50-100nt of homology on each end, or a double-stranded DNA template with homology arms of at least 500bp for optimal efficiency. It has been shown in *Drosophila melanogaster* that individuals carrying both genomically encoded Cas9 and gRNA sources efficiently mutate the target sequence via NHEJ in the great majority of cells, but can also undergo HDR repair in germline cells when exogenous ssODNs (Gratz et al., 2013) or plasmids with appropriate homology arms (Gratz et al., 2014) are co-injected into the polar plasm of the egg.

Given the adaptability to a new systems and the flexibility of the technology I decided to employ the CRISPR-based genome editing technology to generate regulatory

mutations in *Drosophila melanogaster* altering the transcription of genes in the wing GRN (Fig. 1D, Chapter 3) and to explore the possibility of translating such results into the emerging model organism *Megaselia abdita* (Chapter 2.4).

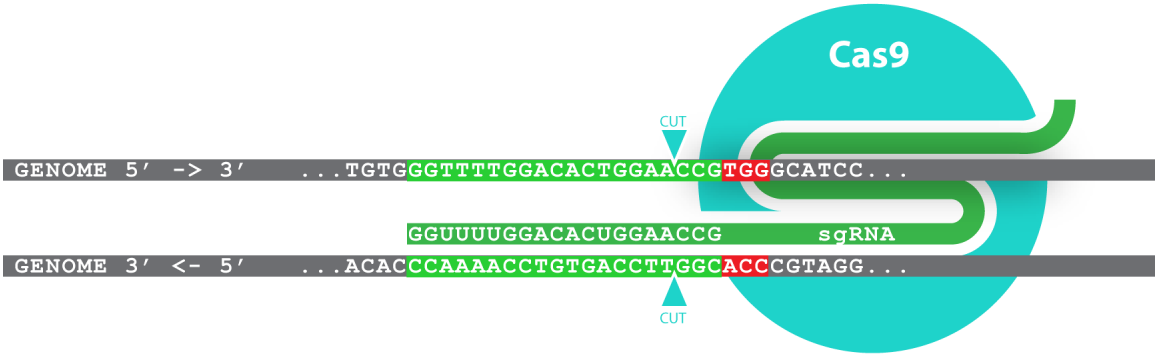

Figure 2 - CRISPR components for genome editing.

Gene editing using CRISPR-based Technology is based on two components: (1) a synthetic guide RNA (sgRNA) which is a fusion of the target sequence (first 20 nt) and the scaffold gRNA which is used to bind the (2) Cas9 nuclease. After recognition of the target site in the genome, determined by correct base pairing with the gRNA, the assembled complex drives double-stranded DNA cleavage 3 nt upstream from the PAM (NGG) sequence found in the genome.

2. Analysis of the *Megaselia abdita* wing gene regulatory network.

An analysis of the wing gene regulatory network (GRN) and cis-regulatory modules (CRMs) associated with it in related species with different wing venation patterns or shape should reveal mechanisms by which developmental systems have adapted during evolution to generate different morphological outcomes.

To gain insight into the evolution of the wing GRN, I chose the scuttle fly *Megaselia abdita* (*phoridae*) as a starting point for a comparative analysis with the well studied fruit fly wing GRN. This choice of species was driven by three main factors: (A) the wing venation pattern is divergent from *D. melanogaster* (i.e. *M. abdita* carries an extra vein that may be M4), in contrast to more closely related species with sequenced genomes, such as *Musca domestica* (house fly), *Glossina morsitans* (Tsetse fly), or *Ceratitis capitata* (mediterranean fruit fly) which have almost identical wing veneation patterns to *D. mel*.. (B) *M. abdita* diverged from *D. mel.* only about 160 Ma, which is significantly closer than other *diptera* such as mosquitoes diverged early in the lineage (250 Ma) during the Permian radiation (Bertone and Wigemann, 2009, Wigeman et al., 2011). The shorter evolutionary distance between *M. abdita* and *D. mel.* increases the chances that CRMs might be identified by sequence similarity or functionally tested for activity in *D. melanogaster*. (C) Laboratory protocols for breeding *Megaselia abdita* have been established (Rafiqi et al., 2011A); its developmental stages and animal morphology are remarkably similar to the ones of *D. mel.* and resources such as transcriptome data

and a genomic BAC library have been previously generated by the efforts of the Urs Schmidt-Ott, Johannes Jaeger, and Pat Simpson labs.

In this chapter I will describe: (1.1) analysis of the gene expression patterns of several *Megaselia abdita* homolog genes known to be involved in the development of the *Drosophila melanogaster* wing, (1.2) the collaborative assembly of the *Megaselia abdita* genome used subsequently to plan (1.3) homologous CRM identification using bioinformatic approaches, and (1.4) functional analysis using the emerging CRISPR/Cas9 technology, to generate analogous mutations to characterized *Dmel* regulatory alleles (e.g.: *radius incompletus, abrupt, vestigial*) resulting in wing vein defects.

2.1. *Megaselia abdita* wing GRN gene expression analysis.

I began my studies by optimizing and conditions for both antibody staining and *in situ* hybridization in early pupal wing primordia of *M. abdita,* when vein stripes are first distinguishable. The decision to analyze gene expression at a slightly later stage than the stage commonly analyzed in *D. mel.* (third larval instar) derived from the observation that *M. abdita* third larval wing discs are extremely small compared to *D.mel.* larvae even if the actual larval size is comparable. Furthermore, I determined empirically that expression of genes involved in venation was only weak and/or sporadic in wing disc stages. This difference between developmental timing in *M. abdita* versus *D. mel.* could be due to a shift of late larval stages in *M. abdita* developmental into pupariation. Indeed, while the larval stage is in both species about 5 days, the pupal stage of *M. abdita* is about 3 times longer than that in *D. mel* (at 25°C).

The first step that I took to optimize dissection and staining of *M. abdita* tissues was to use the Engrailed antibody 4D9, which is well-known to cross-react among arthropod species (Patel et al. 1989), to stain *M. abdita* larval carcasses in order to identify the location and arrangement of the various imaginal discs. Additionally, I proceeded to stain early and late pupal stages with the Engrailed antibody in order to establish a range of time points that could be used for relating *M. abdita* developmental steps to comparable stages in *D. mel.* Furthermore, I tested several antibodies available in the lab hoping for cross-reactivity in *M. abdita*. Among these antibodies only the PS1

(phospho-Smad1) antibody recognizing the phosphorylated form of the Madea transcription factor in organisms ranging from *D. mel.* to mammals (Persson et al., 1998).

After this first optimization step, I proceeded to generate reagents to perform *in-situ hybridization* (ISH) on *M. abdita* tissues. For several genes of the wing GRN I was able to identify the *Megaselia abdita* homologs by BLAST searches of *D. mel.* protein sequences against the newly assembled transcriptome kindly provided prior of publication by Urs Schmidt-Ott and Johannes Jaeger. For genes for which I did not have significant hits, I used multiple sequence alignments from species with sequenced genomes to design degenerate CODEHOP primers (Rose et al. 2003) for PCR amplification of the target coding sequences from cDNA. By such means I was able to successfully amplify coding sequences from the genes: *engrailed* (*en*), *optomotor-blind* (*omb*), *spalt-major* (*salm*), *knirps* (*kni*), *caupolican* (*caup*), *Delta* (*D*), *rhomboid* (*rho*) and *abrupt* (*ab*) and clone them in the dual promoter pCRII backbone provided in TOPO cloning kits. Clones in this vector can be used directly as template for *in vitro* transcription. Since the insert is flanked by the T7 and SP6 bacterial promoters, labeled sense and antisense probes are easily generating by use of the appropriate RNA polymerase whatever the direction of the insert.

So far, I have successfully examined expression of the wing GRN components En, pMAD (a readout of BMP signaling), *salm* and *omb* (two downstream targets of BMP signaling), the vein organizing genes *kni* and *ab*; and the downstream vein markers *rho* (expressed in all veins in *D. mel.*), *caup* (veins L1, L3 and L5 in *D. mel*) and *Delta*

(veins L1, L3-L5 in *D. mel*). The data are summarized in Fig. 3 (figure Fig. 1B can be used as a reference for GRN hierarchy positioning).

Since it was not possible *a priori* to determine which vein stripe observed in the ISH pattern corresponded to which specific adult vein structure, we decided to optimize fluorescent *in situ* hybridization (FISH) to perform dual probe staining and thereby map the vein primordia observed in the ISH stains (e.g., *kni*, *caup*) to the adult wing structures and to the corresponding veins in *D. mel.*.

From the stains I generated it is possible to conclude that *kni* is expressed in two stripes that run along both the anterior and posterior borders of the *sal* expression domain (Fig. 3). These veins presumably correspond respectively to L2 and M4, which most likely form between the L4 and L5 homologous structures of *M. abdita* (Fig. 1C, Fig. 3 and Fig. 4). This finding supports the previously proposed hypothesis that the reduced number of veins in *D. mel.* relative to the ancestral venation pattern reflects silencing of vein formation along quiescent vein inducing borders (or paraveins) in *D. mel.* (Biehs et al. 1998). The paravein hypothesis contrasts with the traditional view in which reduced vein patterns reflect the fusion of ancestral veins or the deletion of entire sectors of the wing between particular veins (Hennig, 1954, Garcia-Bellido and De Celis, 1992). A second conclusion that could be made is that *rho*, *caup* and *Delta* are expressed in stripes likely to correspond to the same veins as in *D. mel.*, suggesting that this tier of the wing GRN is also well conserved across *Diptera*.

While further analysis is needed in order to generate a more comprehensive comparison between *Drosophila melanogaster* and *Megaselia abdita* wing GRN, I have successfully generated several high quality *in situ hybridization* probes, while others could likely be improved by cloning larger genomic coding fragments. These tools would permit additional dual FISH stains to be performed that refine and deepen the developmental comparison between the two species. These comparative expression studies set the stage for functional analysis of components of the wing GRN in *M. abdita.*

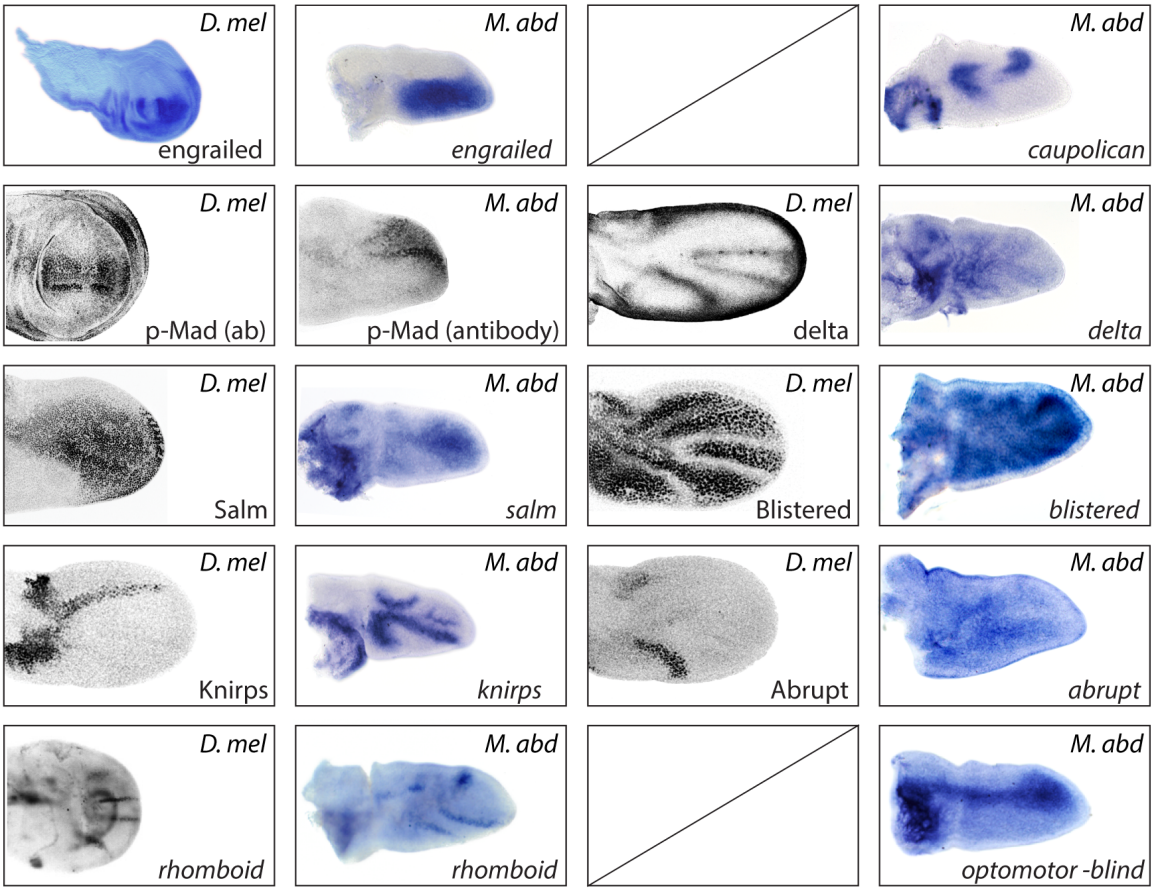

Figure 3 - Gene expression comparison between *D. mel.* and *M. abdita.*

This figure shows the results of the gene expression analysis that I have conducted in *M abdita* and how it relates today known *D. mel.* patterns. The first and third columns displayed expression patterns of *D. mel*, while the second and fourth display the corresponding expression patterns of *M. abdita*. All the *D. mel*. stains, except for *rhomboid*, are antibody stains. While for *M. abdita* the stains are all *in-situ* hybridization except for the p-Mad stain (antibody).

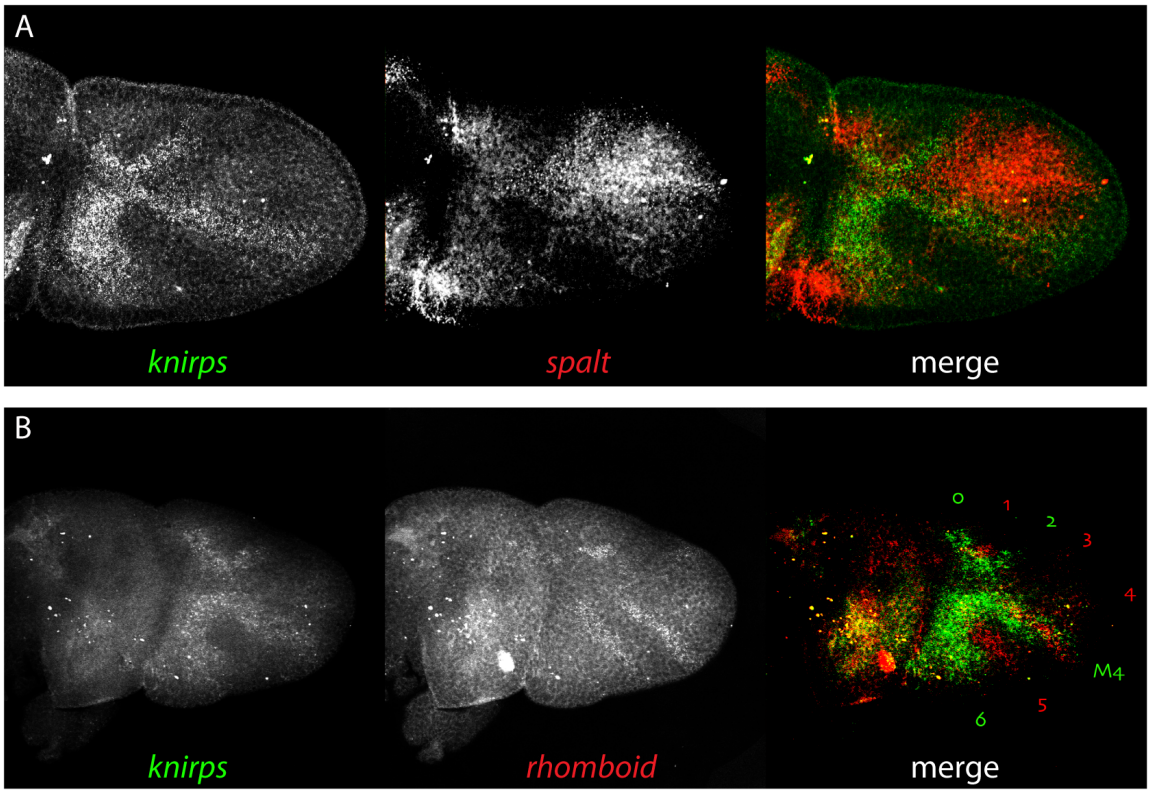

Figure 4 - ***M. abdita*** **FISH stains of the vein determining gene** ***knirps.***

This figure displays the result of FISH stain to understand the topological relationship between the expression of *knirps* with the ones of *spalt* (A) and *rhomboid* (B). As it is possible to see in panel (A), *kniprs* is associated with *spalt* borders, and contrarily to *D. mel., knirps* is clearly expressed in a posterior branch associated with the posterior *spalt* edge, suggesting that in this specie it is responsible for the determination of a vein in the posterior compartment (M4). To support this hypothesis I generated a double stain with the downstream factor *rhomboid* which is known to be expressed in all vein primordia in *D. mel*. As it is possible to see in (B) the posterior branch of *knirps* is indeed associated with *rhomboid* expression (M4).

2.2. *Megaselia abdita* genome assembly.

In order to identify the cis-regulatory modules (CRM) responsible for the observed gene expression patterns observed in *M. abdita* I decided to obtain genomic sequence that could be analyzed by sequence similarity or used for computational CRM prediction. Fortunately, the Schmitdt-Ott Lab (Univ. of Chicago) and the Jaeger Lab (CRG Barcelona, Spain) had already begun assembling a set of newly sequenced *M. abdita* Illumina reads. I joined their efforts establishing a collaboration in which the Bier lab provided an additional set of genomic reads obtained using the novel PacBio sequencing technology. This technology was chosen because the genomic reads generated are in the order of 2-10 kbp, therefore extremely useful for scaffolding of shorter reads such as the one from Illumina. The drawback of this technology consist in its highly error-prone output, a problem that in turn is overcome by comparing the raw PacBio reads to a 20X coverage set of the high fidelity of the shorter reads.

While we successfully contributed 3.5X long-read coverage to the *M. abdita* genome project, the improvement of the genome assembly obtained after the implementation of PacBio sequencing data was only modest. The limitation in getting longer assemblies is most likely due to long repeated sequences found in the *M. abdita* genome. Nonetheless we were able to partially exploit the final *M. abdita* genome assembly for subsequent CRM discovery and analysis.

For this part of the project I would like to acknowledge and thank Long Do for providing his bio-informatics expertise and to perform error correction and filtering of the PacBio reads, which took a tremendous amount of effort and computation time, and involved a sophisticated partitioning of the correction process into manageable segments that were then integrated to create an accurately corrected PacBio output.

2.3. *Megaselia abdita* candidate CRM identification.

The first step that I took towards the identification of homologous CRM from other Dipteran species was to generate alignments of several of the known wing CRM previously identified in *D. mel.* (de Celis et al., 1999, Sivasankaran et al. 2000, Lunde et al. 2003, Barrio and de Celis, 2004) with genomic sequences from the 12 *Drosophila* species that have been sequenced (Clark et al. 2007). Through this analysis, I was able to identify, for each CRM, a few hyper-conserved stretches of DNA sequences that approach 100% identity. I operated under the assumption that such hyper-conservation in sequence should reflect functional constraints refractory to change during evolution.

In order to identify CRM from flies at the higher evolutionary distance from *D. mel.* I used the BLAST tool to find matches of such hyper-conserved sequences in genomes of more divergent species such as *Musca domestica* (*M. dom.,* house fly), *Glossina morsitans* (*G. mor.*,Tsetse fly), or *Ceratitis capitata* (*C. cap.*, mediterranean fruit fly). Remarkably, in most cases, I was able to identify almost perfect matches to such hyper-conserved regions using simple BLAST searches. In cases where I had genomic scaffolds covering large enough contigs, I could confirm that the identified hit, was indeed positioned in close proximity with the expected regulated gene. In Fig. 5A I have summarized the identified CRM homologs from such species and provided an example of graphical alignment generated with Jalview for the *knirps* CRM homologs. To the best of my knowledge the occurrence of such hyper-conserved DNA sequences has not been

reported. While such sequences are likely to be the product of constraints dictated by two or more physically interacting transcription factors essential for a essential developmental process, experimental evidence to confirm this tentative hypothesis is needed.

The species *M. dom, G.mor*. and *C.cap*. from which I successfully identified CRM homologs are all members of the *Schizophora* branch of the order *Diptera*. The *Schizophora* radiated in the early Paleogene about 60-70 Ma (Bertone and Wigemann, 2009, Wigeman et al., 2011). I attempted the same BLAST-based approach to the one described above to similarly identify CRMs from species more distant than *Schizophora* such as the mosquitoes *Anopheles gambiae* and *Culex Pipiens* (*culicidae* diverged from *D. mel.* 250 Ma) for which high quality genomic data is available and has been extensively annotated. As perhaps would be expected in these cases of deep evolutionary divergence, I was unable to identify any sequence similarity by simple BLAST search.

Megaselia abdita diverged from *Drosophila melanogaster* about 165 Ma and is thus positioned at an intermediate evolutionary distance between *Schizophora* and mosquitoes. Since I could not predict *a priori* whether it would be possible to identify *M. abd.* CRMs by simple BLAST searches, we decided to take advantage of published algorithms for efficient CRM identification using sequences from multiple species. We therefore collaborated with Dr. Saurabh Sinha (Univ. Illinois), an expert in CRM discovery. Dr. Sinha has devised sophisticated algorithms using a motif-blind algorithm tuned to identify regulatory sequences using known *D. mel.* CRMs as a training set for the algorithm (Kantorovitz et al., 2009). I provided Dr. Sinha with a training set from the

identified *Drosophilids/Schizophora* CRMs for the gene *knirps*. He first used the *Drosophilids* CRMs to test his algorithm on the *Schizophora* genomic sequences, from which he found that the CRMs that I previously identified where displayed as high scoring regions. Once the algorithm was validated and optimized, Dr. Sinha used both the *Drosophilids* and *Schizophora* CRMs to predict candidate CRM sequences in the corresponding *A. gambiae* loci. Indeed Dr. Sinha successfully identified high probability CRM sequences in the *A. gambiae* region upstream from the *kni* coding region (the same relative position to the coding region as the *knirps* wing CRM does in *D. mel.*) that were interestingly enriched in binding sites for the wing transcription factor *Scallopped* (*Sd*).

I have successfully generated a pipeline to apply Dr. Sinha's motif-blind algorithm to predict CRMs in the *M. abdita* genome or other sequenced *Diptera*. For regions that are not yet fully assembled (e.g., *kni*), I obtained as backup a BAC library that could be screened using a probe from the coding sequence. Sequencing the identified BAC clones should provide enough sequence to cover all the regulatory region of the genes of interest and move the project forward. Furthermore, I have established that for low divergence genomes, small strings of hyper-conserved regulatory DNA (see above) could be successfully used for CRM identification in novel genomes including *M. abdita*.

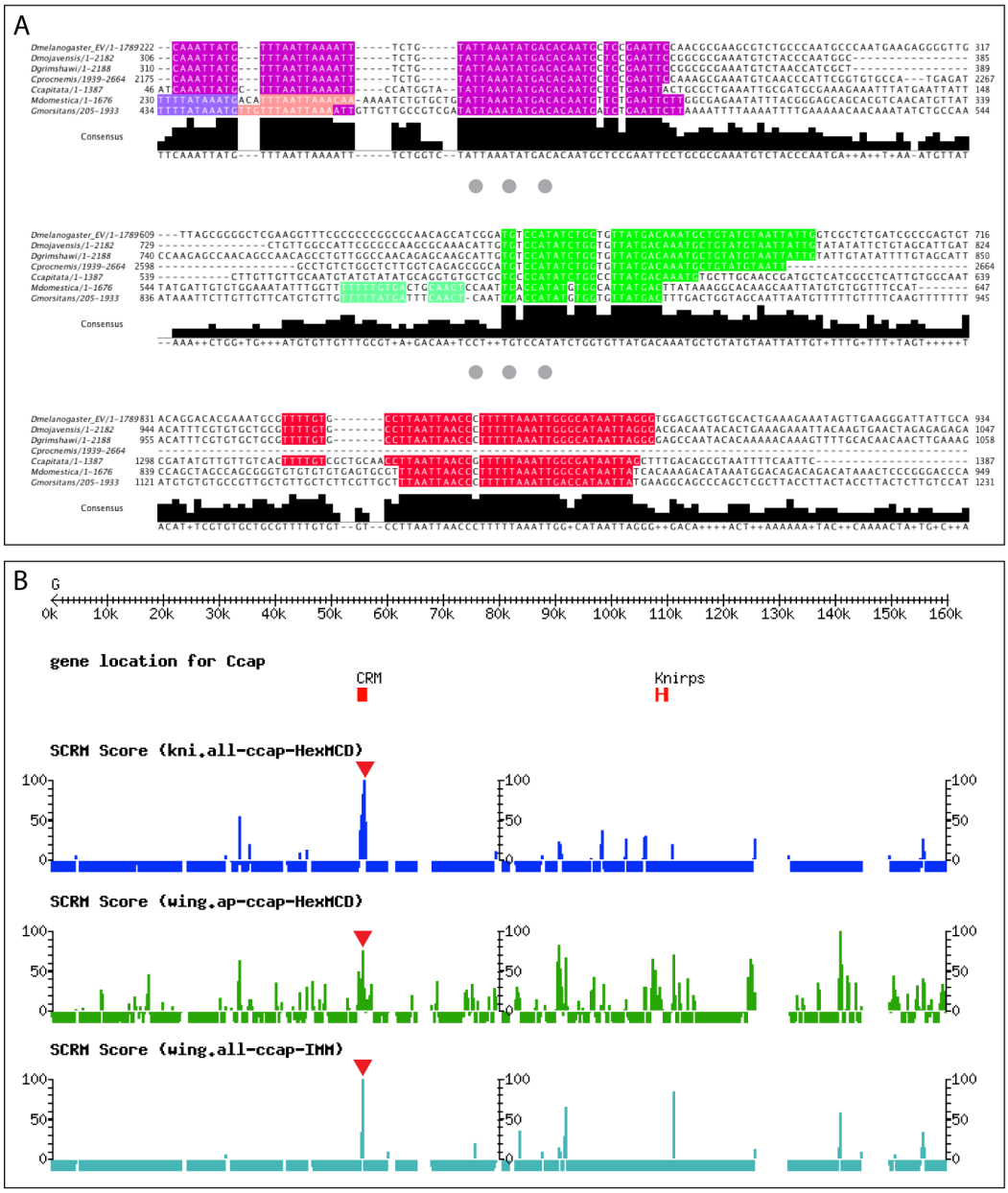

Figure 5 - CRM prediction algorithm training and testing.

(A) multiple sequence alignment of the knirps CRM sequences identified by BLAST searches. (B) initial testing of Dr. Sinha's algorithms reliably identifying the knirps CRM (▼) when using different algorithm training sets (kni.all: *D. mel knirps* CRM, and all the identified CRM homologs, wing.ap: all the identified wing CRMs expressed along the Anterior/Posterior axis and their homologs, wing.all:all the identified wing CRMs and their homologs; -ccap: indicated that the *Ceratitis capitata* sequences were omitted from the training set since the algorithm was run on that species' *knirps* genomic locus).

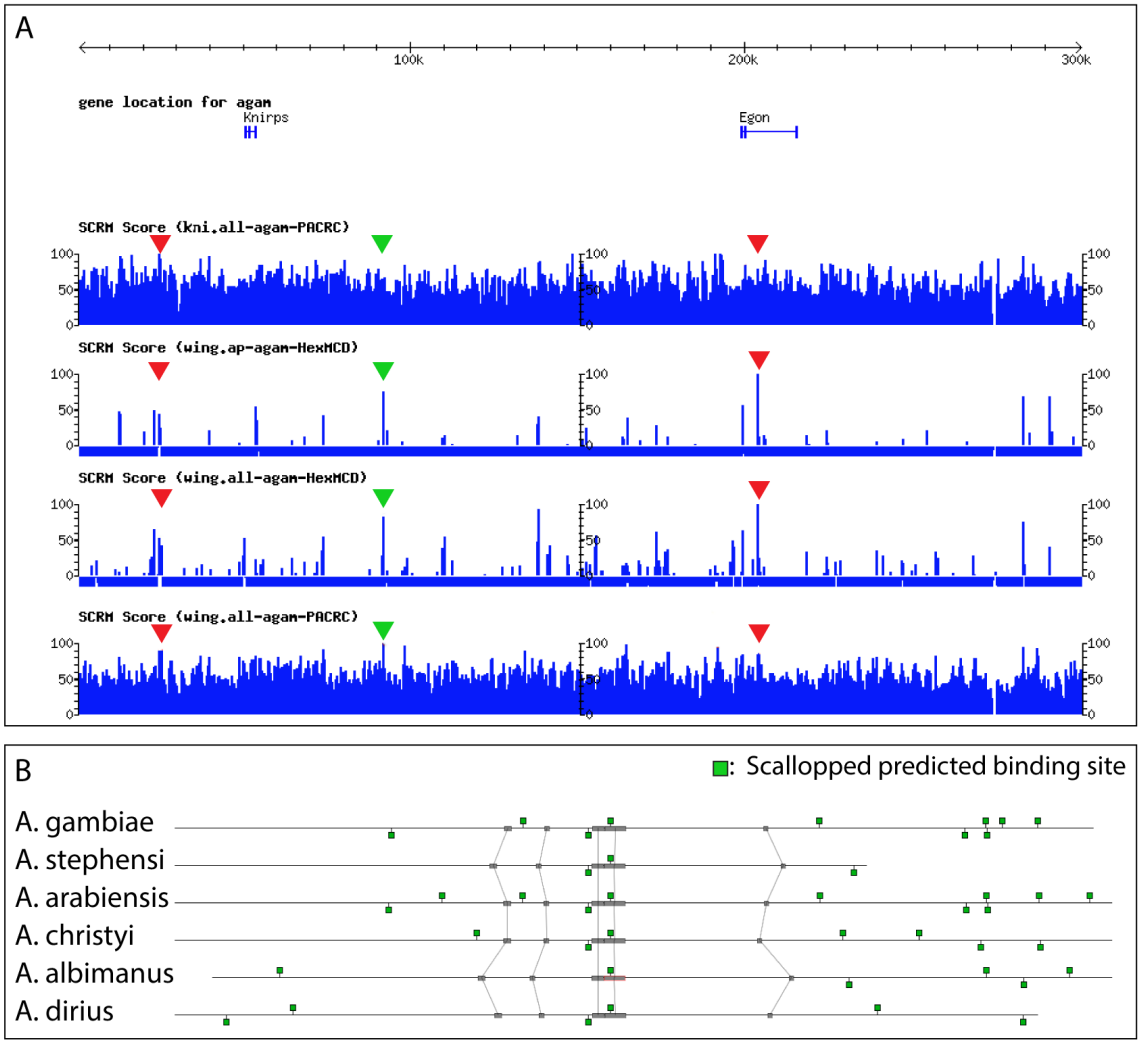

Figure 6 - Prediction of candidate CRMs in the distantly related *A. gambiae.*

(A) shows the output tracks of Dr. Sinha's algorithms when run on the *A. gambiae knirps/ egon* locuses (*egon* is a close homolog of *knirps*), which identify high scoring candidates (▼). By closely looking at these candidates I identified the second one (▼) to be the favorite for future reporter or genome editing analysis. This choice was made on one hand because the sequence (▼) falls in the same relative position and comparable distance respect to the *knirps* coding sequence than in *D. mel*. And on the other hand because (▼) displays an enrichment of *Scallopped* (*Sd*) transcription factor predicted binding sites (B)(■) conserved among related *Anopheles* species.

2.4. *Megaselia abdita* genome editing with the CRISPR/Cas9 system.

After having successfully established the expression patterns for core wing GRN genes in *M. abdita* by *in-situ hybridization*, I considered strategies to obtain functional information that would support the gene expression data. I reasoned that a functional analysis could be conducted in two different ways: (A) by generating transgenic *D. mel*. strains that would carry exogenous CRM sequences driving the expression of a reporter gene such as GFP or (B) by editing the genome of *M. abdita* and creating mutations in regulatory sequences that would abolish the function of putative CRMs, analogously to classical regulatory alleles such as *knirps*$^{ri\text{-}1}$ (*ri = radius incompletus)* which eliminates *knirps* expression in the wing primordium and results in truncated adult L2 veins.

The first of these two strategies (A) would rely on the assumption that the exogenous CRM would be active in *D.mel*. and would respond to the same set of transcriptional inputs that are present in its original genomic background. Since such CRM from distantly related species might not respond to *D. mel.* trans-acting factors, they might not drive any expression, or they might lead to uninformative non-specific expression. For this reason, I thought I might be best to test this approach using CRMs from the more closely related *Schizophora* species for which I identified candidate CRM sequences. Such an analysis in *Schizophora* would provide less information on the GRN evolution, since these species have identical the venation pattern to the one of *D. mel*., however, it might be possible to discern differences in the relative positions of veins since

the shapes and sizes of the wings vary significantly among this group. Since approach (B) depends only on identifying high quality candidate CRM sequence, this method might be expected to yield, if successful highly informative results in more distantly related species and is therefore the approach to which I have devoted the greatest effort.

To create mutants in candidate CRMs from *M. abdita*, I decided to use the emerging and highly efficient CRISPR/Cas9 genome editing technology. As summarized in the introduction, this remarkable technology has been successfully employed in species ranging from yeast to humans. In this method, guide RNA (gRNA) targeted cleavage by the Cas9 endonuclease generates double strand DNA (dsDNA) breaks at high rates. Such breaks are then either repaired by an error prone pathway (Non-homologous End- Joining or NHEJ) or by a highly precise homologous repair pathway (Homology directed Repair or HDR). Other than in the fruit fly, CRISPR/Cas9 has been successfully used in insects including the mosquito *Aedes aegypti* (Kistler et al., 2015) and the silkworm *Bombyx mori* (Wang et al., 2013) suggesting that the system would be highly efficient also in *M. abdita.*

The first step that I took to optimize this technology in *M. abdita* was do identify optimal candidates to target for genome editing. The choice of target was guided by the idea that creating visible phenotypes would be easier to screen, I therefore gave precedence to X-inked recessive mutations or autosomal dominant mutations. I successfully identified the *M. abdita* homologs of the *D. mel.* gene *white*, *yellow*, *vestigial* and *stubble*. *white* and *yellow* because I reasoned that as in *D. mel*. they might

be be X-inked, for *vestigial* because I could identify a high probability candidate for the homologous wing CRM with a hyper-conserved island which if mutated should result vestigial wings without affecting viability. For *Stubble,* I targeted a specific region which should create a truncated protein, mimicking the well-known *D. mel*. dominant marker *Stubble*, which displays shortened bristles.

I choose two different target sequences for each one of the four candidate genes. I generated the gRNAs by in vitro transcription using a PCR template similar to the one described by Shao et al. in 2014, and used commercially available capped Cas9 mRNA. After optimizing the collection of *M. abdita* embryos, I proceeded to inject them with combinations of gRNA and Cas9 mRNA water solutions (Rafiqi et al., 2011B). After obtaining an extremely low survival rate even in control water-only injections, I decided to employ the PCR-based T7 endonuclease assay (Kim et al., 2009) to screen embryos for successful Cas9 activity 24h after the injection. While I was able to obtain encouraging preliminary results suggesting some degree of cleavage at the target site, I realized that the relatively low efficiency of this cleavage combined with the extremely low injection survival (5-10%), and the difficulty of retrieving mutations not immediately associated with a phenotype (e.g., recessive alleles), would mean an extensive period of technique optimization, which could ultimately lead to failure. This reasoning guided me to change my strategy. First, I decided to move to a DNA-based approach, using both sgRNAs and Cas9 encoded on a plasmid for the injection mix. Second, I devised the mutagenic chain reaction (MCR, see chapter 4) a Cas9-based active genetic cassette capable of converting

heterozygous to homozygous mutations (Gantz and Bier, 2015), which should, in principle, dramatically simplify the recovery of recessive mutations.

As mentioned above, in order to increase the efficiency of the gene editing system I decided to generate plasmids that would use endogenous *M. abdita* gene regulatory sequences to drive the expression of Cas9 and gRNAs in *M. abdita*. This plasmid-based approach has been shown to be much reliable in *D. mel.* than simple RNA injection. In order to mimic the strategy previously used in *D. mel*. (Port et al., 2014) I cloned 4 kbp of regulatory sequence upstream of the ATG of the *M. abdita Actin5C* gene and fused it to the coding sequence of Cas9. Similarly, I identified the *M. abdita U6 snRNA* gene locus and created plasmids that would express sgRNA under the control of the *M. abdita* U6, as it has previously been done to create *D. mel.* reagents (Gratz et al., 2013). I successfully built such expression plasmids and cloned the 8 target sequences previously chosen (2 for each gene) into them.

At this point in of the project I paused the injection experiments to focus on the early results obtained with the MCR technology since this second approach turned out to be more successful than might have been anticipated and, as discussed below, lead to my subsequent publication of a manuscript on this new technology (Gantz and Bier, 2015). Nonetheless, all the reagents for a Cas9-based genome editing of *M. abdita* are ready for the next step on this project. A slight improvement of the approach will be to additionally target the *M. abdita* homolog of the *D. mel. cinnabar* gene, also known as *Kynurenine 3-monooxygenase*, and known to be responsible for eye pigment formation in mosquitoes

and resulting in white eyes if mutated (Bhalla, 1968, Han et al. 2003), this gene may be an ideal candidate for the optimization of Cas9-based genome editing in *M. abdita*.

In summary, I have developed the necessary tools for functional analysis of *M. abdita* CRMs. These methods include: 1) optimized dissection, antibody staining, and *in situ* hybridization protocols in developing *M. abdita* wings, 2) generation of CRISPR/ Cas9 mutagenesis tools for *M. abdita* for the future generation of regulatory alleles (Cas9 and sgRNA expressed from *M. abdita* regulatory sequences) and 3) a one-step mutagenesis method using the mutagenic chain reaction (MCR) described in detail in chapter 4.

3. Targeting of the *Drosophila melanogaster knirps* wing CRM with CRISPR/ Cas9.

In order to embark on a CRISPR/Cas9-based approach to edit gene regulatory regions in *M. abdita*, I first decided to test and optimize this strategy in *D. mel*. Given the expression data I had generated on wing GNR components in *M. abdita*, (see Chapter 2) and the prior research into formation of the L2 vein in *D. mel.* that had been done in the Bier lab (Lunde et al.,1998, Lunde et al., 2003), I choose to target the well-studied *knirps* wing CRM. I reasoned that data obtained from such an analysis could also be used to further understand the information encoded in a CRM. Given the observation that in *M. abdita* the gene *knirps* is expressed in two stripes along both anterior and posterior borders of the *spalt* expressing domain, and that *D. mel.* displays weak rudiments of *kni* expression along the posterior *spalt* border (Fig. 3), I reasoned that this approach could possibly identify mutations that would remove a putative posterior repression binding site, therefore restoring the M4 vein lost in *D.mel.* during evolution (Fig. 1C). In addition, having a more refined model for the *kni* L2 CRM should help inform the types of experiments to be designed in *M. abdita*.

3.1. Reproducing the classical *knirps* allele *"radius incompetus"* with CRISPR/ Cas9.

The approach that I took to generate regulatory alleles of the *D.mel. knirps* wing CRM resulting in a phenotype analogous to that of the $kni^{ri\text{-}1}$ allele was twofold. Firstly, I decided to generate a 2.2 kb deletion that would cover the 1.4 kb EX fragment identified by Karen Lunde (Lunde et al., 2003), along with adjacent sequences which I believed relevant due to sequence conservation among *Schizophora* (Fig. 5A). Secondly, I decided to target one of the hyper-conserved regions identified by sequence alignment among *Schizophora* (Fig. 5A), which is known is known to be deleted in $kni^{ri\text{-}1}$ and altered in sequence in the kni^{ri53j} allele (Lunde et al., 2003) with the hope of generating small indels (insertion/deletions) that would span such region. I hoped that these two strategies would complement each other. On the one hand, deletions (particularly large ones such as a 2.2kb) could abolish the CRM function, but at the same time, might also result in a homozygous lethal mutation. On the other hand, small indels should be viable but might not generate strong visible phenotypes.

For the deletion analysis I generated 2 gRNAs targeting the sequences, referred here as A and B, chosen to target non-conserved regions flanking the EX fragment, previously defined in the Lunde et al (2003) reporter analysis, to generate a complete CRM deletion. In contrast, targeted mutation of the hyper-conserved region using the single gRNA targeting sequence C represented in figure 7A was hoped to generate small

indels in a region known to be critical to CRM function (i.e., the single bp C->A mutation in *kni*^{ri53j} is in the immediate vicinity of the C-cleavage site). The A, B and C target sequences gRNAs where cloned into the pU6-BbsI-chiRNA vector (Gratz et al., 2013). The pVG051-gRNA-A and pVG052-gRNA-B constructs were co-injected along with the single-strand bridging oligo deoxynucleotide (ssODN) vg138 that comprised 120 bp homology sequence flanking the sequence intended o be deleted. I hoped that including this bridging oligo would increase the frequency of clean deletions mediated by homology directed repair (HDR) between A and B, given the ssODN sequence homology with the first 60nt of each genomic end left after excision of the AB fragment. The pVG050-gRNA-C was instead injected alone to promote error prone repair via non-homologous end joining (NHEJ). In both cases the DNA mixes where injected into a fly stock carrying a germline-specific Cas9-expressing transgene driven by the *vasa* regulatory region (Gratz et al. 2013). After eclosion the injected animals were crossed to a *kni*$^{ri\text{-}XT2}$ / TM3, *kni*$^{ri\text{-}1}$ stock to facilitate phenotypic recovery of alleles by screening for L2 vein loss phenotypes (*kni*$^{ri\text{-}XT2}$ is a deletion of most of the *kni* regulatory region). This method was quite efficient as 10% of the AB and 12% of the C injected animals yielded offspring with truncations of the L2 vein. Remarkably half of the recovered mutations from the AB injection carried the precise deletion between the the A and B target cut sites. As shown in figure 7B, flies carrying a homozygous AB mutation displayed an extreme *ri* phenotype, equal to that reported for *kni*$^{ri\text{-}XT2}$ homozygotes (Lunde et al., 2003). In the case of the sgRNA-C, as expected, I was able to recover several small

indels covering the hyper-conserved region that resulted in gaps in the L2 vein. Along with the expected indels, several deletions in the 0.5-1.4 kb range were also recovered. This last set of mutations may have been generated by the non-canonical single-stranded annealing (SSA) DNA repair mechanism (Ivanov et al., 1996), which generates deletions between short repeated sequences.

The success of these first editing attempts of the *knirps* CRM lead me to conceive of a much more detailed functional dissection of this CRM. In the next section (2.2) of this chapter I will outline how Shannon Xu, a master student in the lab whom I mentored, and I worked together to push this functional analysis further.

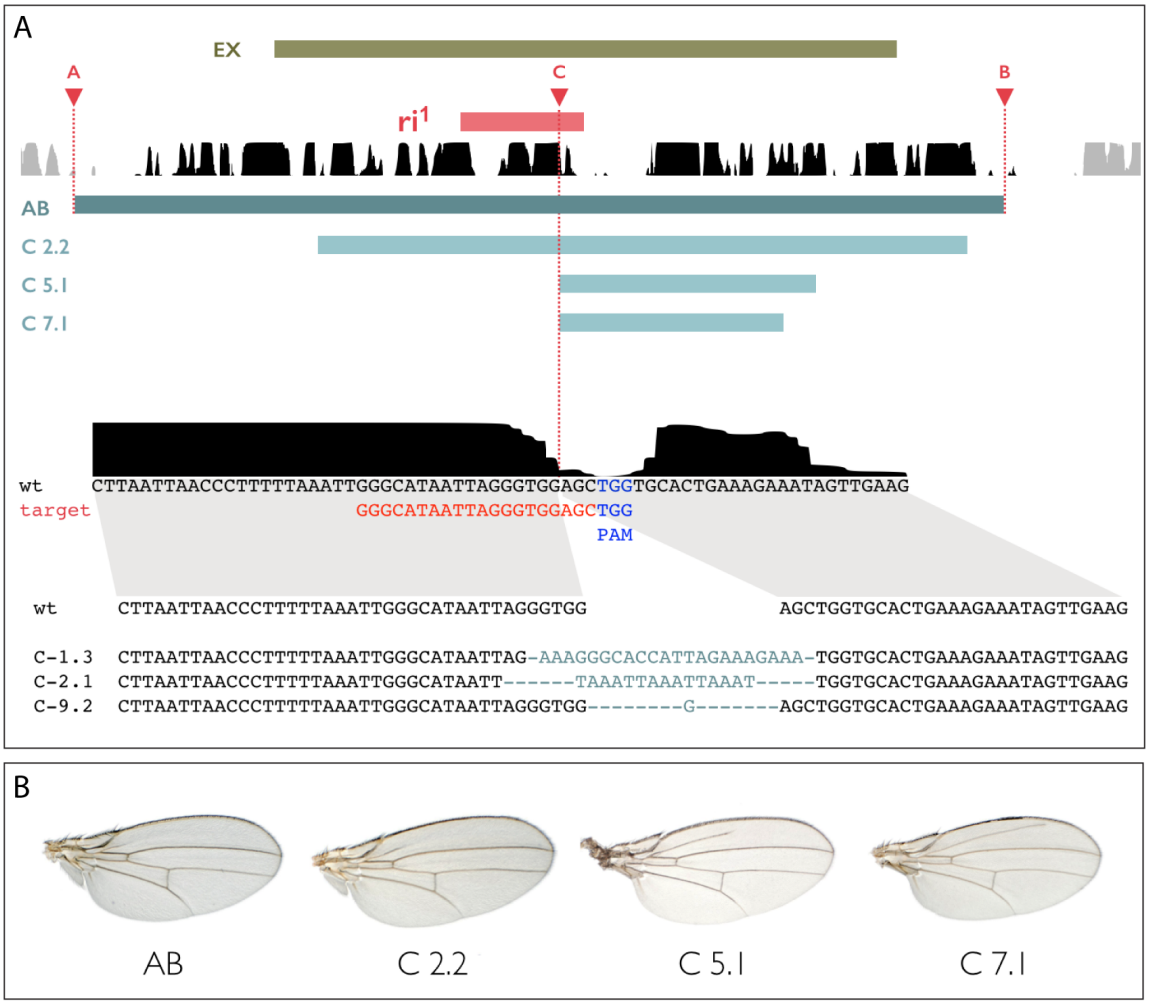

Figure 7 - Editing the *knirps* wing CRM with CRISPR.

(A) Schematic of the *knirps* wing L2 CRM, the position of the A, B and C sgRNAs (▼) used in my study. The EX fragment represented on top indicate the position of the published *knirps* minimal L2 CRM (Lunde et al. 2003), along with a known deletion associated with the knirps$^{ri\text{-}1}$ (ri^1) mutation. Indicated with pale blue horizontal bars (■) and sequences below are some of the mutations obtained with their injections of the A/B and C sgRNAs. (B) shows the phenotype range of some of the recovered mutants; the AB deletion is associated with an extreme *ri* phenotype lacking the entire L2 vein.

3.2. Functional analysis of the *knirps* wing CRM using CRISPR/Cas9.

The *D. mel. knirps* wing CRM most likely contains a combination of separable functional elements that perform different transcriptional activities. In the Lunde et al. 2003 publication several results were reported that support this possibility:

- The minimal 1.4 kb L2 *knirps* enhancer, called EX fragment, displays ectopic expression in the posterior compartment of the wing disc, and so does the 5kb E fragment that extends further upstream relative to coding sequences of the gene. This observation suggests that part of the regulatory region responsible for posterior repression should lie more proximally to protein coding sequences. Additionally, one of the hyper-conserved sequence that I have identified lies just outside the proximal border shared by the E and EX fragments.
- The EC fragment, a distally truncated version of the EX sequence, drives ectopic reporter gene expression in both the distal anterior and the posterior portions of the wing disc. This suggests that the distal 0.7 kb of the EX fragment carries binding sites for transcriptional repressors, and that this portion of the *knirps* CRM could have a separable suppressor function. In turn the EC should contain most of the activating transcription factor binding sites.
- When the single C → A point mutation, identified in the $kni^{ri\text{-}53j}$ allele, is inserted on the EX reporter construct, the central wing disc expression is abolished (L2) while

peripheral expression associated with the L1 and L6 vein primordia is retained. These results suggest that two different activating transcription factors may act independently on the CRM.

- Lastly, a repression input on the CRM, determined by the transcription factors *salm/ salr* (expressed in the central wing disc domain), is likely to act on the EC fragment. This is supported by the fact that none of the reporter constructs generated have ectopic expression in the central region. During my studies I have obtained additional evidence supporting this point. For example, when a *salm*-RNAi construct is expressed in the wing, *knirps* expression and the L2 vein move posteriorly into the *salm/salr* domain (data not shown).

Given the above and other considerations, Shannon (Xiang-Ru) Xu and I crafted a plan for a much more detailed CRISPR/Cas9 deletion scheme of the *knirps* L2 regulatory region. We subdivided the AB region into four segments to be individually deleted (Del1-4). Each of the Del1, 2 and 3, was meant to remove one of three hyper-conserved regions outlined in figure 8. We predicted that the Del1 should abolish the posterior repression causing ectopic *knirps* expression in the posterior compartment, and therefore extra vein tissue. Similarly we predicted that Del2 might abolish repression in the central domain leading to vein tissue formation in between L2 and L5. We also predicted that Del-4 would have an opposite phenotype associated with extra veins anterior to L2 and

posterior to L5. The deletion Del3 would remove about 200 bp around the *kni*$^{ri\text{-}53j}$ allele, functioning as a positive control that was expected to generate an *ri* mutant phenotype.

We injected sgRNA mixes as described above for generating the *kni*^{AB} allele. Since we did not know *a priori* whether a given deletion would display a phenotype, we devised a practical protocol to quickly screen positive injectants without relying on a visible phenotype. Injected animals where first crossed to the homozygous *kni*^{AB} stock, after 4 to 5 days, 20 wandering larvae where pooled from each vial and DNA was prepared for PCR analysis. These DNA preps where used as a template for PCR amplification of the region containing each deletion. An advantage of recovering candidate mutations in trans to the *kni*^{AB} was that successful amplification of a DNA product could only derive from the mutated chromosome inherited from the injected parent. Also, by using the shortest elongation time possible for the PCR elongation step, we biased amplification of the shorter template (deletion) over the longer one (wild type) which aided in the recovery of deletions even if they were underrepresented in the DNA pool. This strategy proved successful in reliably identifying positive vials. The offspring recovered from such positive vials were screened for wing phenotypes if present, or crossed to a TM3/TM6 double balancer (~30 single crosses per positive vial) to recover the mutation in homozygosity in other cases.

We successfully obtained all the four planned deletions, which displayed several unexpected phenotypic features:

- Our positive control Del3 displayed an *ri* phenotype in combination with the AB deletion (kni^{Del3} / kni^{AB}), but surprisingly did not do so when homozygous (kni^{Del3} / kni^{Del3}).
- Both deletions Del1 and Del4 had no defects in trans to the kni^{AB} full CRM deletion (kni^{Del1} / kni^{AB} and kni^{Del4} / kni^{AB}). However, as homozygotes, we observed mild extra vein phenotypes with low penetrance as we had predicted. One potential explanation for these observations is that under normal circumstances homologous chromosomes may pair in somatic cells such that each allele aids in the cis-acting assembly of protein-CRM complexes on the sister chromosome.
- Del2 displayed a consistent *ri* phenotype, with the exception of the deletion $kni^{Del2.01}$ which again does not show any phenotype when homozygous. In addition to the 402 bp deleted in the correct isolate, this mutation carries only an additional -8 bp deletion and +4 bp insertion, which for some yet unknown reason can dramatically alter its function. Thus, the precise endpoints of the deletion have a very significant impact on the resulting phenotype.
- Lastly looking at the overall collection of alleles generated we noticed that different alleles affect distinct segments of the L2 vein along the proximal/distal (P/D) axis, suggesting that P/D inputs act on the *knirps* L2-CRM.

Altogether this round of deletions of the *knirps* wing L2 CRM was extremely successful leading to some expected results while other effects of these mutations were

unanticipated. Importantly, these experiments targeting the L2 CRM in its native chromosome environment have revealed important new regulatory features of this sequence than were obtained previously using standard CRM-reporter fusion constructs. These newly revealed features are mainly due to the much greater sensitivity of the phenotypic readout. Additionally such an approach allows one to carefully analyze cis-regulatory function in the CRM's "natural" environment without the inevitable bias of positional effects.

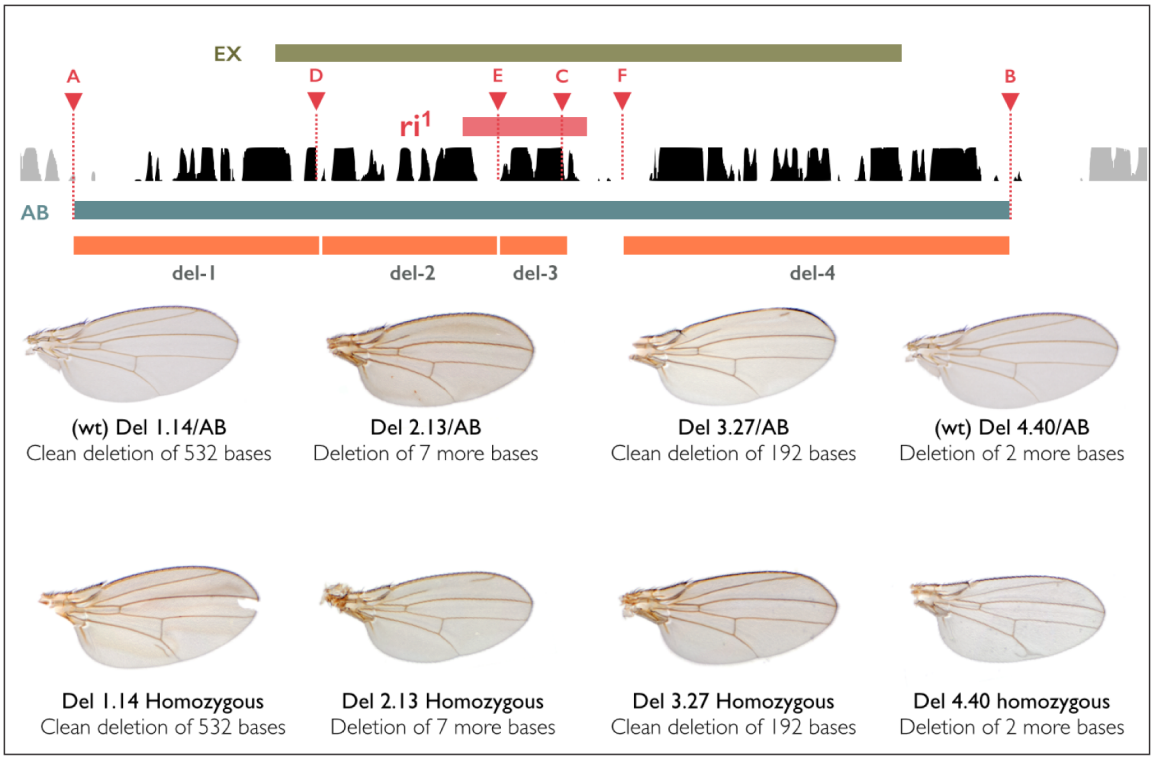

Figure 8 - Functional dissection of the *knirps* wing CRM with CRISPR.

Schematic of the CRISPR-based deletion analysis conducted by Shannon Xu and me. We targeted for deletion four different regions (del1-4) covering most of the previously published knirps CRM (EX fragment, Lunde et al. 2003). The obtained alleles have all associated wing phenotypes: while only del-2 shows a phenotype in combination with $kni^{ri\text{-}AB}$ all four deletions display phenotypes when in homozygous condition (bottom row). As we predicted del-1 and del-4 do not affect negatively the L2 vein, while displaying mild extra vein tissue associated with both the anterior (L2) and posterior (L5) *spalt* borders (refer to figure 1A for vein nomenclature).

4. The mutagenic chain reaction (MCR): a Cas9-based gene drive.

As previously mentioned in Chapter 2, I had encountered technical difficulties in creating mutant alleles in *M. abdita* using standard Cas9/gRNA methods. To overcome these problems, I conceived of an alternative mutagenesis strategy designed to convert otherwise heterozygous recessive alleles to homozygosity. This approach was intended to create cis-regulatory mutations in *M. abdita,* which as soon as they were created in a heterozygous condition, would automatically convert the sister allele to the same condition resulting in a homozygous mutation. Such auto-catalytically acting mutations should both greatly facilitate recovery of mutations based on direct scoring of the phenotype in G1 transformant individuals and aid in the downstream maintenance of stocks (no need for balancer chromosomes).

I reasoned that an autocatalytic genetic behavior could be achieved if insertional mutants were generated using a construct having three components: 1) a central segment encoding the Cas9 gene (expressed in both somatic cells and the germline), 2) a gRNA targeted to a genomic sequence of interest, and 3) homology arms flanking the Cas9/ gRNA cassette that match the two genomic sequences immediately adjacent to either side of the target cut site (Fig. 9). Such a tripartite construct should result in Cas9 cutting the genomic target at the site determined by the gRNA followed by insertion of the Cas9/ gRNA-bearing cassette into that locus via homology directed repair (HDR) triggered by the homology with the flanking sequences (Fig. 9B). Expression of Cas9 and the gRNA

from the insertion allele should then lead to cleavage of the other allele (Fig. 9D) followed by HDR-driven copying of the Cas9/gRNA cassette into the companion chromosome (Fig. 9E-F). In analogy to the polymerase chain reaction (PCR), which doubles DNA templates following each chemical amplification cycle, we refer to this *trans*-acting in vivo mutagenesis scheme as a Mutagenic Chain Reaction (MCR). In principle, MCR should be very efficient in both somatic and germline precursor cells given the high frequency and specificity of mutagenesis (Port et al., 2014) and efficacy of homology-based integration (Gratz et al.,2014) mediated by separate genome-encoded Cas9 and gRNA genes reported in previous studies.

We decided to test the MCR mutagenesis scheme in *D. mel.* by targeting the *yellow* (section 3.1) and *ebony* (section 3.2) genes for which efficient CRISPR mutagenesis tools had been previously characterized (Bassett et al., 2013, Port et al., 2014).

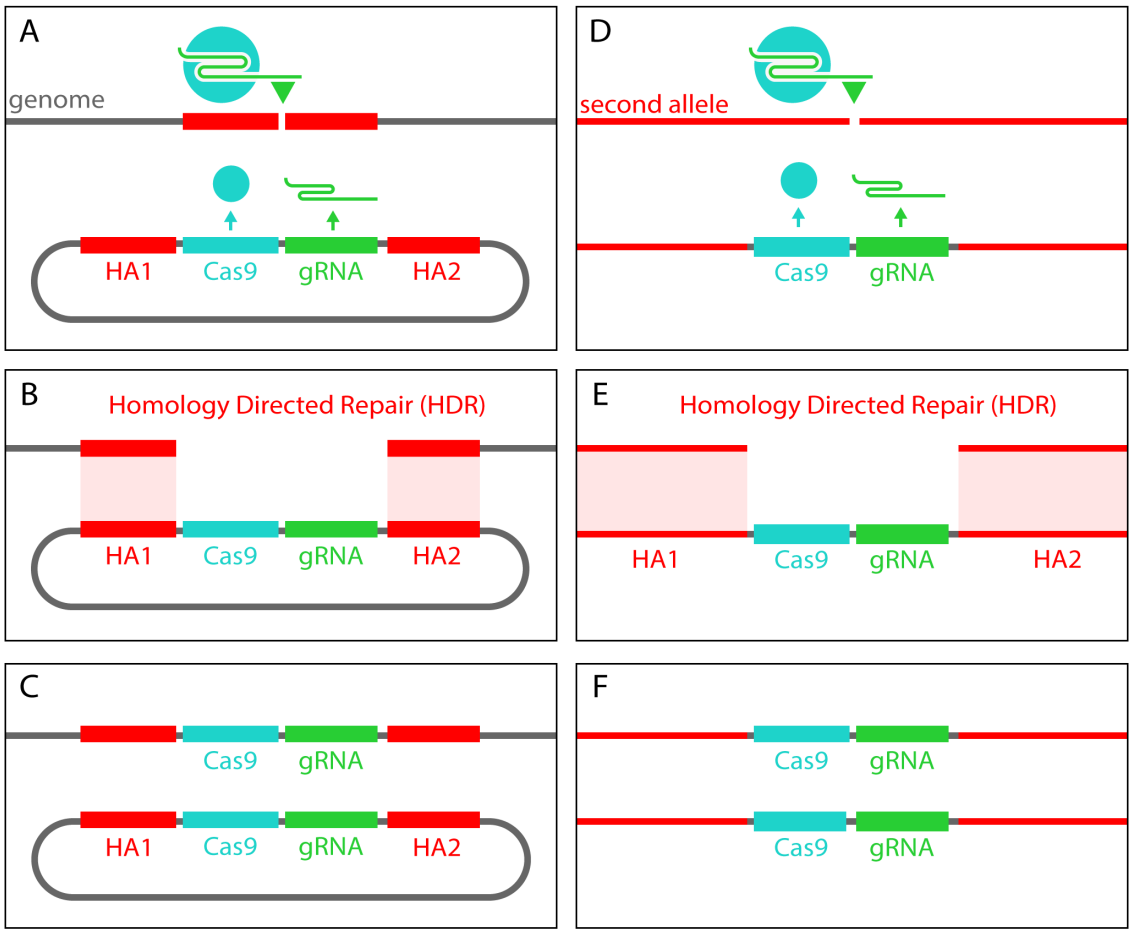

Figure 9 - Scheme outlining the mutagenic chain reaction (MCR).

(A to C) A plasmid consisting of a core cassette carrying a Cas9 transgene, a gRNA targeting a genomic sequence of interest, and flanking homology arms corresponding to genomic sequences abutting the target cleavage site (A) inserts the core Cas9-gRNA cassette into the targeted locus via HDR [(B) and (C)]. (D to F) In turn, the inserted cassette expresses both Cas9 and the gRNA, leading to cleavage (D) and HDR-mediated insertion of the cassette into the second allele, thereby rendering the mutation homozygous [(E) and (F)]. HA1 and HA2 denote the two homology arms that directly flank the gRNA-directed cut site.

4.1. Development of the MCR: targeting the *Drosophila melanogaster yellow* gene.

I decided to empirically test the MCR concept in *D. mel.* by targeting the transgene insertion in the X-linked *yellow* (*y*) locus using the previously characterized "y1" sequence as the gRNA target (Bassett et al., 2013), a *vasa-Cas9* transgene as a source of Cas9 protein (Gratz et al., 2014), and two homology arms of ~1 kb flanking either side of the core vasa-Cas9/y-gRNA cassette that precisely abut the expected gRNA cut site (Fig. 10C, Fig 11A).

Wild-type embryos (*y*+/*y*+ ♀ or *y*+/Y chromosome ♂ genotypes) were injected with the y-MCR element and emerging F0 flies were crossed to a *y*+ control stock. Even if a traditional *y*- allele were to be generated, according to standard Mendelian inheritance, all F1 female progeny of such a crosses should display a wild-type *y*+ phenotype. However in 9 cases (2♂ and 7♀ injected animals), we recovered F1 progeny with a whole-body *yellow* (*y*-) phenotype. Furthermore, when these *y*- F1♀ individuals were crossed back to wild-type *y*+ ♂ flies, 95-100% of their F2 progeny (average = 97%) exhibited a whole-body *y*- phenotype (Table 1) in contrast to the expected Mendelian ratio of 50% (i.e. males only). We also found rare y-MCR F2 ♀s mosaics (~ 4%) with *y*+ patches, and in two instances, recovered *y*+ male progeny from a *y*-MCR F1 female. Thus, while y-MCR transmission via HDR is a highly efficient process in both somatic and germline lineages, chromosomal targets occasionally evade conversion.

PCR analysis of the *y* locus in individual *y*- F1 progeny confirmed the precise expected gRNA-driven genomic insertion of the *y*-MCR construct in all flies that gave rise to *y*- female F2 progeny (Fig. 10D). As expected, males carried only this single allele, while females also carried a DNA fragment corresponding to the size of the wild-type *y* locus, which varied in intensity between individuals indicating that females are mosaic for the MCR conversion. In addition, sequence analysis of rare non-converted *y*+ alleles revealed, in one case, a single synonymous nucleotide change at the gRNA-determined cut site which most likely resulted from a NHEJ repair event, and an in-frame indel in the other instance. The high recovery rate of whole-body *y*- female F1 and F2 progeny from single parents containing a *y*-MCR element, the efficient germline transmission of the MCR element between F1 and F2 generations, and the presence of expected MCR-derived fragments in all F1 *y*- progeny transmitting a converting *y*- allele to their F2 progeny, indicated that the conversion process at this locus was highly efficient in the germline and possibly in the somatic lineage.

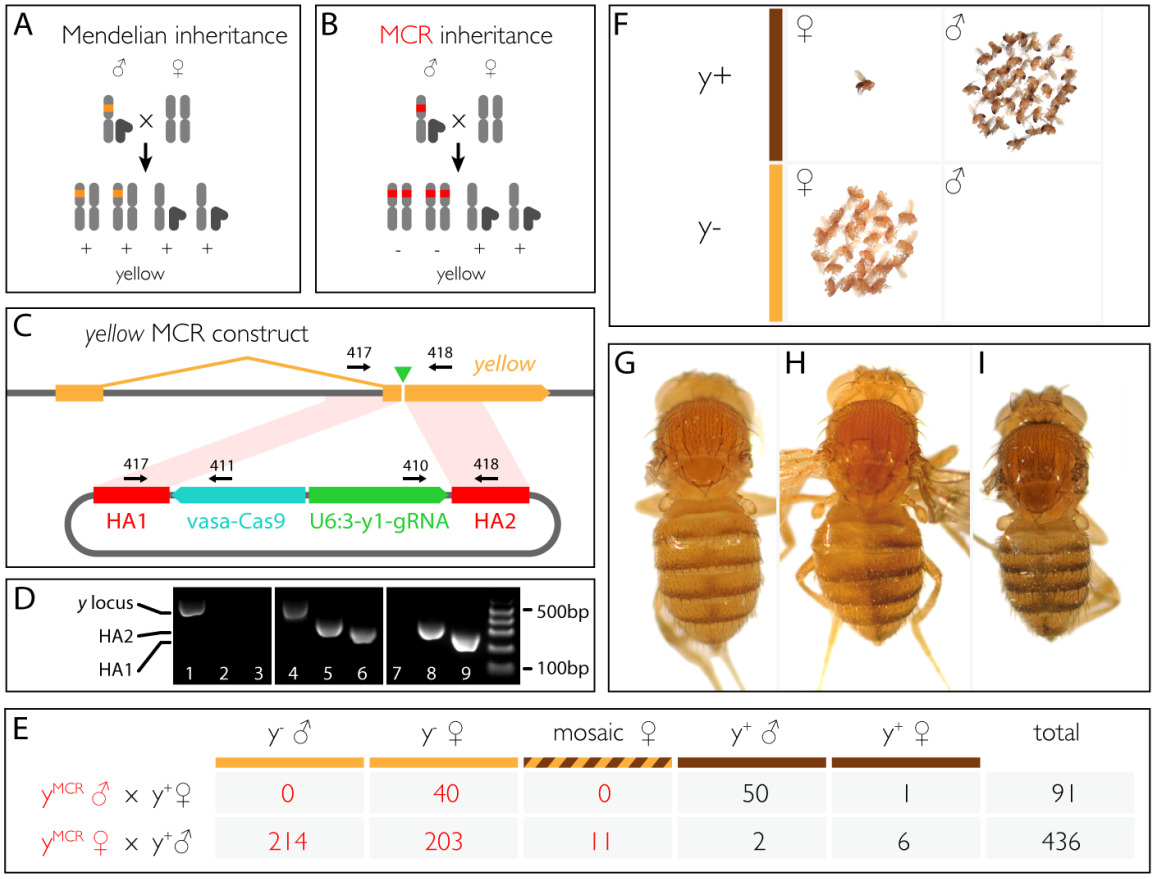

Figure 10 - Experimental demonstration of MCR in *Drosophila melanogaster.*

(A) Mendelian male inheritance of an X-linked trait. (B) Theoretical MCR-based inheritance results in the initially heterozygous allele converting the second allele, thereby generating homozygous female progeny. (C) Diagram of y-MCR construct. Two y locus homology arms flanking the vasa-Cas9 and y-gRNA transgenes are indicated, as are the locations of the PCR primers used for analysis of the genomic insertion site (see supplementary materials). (D) PCR analysis of a y+ MCR-derived F2♂ (lanes 1 to 3; see Gantz and Bier, 2005, fig. S1 for sequence), yMCR F1♀ (lanes 4 to 6), and yMCR F1♂ (lanes 7 to 9) showing junctional bands corresponding to y-MCR insertion into the chromosomal y locus (lanes 2, 3, 5, 6, 8, and 9) and the presence (lanes 1 and 4) or absence (lane 7) of a PCR band derived from the y locus. Although the yMCR F1♂ (carrying a single X chromosome) displays only MCR-derived PCR products (lanes 8 and 9), yMCR F1♀s generate both MCR and non-insertional amplification products. (E) Summary of F2 progeny obtained from crosses described in table 1. (F) Low-magnification view of F2 progeny flies from an yMCR ♂× y+♀ cross. Nearly all female progeny display a y– phenotype. (G) High- magnification view of a full-bodied yMCR F1♀. (H) A rare 50% left-right mosaic female. (I) A y+ control fly.

Table 1 - Propagation of the *y*- phenotype among progeny of y-MCR parents.

Summary of the genetic transmission of the *y*- phenotype through two generations carrying the y-MCR construct. Two F0 parents were selected for this analysis, one male (M3) and one female (F5) which when mated to *y*+ flies gave rise to *y*- female F1 progeny, and hence were scored as carrying the y-MCR construct. For M3 (who had no male *y*- F1 progeny as expected), 6 of his y^{MCR} F1 female progeny (f1-6) were then crossed to y+ males to generate an F2 generation. Female F5 gave rise to 14 *y*- females and 18 *y*- males, of which we tested two males (m1, m2) for potential inheritance and propagation of the y-MCR construct by crossing them to *y*+ females and scoring the F2 generation for the *y*- phenotype. Female F2 *y*- progeny were each examined closely for visible mosaicism. The percent of *y*-MCR progeny was calculated by dividing the number of *y*- F2 progeny (including mosaics) by the total number of female progeny. The percent of successful germline MCR conversion via HDR (homology directed repair) was estimated in female progeny from F1 crosses by assuming that half of them would be expected to inherit the MCR element by Mendelian segregation and would thus give rise of at least 50% *y*- progeny (perhaps with some mosaicism) while the other half would bear a *y*+ chromosome unless it had been converted in the germline of the F1 parent via HDR [percent conversion = 2(X – 0.5N)/N, where N = total number of flies and X = number of mutant progeny]. This is likely to be an underestimate of the actual germline conversion rate since some females inheriting the F1 y^{MCR} allele might not give rise to y- progeny. Indeed, as indicated in the male crosses, where all female progeny would be expected to inherit the MCR construct by simple Mendelian transmission, we found one *y*+ female (from m2), suggesting that the *y*+ allele inherited from the female F1 parent somehow evaded HDR conversion. We also observed two instances in which male progeny inherited *y*+ alleles from y-MCR carrying females (asterisks). These alleles may either have escaped MCR conversion altogether or perhaps were the result of non-homologous end-joining repair that generated in frame deletions that carry out *y* gene function but that are protected from further gRNA directed cleavage. The latter case is strongly suggested by the *y*+ male derived from the female f3, which sequence analysis revealed carries a single nucleotide change at the gRNA cut site within the *y* locus resulting a T->I substitution (Supporting Fig. 2 in Gantz and Bier, 2015). This gRNA-resistant allele is unlikely to be a rare sequence polymorphism since if were, it should have resulted in 50% of the F2 offspring being *y*+. We also analyzed the sequence of one of the two *y*+ females derived from the same MCR parent (F3) and identified a combined in-frame deletion (7 nucleotides) and insertion (4 nucleotides), the net effect of which is the substitution of three amino acids (TVG) with two residues (IY) (Supporting Fig. 2 in Gantz and Bier, 2015). We note that the percent of *y*- males among total male progeny (2%) is less than that for *y*+ females (6%) raising the possibility that *y*+ females consist of both *y*- (guide-cleaved mutant)/+ and *y*+ (guide-resistant mutant)/+ genotypes. PCR data for entries indicated in **bold red text** are shown in Fig. 10D. F2 progeny from male m2 (bold blue text) are shown in Fig. 10F. Green text indicates averages of % y-MCR and % HDR germline conversion for all lines tested in this table.

F0 progenitor	Sex of F1	Total F2 offspring	F2 y- ♂	F2 y+ ♂	F2 y- ♀	F2 y+ ♀	Mosaic ♀	% ♀ Mosaic	Tot. F2 ♀	Tot. HDR ♀	% y- MCR ♀	HDR germline conversion rate (%)
M3	f1	55	30	0	22	0	3	12	25	25	100	100
M3	f2	73	39	0	33	0	1	3	34	34	100	100
M3	f3	74	35	1*‡	35	2	1	3	38	36	94.7	89
M3	f4	69	31	1*	34	2	1	3	37	35	94.6	89
M3	f5	66	28	0	33	1	4	11	38	37	97.4	95
M3	f6	99	51	0	46	1	1	2	48	47	97.9	96
F5	m1	30	-	15	15	0	0	0	15	15	100	100
F5	m2	61	-	35	25	1	0	0	26	25	96.2	92
Total/ Ave.	-	527	214	52	243	7	11	4.2	261	254	97.3	94.5

4.2. MCR targeting the *Drosophila melanogaster ebony* gene.

While I was in the process of testing y-MCR, I generated in parallel a similar construct that would analogously target its insertion into the *D. mel* autosomal *ebony* (*e*) locus. Since I could not foresee *a priori* whether the *vasa*-Cas9 transgene would be optimal for this kind of approach, I decided to build this second MCR using the published *actin5C*-Cas9 transgene (Port et al. 2014), which should principle be expressed in all cells. Furthermore, because of the possibility that endogenous transcription of a locus could interfere with Cas9 transcribed in the opposite orientation (e.g., as in the case *y*-MCR), I decided in this case to create the *e*-MCR construct carrying the Cas9 transgene inserted in the same direction as the *ebony* transcription. Additionally, since the *actin5C* regulatory region driving Cas9 is fairly large (~4 kbp), it should in principle act as a buffer against any transcription initiating at the *ebony* promoter.

As in the case of the y-MCR, F0 flies injected with the e-MCR construct were crossed to wild-type *e*+/*e*+ animals and, also in this case, I was able to identify F1 progeny displaying the *ebony* phenotype (not expected in Mendelian inheritance). These flies where then pair-crossed to a double balancer TM3/TM6 line, in order to establish homozygous stocks, in subsequent crosses. F0 injected × e+/e+ crosses where performed in pools and I successfully established three independent homozygous *e*-MCR stocks. I then confirmed precise insertion at that *e*1-gRNA target cut site, by sequencing of PCR products spanning each edge of the insertion.

While I have not yet extensively evaluated the *e*-MCR conversion rate (as it was done for the *y*-MCR), I indeed observed allelic conversion. Furthermore the offspring of *e*-MCR /*e*-MCR(?) × e+/e+ crosses, displayed different phenotypical classes: dark ebony flies (*e*-) intermediate ebony flies (*e*+/-) and wild-type looking flies (*e*+). While in some crosses close to 100% of the progeny was dark ebony (*e*-), in other crosses the other two phenotypical classes (intermediate ebony or wild-type) where highly represented. In the case of the *e-MCR* evaluation of allelic conversion efficiency will be complicated two main factors: 1) *ebony* is partially dominant (i.e.: heterozygous animals display a mild phenotype), 2) the *ebony* phenotype is not cell-autonomous, thus scoring mosaic animals could be difficult.

In conclusion, I have successfully generated a second fully functional MCR targeting the autosomal *ebony* locus. While in this case I used a different Cas9 transgene, targeted a different locus and used an overall larger MCR insertion fragment (~10 kb vs 7.8 kb of the y-MCR), I created a second construct fully capable of allele conversion by a mutagenic chain reaction scheme. Additional analysis will be needed, however, to quantitate its germline conversion frequency.

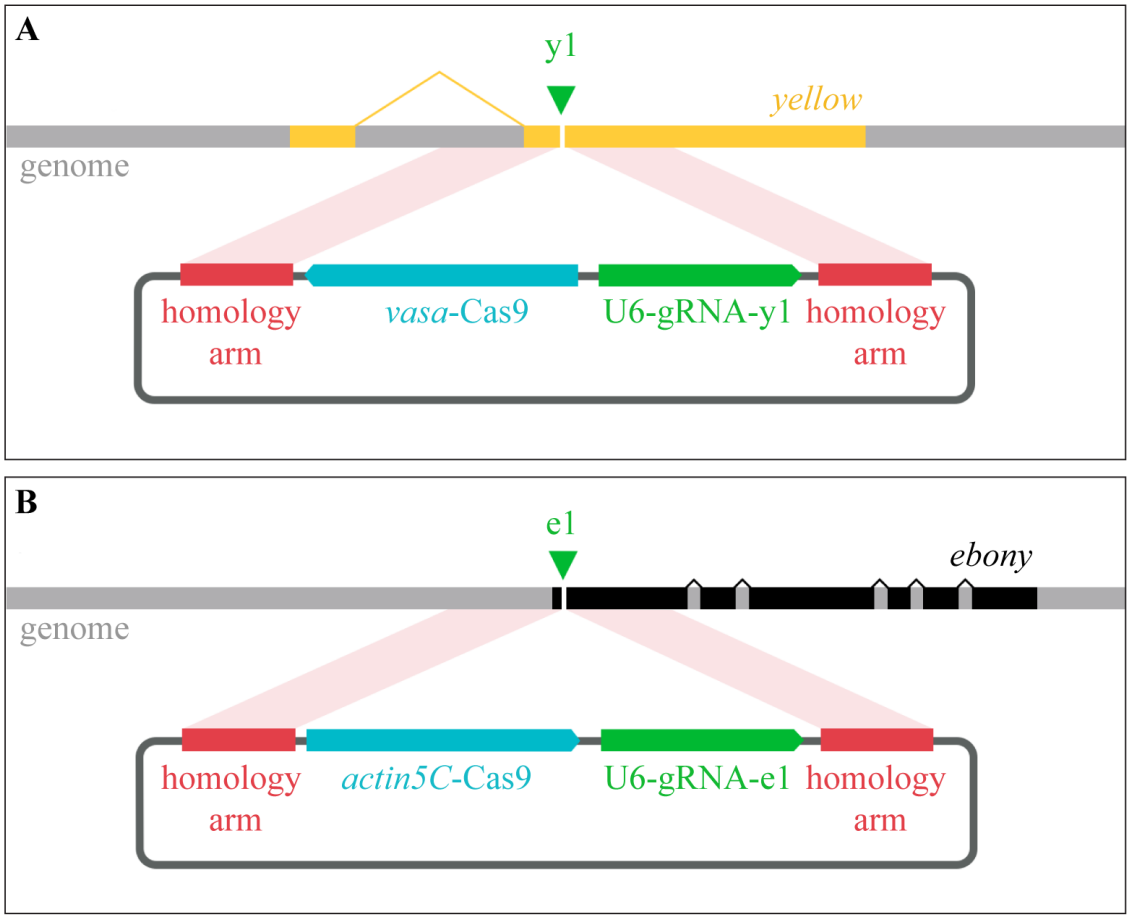

Figure 11 - MCR construct targeting the *ebony* gene.

(A) outlines the insertion of the *y*-MCR construct in the *yellow* locus (see chapter 4.1) while (B) the *e*-MCR in the *ebony* locus (chapter 4.2). In (B), differently from (A), I used a Cas9 transgene which uses a constitutively active promoter (actin5C).

4.3. ERACR: Element for the Reversal of the Autocatalytic Chain Reaction.

The MCR scheme represented described in chapter 4.1 is one of the strongest forms of gene drive ever reported. Gene drives are genetic components that have the intrinsic capability of being inherited at a higher rate than Mendelian segregation would predict. Therefore gene drives have the potential of exponential spread in populations leading to complete take-over under idealized conditions.

A concept similar to the MCR was previously proposed (Esvelt et al., 2014), and as these authors suggest, it is advisable to create a construct (called in their case "immunizing drive") capable of stopping the autocatalytic activity of a gene drive such as an MCR. The construct proposed and later tested by the same group using a Cas9-based gene drive in yeast (DiCarlo et al., 2015), efficiently stops the propagation of a gene drive but, after "immunization", the Cas9 gene is left behind along with its potential of increasing the mutagenesis background due to non-specific Cas9 activity or the enzyme acting on off target sites.

At the time I devised the mutagenic chain reaction, I also planned a strategy that could be used to reverse the phenotype generated by an MCR and additionallysuch reversal construct should have the property of exponentially spreading throughout an MCR population (in the same way that an MCR can spread through a wild-type population). As I mentioned above the option proposed by Esvelt et al. leaves and integrated Cas9 gene after "immunization". I decided to improve upon this concept and

devised a more advantageous strategy to reverse the autocatalytic spread of an MCR without leaving any source of Cas9 behind which is called the Element Reversing the Autocatalytic Chain Reaction or ERACR. While the "immunizing drives" rely on a Cas9 source in cis (carried on the "immunizing drive"), the key idea behind the ERACR construct is that it relies on a Cas9 source provided in trans (i.e. from the MCR) and therefore is freed from carrying its own source of this potentially persistently mutagenic enzyme (Fig. 12).

An ERACR construct is built with the following essential components: 1) a central domain containing two gRNAs that would direct Cas9 cleavage to two genomic targets, one on each side of the site previously used by the MCR's gRNA for targeted insertion, and 2) homology arms on each side of the central domain that perfectly about the genomic ends left after excision of the MCR construct. This strategy takes advantage of the Cas9 produced by the MCR to drive the replacement of the MCR, with a copy of the ERACR. The Cas9 protein is needed only before the MCR construct excision, once this event has taken place the endogenous cell machinery takes over, repairing the break using as a template the ERACR-containing chromosome (Fig. 12D-F). In addition to these minimal elements an ERACR construct could also carry a recoded version of the gene disrupted by the MCR it is designed to delete (thereby restoring gene activity) or other desired marker or effector genes.

The first version of an ERACR that I built (y-ERACR) was designed to excise and replace the y-MCR construct described in chapter 4.1 (Fig. 10C and Fig. 11A).

Additionally to the gRNAs (1) and the flanking homology arms (2) described above, I added two extra components: 3) a partially recoded *yellow* gene sequence, that would be inserted in frame with the genomic transcript, therefore restoring the *yellow* transcript and 4) a 3xP3-DsRed transgene (red fluorescent eye marker) that could be used to follow the y-ERACR.

The y-ERACR construct was injected along with a plasmid source of Cas9 (pHsp70-Cas9, Addgene #46294) in a w- *D. mel.* stock. The injected animals where crossed to each other in pools of 2♂ and 2♀. The F1 progeny of such crosses was screened for the red fluorescent eye marker and from positive vials single y-ERACR/Y ♂ where crossed to FM7/FM7 ☿ to establish stocks.

In parallel of the stock establishment, I proceeded to test the conversion efficiency of the y-ERACR: I single-crossed 3 emerging F1 ☿ (ME1, ME2, ME3, see Table 2) that displayed eye fluorescence (y-ERACR/+) to y-MCR/Y ♂. From each of the crosses, I collected a total of 15 F2 virgin female flies that carried the fluorescence eye marker (y-MCR/y-ERACR) and crossed these trans-heterozygotes to a w-y+ stock (see Table 2). After eclosion I analyzed the phenotype of the emerging F3 progeny in order to assess the conversion efficiency of the *y*-ERACR. In Table 2 displays the phenotypical classes and their frequencies for the F3 progeny of such crosses. By focusing on the *y*- female progeny it is possible to estimate the frequency of MCR deletion as the fraction of females with a yellow mutant phenotype can only be explained by presence of the MCR). This analysis revealed that only ~5% of the female progeny retained the MCR, or

conversely that the MCR was neutralized ~95% of the time (Table 2, † and ‡). From the analysis of the male progeny it is possible to estimate the efficiency of ERACR conversion of the MCR: the M Y R column are reported males that display a phenotype expected if they had a *y*-ERACR/Y genotype. The frequency of this phenotypical class is expected to be 50% for Mendelian inheritance alone. However, I observed 81.4% males (Table 2, §) displayed the expected ERACR marker phenotype which indicates that the conversion rate of the MCR allele into an ERACR was ~60%. As it is evident in the last column (male to female ratio) I observed, in many cases, a value lower than the expected ratio of 1.00. I believe possible that one or more of the stocks that I used for these crosses may have carried an X-linked lethal mutation therefore resulting in altered ratios (male lethality) and associated skewed evaluation of conversion rate. I also observed that the F y R class exists (Table2, †) and, while rare, it can be explained in principle by chimeric fusion of the *y*-MCR (responsible for the *y*- phenotype) and the *y*-ERACR (responsible for the red eye marker). *A posteriori* I realized that some regions of the y-ERACR had sequence homology to part of the MCR or neighboring genomic region. I believe such homology regions could be responsible for the unexpected classes observed in Table 2, since they could have driven homology directed repair after only one the two y-ERACR's gRNAs generated a cut. This hypothesis should be confirmed by molecular analysis of the *yellow* locus in all such unexpected classes (M Y r, M y R, M y r). Furthermore, based on that assumption, I have designed an y-ERACR2.0 construct which eliminates virtually all cross-homology between the ERACR and MCR/yellow locus sequences: the *yellow*

sequence has been completely recoded, the yellow 3'UTR has been replaced by that one of the evolutionary distant *Drosophila grimshawi* (*D. gri*) as I did for the U6 regulatory regions driving the expression of the gRNAs (now from *D. gri*). Such construct should, in theory, dramatically increase the correct conversion of the y-MCR to the y-ERACR2.0.

In conclusion, while the ERACR strategy needs further optimization in order to reach higher conversion rate, I have successfully generated a construct which is capable of targeting the y-MCR for removal 95% of the time and successfully converts it to the ERACR allele with ~60% efficiency.

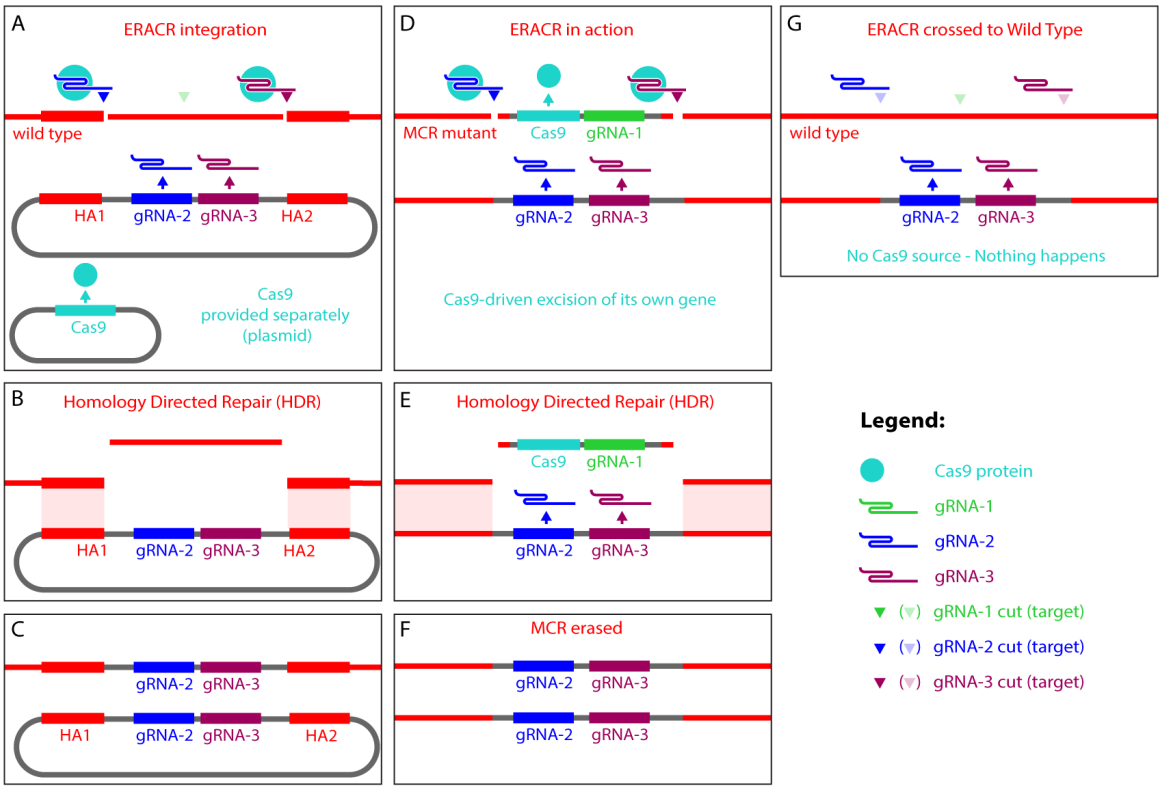

Figure 12 - ERACR strategy.

(A-C) the insertion of a ERACR construct makes use of the Cas9 provided by a second plasmid (A) to drive insertion of the ERACR in the genome by homology directed repair (HDR) in between the cut sites generated by the gRNA-2 and gRNA-3 carried on the ERACR construct (B & C). When an ERACR baring animal mates with an MCR one the generated heterozygous cells (D) also undergo ERACR conversion driven by the Cas9 carried on the MCR to drive replacement of the MCR cassette itself by an HDR mechanism (E-F). The advantage of the ERACR approach consists in its dependence form an external source of Cas9; when an ERACR is combined with a wild type allele (G) it is incapable of conversion due to lack of Cas9 protein.

Table 2 - Evaluation of the conversion efficiency of the y-ERACR construct.

F2 Cross ↓	Males					Females					M / F
Phenotype →	M Y R	M Y r	M y R	M y r	Total	F Y R	F Y r	F y R	F y r	Total	Ratio
ME1-1	38				38	38	33			71	0.54
ME1-2	39	1			40	34	43			77	0.52
ME1-3	29	1	1		31	44	18			62	0.50
ME1-4	41	6			47	45	7		4	56	0.84
ME1-5	37		1	12	50	30	19		10	59	0.85
ME1-6	32			15	47	32	16	1	12	61	0.77
ME1-7	14				14	20	3		1	24	0.58
ME1-8	15	5		1	21	15	3	1	3	22	0.95
ME1-9	42	16			58	52	11		1	64	0.91
ME2-1	27	11		2	40	28	22		2	52	0.77
ME2-2	42	6		1	49	33	30		2	65	0.75
ME3-1	24	7		2	33	20	19		2	41	0.80
ME3-2	44	6	1		51	43	23		1	67	0.76
ME3-3	32		2	3	37	45	24		1	70	0.53
ME3-4	38	10		3	51	33	18		1	52	0.98
Total	494	69	5	39	607	512	289	2	40	843	
Percentage	§ 81.4	11.4	0.8	6.4	100	60.7	34.3	† 0.24	‡ 4.7	100	

Crosses: MEx-y; x refers to the F1 female; y refers to the progenitor F2 female.
Phenotypes: M=male, F=female, Y= *yellow*+, y=*yellow*-, R=red eye marker, r=no eye marker

The table summarizes the phenotypical classes obtained from crossing *y*-ERACR/*y*-MCR♀ to y+/ Y♂ flies (15 independent crosses numbered MEx-x). As it is possible to observe in the table, while in several cases I obtained unexpected phenotypical classes, the y-MCR construct was neutralized at about 95% rate, inferred by the amount of yellow females collected (‡).

4.4. MCR in *Anopheles stephensi*: an approach for malaria eradication.

An autosomal allele is defined as being under genetic drive if more than 50% of the progeny inherit the allele from an individual carrying a single copy of that allele. A wide variety of genetic elements or symbiotic/parasitic organisms have been identified that generate drive and are often referred to as selfish genes since the can spread through a population and become fixed as a result of their steady increase in frequency (James 2005, Sinkins et al. 2006, Marshall 2009, Alphey et al. 2013, Burt 2014). Well-studied examples of such selfish elements or organisms include: transposons (Skipper et al. 2013), Medea elements (Chen et al. 2007, Ward et al, 2011, Akbari et al. 2014), homing endonuclease genes (HEGs) (Deredec et al. 2011, Windbichler et al. 2011, Alphey et al. 2014), maternal effect underdominant elements (Akbari et al. 2013), and the bacterial endosymbiont/parasite Wolbachia (Sinkins et al. 2004, Huang et al. 2007). CRISPR-based self-propagating elements such as MCRs or similar constructs are newcomers to this established selfish DNA clan.

Among the previously studied gene drive systems, HEGs share great parallels with MCR elements (Deredec et al. 2011, Windbichler et al. 2011, Alphey et al. 2014). HEGs encode restriction endonucleases inserted at a site in the genome of host organisms that can be cut by the HEG nuclease. The combination of Cas9 and a gRNA in an MCR effectively performs exactly the same function. The big advantage that MCR is that it can be targeted for insertion into virtually any locus in the genome, while in the case of

an HEG one must either make do with its single defined target sequence or go through significant effort to select for enzyme variants that target a new site of interest. In addition, the ability to separately control delivery of Cas9 and gRNAs ad two separable genetic elements offers great versatility to develop additional functionalities.

The ability of gene-drive constructs, such as the MCR, to spread exponentially through a population is a feature that can be exploited to disperse disease-resistance cassettes that would lower transmission rates of parasites such as *Plasmodium spp.* or viruses (eg.: Dengue virus). Gene cassettes that would lower transmission of the malaria-causing mosquito parasite *Plasmodium* have been previously generated and shown to decrease transmission by different strategies (Ito et al., 2002, Moreira et al., 2002, Isaacs et al., 2011). The single chain antibody (scFv) approach, developed by Isaacs et al., was the product of a fruitful collaboration between the Vinetz lab (here at UCSD) and the James lab at UC Irvine (Isaacs et al., 2011, Isaacs et al., 2012). In the discussion section of their 2012 manuscript these authors state that " *If coupled with a mechanism for gene spread, scFv-expressing, malaria-resistance transgenes could become a self-sustaining disease control tool.*" (Isaacs et al., 2012). My advisor Ethan and I realized that the MCR is exactly the mechanism that this authors are calling for. With this idea in mind we took immediate steps to establish a collaboration with the James and Vinetz Labs in order to test the MCR technology mosquitoes and potentially generated tool to distribute malaria-resistance cassettes in wild *Anopheline* populations.

In consultation with our collaborators we decided to build tools (Cas9 and sgRNA genes) that would be optimal for expression in *Anopheles stephensi* (*A. ste.*) as well as a full MCR that would target its insertion in the *kynurenine 3-monooxygenase* (*kmo* or *kh*) gene of *A. ste.* homologous of the *D. mel. cinnabar* (*cn*) gene (Bhalla, 1968, Han et al. 2003). I started this process by cloning the regulatory regions of the *A. ste. vasa* (4kb) and *actin5C* (9kb) genes and fusing them to an *A. ste.* codon optimized Cas9 gene, provided by the James lab, in order to generate transgenes that would have optimal expression in such species. Similarly, I identified the genomic locations of two different U6-shRNA genes (U6A and U6B) and then cloned a ~1 kb fragment from such loci and replaced the U6-shRNA with a sgRNA sequence preceded by two BbsI restriction enzyme cut sites for rapid cloning of target sequences (Gratz et al., 2013). Most of the obtained transgenes, represented in figure 13, were also cloned into a pBacDsRed-attB backbone (Isaacs et al., 2012) which permits the generation of stable transformants, using either traditional piggyBac or φC31 site specific integration transgenesis protocols.

As mentioned above we decided to first target the coding sequence of the *kmo* (or *kh*) gene which is homologous to the *D. mel*. eye pigment locus *cinnabar*; this choice was guided by two factors: 1) mutation in *kmo* results in the easily scored white eye phenotype in *Aedes aegypti* (Bhalla, 1968, Han et al. 2003), 2) *kmo* has been successfully targeted for disruption using at TALEN-based approach (Aryan et al., 2013). Using the transgenes that I previously generated, I built the *kh1*-MCR construct (*kh* is a synonym of *kmo*) that should integrate in the *kh2* target site represented in figure 13.

Although I had to overcome a variety of new technical cloning challenges to generate the >20 kb *kh1*-MCR construct, I recently verified its full sequence and sent it to Nijole Jasinskiene, in the James Lab to be injected into embryos for transgenesis. In the event that it is difficult to recover such transformant (e.g., due to its large size), Nijole is also injecting the pBacDsRed-*AsVasa*-AsCas9 (pVG160) construct to generate stable transformant with a stable genomic source of Cas9 via piggyBac transposition. Additionally two gRNA plasmid sources, targeting either the *kh1* or the *kh2* were co-injected (pVG162 and pVG163 respectively) in order to acquire information that may help us optimize the efficiency of second generation constructs.

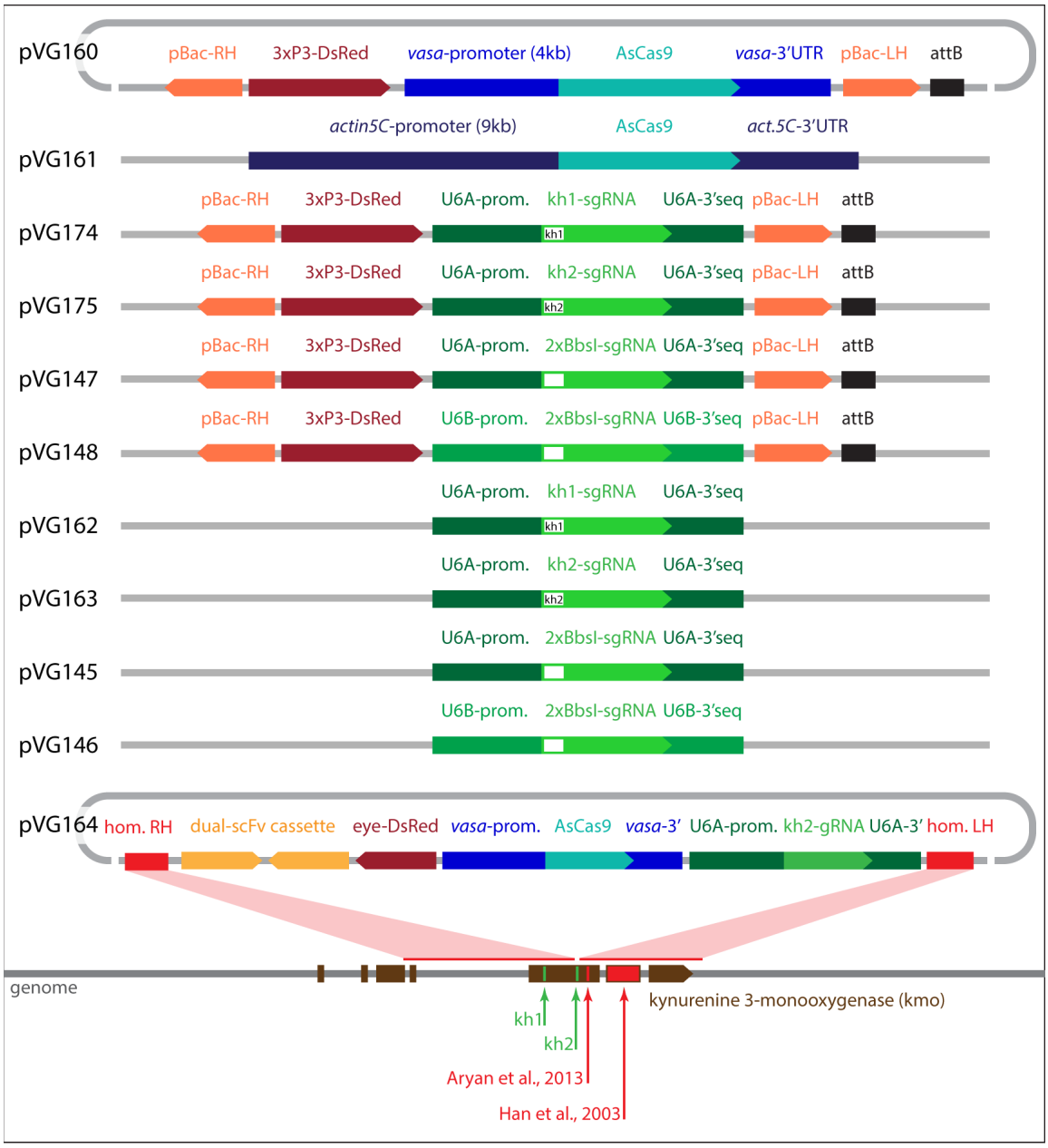

Figure 13 - *Anopheles stephensi* Cas9, gRNA and MCR constructs.

Constructs generated in order to optimize CRISPR/Cas9 genome editing in *Anopheles stephensi* in order to test the MCR technology for future use to spread malaria-resistance cassettes in wild populations.

4.5. Safety concerns and containment strategies for MCR genetic elements

Recently there have been fruitful exchanges in the literature discussing strategies for regulating transgenic constructs designed to achieve genetic drive in populations including transposable and selfish genetic elements or organisms (Marshall 2009), homing endonuclease genes (Derdec et al. 2011, Windbichler et al. 2011, Alphey et al. 2014) or a similar CRISPR/Cas9 based strategy (Oye et al. 2014, Esvelt et al. 2014) to that proposed and demonstrated here. We ardently join others who have considered risks associated with drive elements in pointing out imperative Biosafety and ethical concerns associated with genetic drive. Such issues are all the more critical to address now that we have explicit evidence for the remarkable efficiency of potential MCR transmission in metazoan species and the inherent ability of such drive elements to spread rapidly and pervasively through populations, potentially converting an entire species to a new genetic condition. Contagious dispersion of such mutations into wild populations could arise unintentionally without adequate safety measures to assure that a transgenic organism carrying an MCR construct has no opportunity to be released or mate with free individuals. There are also more sinister scenarios in which such constructs could be used intentionally for ill purposes. Therefore among with others we started a dialogue on this question (Akbari et al. 2015) and participated in the discussion recently initiated by the National Academy of Sciences on gene drives. On the practical side we have

developed a protocol (BUA R461) for our use of MCR elements in *Drosophila* that has been approved by the UCSD Institutional Biosafety Committee and reported here:

BUA R461 MCR AMENDMENT: Approved: 01/16/2015

Biosafety measures: To prevent any unintentional release of MCR flies into the environment we have taken stringent precautions including the following measures:

1) All fly vials are tightly plugged with cotton and are in turn inserted inside a 50 ml Corning tube with a Nitex mesh covered hole (0.5 cm diameter) in the cap and placed in racks within a sealed Tupperware box with four mesh covered holes (1 cm diameter) for air exchange. We refer to these vials stocks as triple contained.

2) Boxes with triple contained MCR flies are kept in locked facilities at all times.

3) MCR flies are either killed by freezing directly within triple contained vials prior to inspection or use for molecular analysis (e.g., DNA extraction and PCR) or are transported in triple contained vials to a BSL2+ Insectary with triple doors for manipulations of live flies. Live flies are anesthetized and then a few easily countable number of flies (10-20) are placed on a CO2 pad, sorted, counted, and then either placed in a new vial or killed immediately by emersion in oil. The number of flies within the receptive vial are then counted to assure they match the number on the CO2 pad.

4) Detailed records are maintained for each cross of MCR flies indicating how many F0, F1, or F2 flies were examined in a living or dead condition.

5) Microfuge tubes containing MCR DNA constructs and bacterial stocks carrying such constructs are well labeled and kept in separate boxes of refrigerator or freezer compartments. Following growth of any MCR containing bacteria, plates are immediately collected and autoclaved.

6) Only a single highly expert investigator (V. Gantz) handles MCR reagents, bacterial stocks and flies to avoid any possible confusion arising from multiple investigators.

BUA R461 MCR AMENDMENT END

4.6. The mutagenic chain reaction: A method for converting heterozygous to homozygous mutations.

This chapter, in full, is a reprint of the material as it appears in Science 2015, Gantz, Valentino M and Bier, Ethan. The dissertation author was the primary investigator and author of this paper.

Figure 1: please refer to the dissertation's figure 9.

Figure 2: please refer to the dissertation's figure 10.

ABSTRACT

An organism with a single recessive loss-of-function allele will typically have a wild-type phenotype, whereas individuals homozygous for two copies of the allele will display a mutant phenotype. We have developed a method called the mutagenic chain reaction (MCR), which is based on the CRISPR/Cas9 genome-editing system for generating autocatalytic mutations, to produce homozygous loss-of-function mutations. In Drosophila, we found that MCR mutations efficiently spread from their chromosome of origin to the homologous chromosome, thereby converting heterozygous mutations to homozygosity in the vast majority of somatic and germline cells. MCR technology should have broad applications in diverse organisms.

It is often desirable to generate recessive loss-of-function mutations in emergent model organisms; however, identifying such mutations in the heterozygous condition is challenging. Taking advantage of the CRISPR/Cas9 genome-editing method (1, 2), we have developed a strategy to convert a Drosophila heterozygous recessive mutation into a homozygous condition manifesting a mutant phenotype. We reasoned that autocatalytic insertional mutants could be generated with a construct having three components: (i) A Cas9 gene (expressed in both somatic and germline cells), (ii) a guide RNA (gRNA) targeted to a genomic sequence of interest, and (iii) homology arms flanking the Cas9-gRNA cassettes that match the two genomic sequences immediately adjacent to either side of the target cut site (Fig. 1A). In such a tripartite construct, Cas9 should cleave the genomic target at the site determined by the gRNA (Fig. 1A) and then insert the Cas9-gRNA cassette into that locus via homology-directed repair (HDR) (Fig. 1, B and C). Cas9 and the gRNA produced from the insertion allele should then cleave the opposing allele (Fig. 1D), followed by HDR-driven propagation of the Cas9-gRNA cassette to the companion chromosome (Fig. 1, E and F). We refer to this trans-acting mutagenesis scheme as a mutagenic chain reaction (MCR).

We expected that autocatalytic allelic conversion by MCR should be very efficient in both somatic and germline precursor cells, given the high frequency and specificity of mutagenesis (3) and efficacy of homology-based integration (4) mediated by separate genome-encoded Cas9 and gRNA genes observed in previous studies. We tested this prediction in D. melanogaster with the use of a characterized efficient target sequence

(y1) (5) in the X-linked yellow (y) locus as the gRNA target and a vasa-Cas9 transgene as a source of Cas9 (Fig. 2C) because it is expressed in both germline and somatic cells (4). As the defining element of our MCR scheme, we also included two homology arms, ~1 kb each, flanking the central elements (Fig. 2C) that precisely abut the gRNA-directed cut site. Wild-type (y+) embryos were injected with the y-MCR element (see supplementary materials), and emerging F0 flies were crossed to a y+ stock. According to Mendelian inheritance, all F1 female progeny of such a cross should have a y+ phenotype (i.e., F1 females inherit a y+ allele from their wild-type parent).

From two independent F0 male (♂) × y+ female (♀) crosses and 7 F0♀ × y+♂ crosses, we recovered y– F1♀ progeny, which should not happen according to Mendelian inheritance of a recessive allele. Six such yMCR F1♀ were crossed individually to y+♂, resulting in 95 to 100% (average = 97%) of their F2 progeny exhibiting a full-bodied y– phenotype (Fig. 2, E and G, and table S1), in contrast to the expected rate of 50% (i.e., only in males). We similarly tested MCR transmission via the germline in two y– F1♂ recovered from an F0♀ cross that also yielded y– female siblings. These y– F1♂ were considered candidates for carrying the y-MCR construct and were crossed to y+ females. All but one of their F2 female progeny had a full-bodied y– phenotype (Fig. 2, E and F). Occasionally among yMCR F2♀ we also recovered mosaics (~4%) with a few small y+ patches as well as a lone example of a 50% chimeric female (Fig. 2H), and in two instances, we recovered y+ male progeny from a yMCR F1♀ mother (Fig. 2E and table S1). These infrequent examples of imperfect y-MCR transmission indicate that although

HDR is highly efficient at this locus in both somatic and germline lineages, the target occasionally evades conversion.

Polymerase chain reaction (PCR) analysis of the y locus in individual y– F1 progeny confirmed the precise gRNA- and HDR-directed genomic insertion of the y-MCR construct in all flies giving rise to y– female F2 progeny (Fig. 2D). Males carried only this single allele, as expected, whereas females in addition possessed a band corresponding to the size of the wild-type y locus (Fig. 2D, lane 4), which varied in intensity between individuals, indicating that females were mosaic for MCR conversion. The left and right y-MCR PCR junction fragments were sequenced from y– F1 progeny from five independent F0 parents. All had the precise expected HDR-driven insertion of the y-MCR element into the chromosomal y locus. In addition, sequence analysis of a rare nonconverted y+ allele recovered in a male offspring from a yMCR F1♀ (Fig. 2E) revealed a single-nucleotide change at the gRNA cut site (resulting in a T→I substitution), which most likely resulted from nonhomologous end-joining repair, as well as an in-frame insertion-deletion (indel) in a y+♀ sibling of this male (fig. S1 and table S1). The high recovery rate of full-bodied y– F1 and F2 female progeny from single parents containing a yMCR allele detectable by PCR indicates that the conversion process is remarkably efficient in both somatic and germline lineages. Phenotypic evidence of mosaicism in a small percentage of MCR-carrying females and the presence of y locus–derived PCR products of wild-type size in all tested y– F1 females suggest that females may all be mosaic to varying degrees. In summary, both genetic and

molecular data reveal that the y-MCR element efficiently drives allelic conversion in somatic and germline lineages.

MCR technology should be applicable to different model systems and a broad array of situations, such as enabling mutant F1 screens in pioneer organisms, accelerating genetic manipulations and genome engineering, providing a potent gene drive system for delivery of transgenes in disease vector or pest populations, and potentially serving as a disease-specific delivery system for gene therapy strategies. We provide an example in this study of an MCR element causing a viable insertional mutation within the coding region of a gene. It should also be possible, however, to efficiently generate viable deletions of coding or noncoding DNA by including two gRNAs in the MCR construct targeting separated sequences and appropriate flanking homology arms. Using the simple core elements tested in this study, MCR is applicable to generating homozygous viable mutations, creating regulatory mutations of essential genes, or targeting other nonessential sequences. The method may also be adaptable to targeting essential genes if an in-frame recoded gRNA-resistant copy of the gene providing sufficient activity to support survival is included.

In addition to these positive applications of MCR technology, we are also keenly aware of the substantial risks associated with this highly invasive method. Failure to take stringent precautions could lead to the unintentional release of MCR organisms into the environment. The supplementary material includes a stringent, institutionally approved barrier containment protocol that we developed and are currently adhering to for MCR

experiments. Since this study was submitted for publication, a preprint has been posted on the bioRxiv web server showing that a split Cas9-gRNA gene drive system efficiently biases inheritance in yeast (6). The split system was used to avoid accidental escape of the gene drives. The use of a similar strategy in future MCR organisms would reduce, but not eliminate, risks associated with accidental release. We therefore concur with others (7, 8) that a dialogue on this topic should become an immediate high-priority issue. Perhaps, by analogy to the famous Asilomar meeting of 1975 that assessed the risks of recombinant DNA technology, a similar conference could be convened to consider biosafety measures and institutional policies appropriate for limiting the risk of engaging in MCR research while affording workable opportunities for positive applications of this concept.

Chapter 4 and 4.1 contain part of, and Chapter 4.6 is in full a reprint of the material as it appears in Science "*The Mutagenic Chain Reaction: a Method for Converting Heterozygous to Homozygous Mutations*." Gantz, Valentino M, Bier, Ethan (2015) Science, 348(6233): 442–44. The dissertation author was the primary investigator and author of this paper.

5. Discussion

The the wing GRN displays extensive similarities in several gene expression patterns across the enormous evolutionary distance that spans between flies (my studies here) and beetles (Tomoyasu et al., 2005, Tomoyasu et al., 2009). These findings support the idea that the wing GRN evolved only once when the first winged insects (*pterygota*) appeared on earth around 400 Ma (Averof and Cohen 1997, Engel and Grimaldi 2004). While existing data indicates that genes acting at the bottom of the GRN hierarchy such as *knirps*, *abrupt* and *caupoliacan* might have been established as vein inducing genes as early as 350 Ma (*Neoptera*), further analysis of the wing GRN of the *Ephemeroptera* (mayflies) and *Odonata* (dragonflies) could confirm whether or not their vein-inducng role is likely to have been established more basally in the *pterigota* tree.

The developmental comparison reported here between *Megaselia* and *Drosophila* regulation of *knirps* in the vein L2 might to be more complicated at the molecular level: on the one hand, as shown in chapter 3, different alterations of the *knirps* CRM can lead to defects in different proximal/distal positions on the L2 vein. On the other hand, I have also observed that RNAi down-regulating the *aristaless* gene is capable to generate an L2 phenotype in which the proximal portion of the vein is missing (data not shown here) in line with results published by Campbell and Tomlinson et al. (1998). The phenotype observed in the proximal region of L2 observed in the *aristaless* knock-down experiment is qualitatively separable the phenotype observed in *knirps* alleles which tends to create

gaps in the distal part of the L2 vein. While I am planning to analyze *aristaless* expression in *M. abdita*, I believe that the differential regulation of the *knirps* and *aristaless* genes in the L2 vein primordium might track back to the ancestral condition when L2 was split into the two veins R2 and R3 (e.g., as in mosquitoes). By looking at flies with a more ancestral venation pattern (i.e., having both R2 and R3 veins) it might be possible to understand how the differential effect of these two genes came to evolve in the L2 vein. The developed MCR Technology could be used to analyze in detail the cis-regulatory modules the of *M. abdita* wing GRN, and to establish genetic tools in this organism.

The MCR approach offers a potential solution for insect borne diseases and is a technology within reach. The biggest hurdles other than optimization of MCR tools in mosquito species will be to establish national and international guidelines for the application of MCR-based treatments in affected areas and involvement and education of the public regarding the potential positive impacts and safety issues of this technology.

Another application that I envision for the MCR method is a solution for crop pest management. Massive exploitation of the land for agricultural purposes has resulted in unwanted outcomes such as generating favorable conditions for seasonal infestation due to pest population explosions of unprecedented sizes or spread of many invasive species into non-native environments. MCR technology offers a means for suppressing such populations. As in the case of mosquitoes, MCR constructs could be engineered that should sweep through insect pest populations, conferring sensitivity to transgenic plants

purposely engineered to be noxious to transgenic carriers of an MCR-associated effector cassette. Such an approach would not only reduced effect of pests on agriculture but lower, if not eliminate, the need for pesticides. Furthermore the insect species would not be in any danger since its populations could thrive on a non-agricultural vegetation, therefore never be able to reach population sizes that would trigger infestations. This fact should also lower the evolutionary pressure selecting mutants able to evade the system. Alternatively, MCRs could target specific non-essential loci in invasive species that are selectively involved in causing crop damage.

The highly efficient autocatalytic conversion property of MCR-type constructs, the flexibility of insertion into virtually any genomic location, and the completely new mode of genetic inheritance, opens the doors to novel genetics possibilities. Since MCR construct actively copy themselves onto the other chromosome with high efficiency, should two such constructs be combined in a double trans-heterozygous condition, the resulting animal would be expected to become a double mutant. If two different MCR double-mutants were then crossed to each other, the resulting offspring would be ~100% quadruple-mutant. Mutants carrying 4 different construct are fairly difficult to built even in a model organism as *Drosophila* with extremely developed genetic tools, particularly in situations where the construct locations are in close linkage. MCR elements would accelerate creation of such complex stocks by: 1) halving the number of generations needed to obtain mutant combinations, 2) increasing the rate of recovery of multiple mutants (e.g., ~100% MCR efficiency compared to 1/64 for standard Mendelian

inheritance segregating heterozygous quadruple-mutant alleles) and 3) dramatically reducing limits imposed by recombination (i.e., tight linkage of alleles). Taken together, these advantages allow a new era of active genetic elements to begin.

The enormous advantages of active genetic elements comes with some risk that needs to be carefully considered. The same population dispersal property that has the potential to permanently eradicate malaria from earth, is intrinsically present in all constructs used for any other purpose. This means that any avenue using active genetic elements should use additional precautions to prevent unintended contact of any such animal with the species wild population. As described in chapter 4.5 I have developed a physical containment protocol for *Drosophila melanogaster* and have contributed to a lengthy discussion among prominent researchers in the field to establish a consensus on laboratory safeguards for using active genetic elements. I have also envisioned a new type of split-element system that while working as a molecular containment strategy (Akbari et al. 2015) would allow the combination of the two genetic elements to form have the full MCR capability of spreading into a population, that gRNA-only drives do not possess. A gRNA-only drive would be supplemented with an additional gRNA targeting a second locus at which the Cas9 protein source would be inserted. Such elements are mutually dependent for function and only when combined are capable of generating a mutagenic chain reaction. This configuration should dramatically reduce the need for additional safeguards needed at the reagent generation and optimization stage, although subsequent experiments involving the crossing of such strains to bring the two

elements together should be performed in high containment facilities (Arthropod Containment Level 2-3) prior to release into the wild (protocols for such release will of course require additional levels of deliberation and transparent public discussion). Nonetheless the MCR technology opens brand new avenues for both basic research, agricultural and human applications; while some of its applications are within reach, its benefits have the potential to spread well beyond the breath of applications proposed here, and lead to dramatical changes in how research is performed.

REFERENCES

Akbari, B.O.S., Bellen, H.J., Bier, E., Bullock, S.L., Burt, A., Church, 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, F., Russell, S., Ueda, R. & Wildonger, J. (2015) Safeguarding gene drive experiments in the laboratory. Science (New York, NY), aac7932.

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 Synth Biol 3, 915–928.

Akbari, O.S., Matzen, K.D., Marshall, J.M., Huang, H., Ward, C.M. & Hay, B.A. (2013) A synthetic gene drive system for local, reversible modification and suppression of insect populations. Current biology : CB 23, 671–677.

Alphey, L., McKemey, A., Nimmo, D., Neira Oviedo, M., Lacroix, R., Matzen, K. & Beech, C. (2013) Genetic control of Aedes mosquitoes. Pathog Glob Health 107, 170–179.

Alphey, N. & Bonsall, M.B. (2014) Interplay of population genetics and dynamics in the genetic control of mosquitoes. Journal of the Royal Society, Interface / the Royal Society 11, 20131071–20131071.

Amenya, D.A., Bonizzoni, M., Isaacs, A.T., Jasinskiene, N., Chen, H., Marinotti, O., Yan, G. & James, A.A. (2010) Comparative fitness assessment of Anopheles stephensi transgenic lines receptive to site-specific integration. Insect molecular biology 19, 263–269.

Averof, M. & Cohen, S.M. (1997) Evolutionary origin of insect wings from ancestral gills. Nature 385, 627–630.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. & Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science (New York, NY) 315, 1709–1712.

Barrio, R. & de Celis, J.F. (2004) Regulation of spalt expression in the Drosophila wing blade in response to the Decapentaplegic signaling pathway. Proceedings of the National Academy of Sciences of the United States of America 101, 6021–6026.

Bassett, A.R. & Liu, J.-L. (2014) CRISPR/Cas9 and genome editing in Drosophila. Journal of Genetics and Genomics 41, 7–19.

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, 220–228.

Bertone, M.A. & Wiegmann, B.M. (2009) True flies (Diptera). The timetree of life. Available from: http://www.timetree.org/pdf/Bertone2009Chap33.pdf.

Bhalla, S. (1968) White eye a new sex-linked mutant of Aedes aegypti. Mosquito News 28, 380–385.

Biehs, B., Sturtevant, M.A. & Bier, E. (1998) Boundaries in the Drosophila wing imaginal disc organize vein-specific genetic programs. Development (Cambridge, England) 125, 4245–4257.

Bier, E. (2000) Drawing lines in the Drosophila wing: initiation of wing vein development. Current opinion in genetics & development 10, 393–398.

Blair, S.S. (1995) Compartments and appendage development in Drosophila. BioEssays 17, 299–309.

Burt, A. (2003) Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proceedings Biological sciences / The Royal Society 270, 921–928.

Burt, A. (2014) Heritable strategies for controlling insect vectors of disease. Philos Trans R Soc Lond B Biol Sci 369, 20130432–20130432.

Campbell, G. & Tomlinson, A. (1998) The roles of the homeobox genes aristaless and Distal-less in patterning the legs and wings of Drosophila. Development (Cambridge, England) 125, 4483–4493.

Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., Kafatos, F.C. & Crisanti, A. (2000) Stable germline transformation of the malaria mosquito Anopheles stephensi. Nature 405, 959–962.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N., Pollard, D.A., Sackton, T.B., Larracuente, A.M., Singh, N.D., Abad, J.P., Abt, D.N., Adryan, B., Aguade, M., Akashi, H., Anderson, W.W., Aquadro, C.F., Ardell, D.H., Arguello, R., Artieri, C.G., Barbash, D.A., Barker, D., Barsanti, P., Batterham, P., Batzoglou, S., Begun, D., Bhutkar, A., Blanco, E., Bosak, S.A., Bradley, R.K., Brand, A.D., Brent, M.R., Brooks, A.N., Brown, R.H., Butlin, R.K., Caggese, C., Calvi, B.R., Bernardo de Carvalho, A., Caspi, A., Castrezana, S., Celniker, S.E., Chang, J.L., Chapple, C., Chatterji, S., Chinwalla, A., Civetta, A., Clifton, S.W., Comeron, J.M., Costello, J.C., Coyne, J.A., Daub, J., David, R.G., Delcher, A.L., Delehaunty, K., Do, C.B., Ebling, H., Edwards, K., Eickbush, T., Evans, J.D., Filipski, A., Findeiß, S., Freyhult, E., Fulton, L., Fulton, R., Garcia, A.C.L., Gardiner, A., Garfield, D.A., Garvin, B.E., Gibson, G., Gilbert, D., Gnerre, S., Godfrey, J., Good, R., Gotea, V., Gravely, B., Greenberg, A.J., Griffiths-Jones, S., Gross, S., Guigo, R., Gustafson, E.A., Haerty, W., Hahn, M.W., Halligan, D.L., Halpern, A.L., Halter, G.M., Han, M.V., Heger, A., Hillier, L., Hinrichs, A.S., Holmes, I., Hoskins, R.A., Hubisz, M.J., Hultmark, D., Huntley, M.A., Jaffe, D.B., Jagadeeshan, S., Jeck, W.R., Johnson, J., Jones, C.D., Jordan, W.C., Karpen, G.H., Kataoka, E., Keightley, P.D., Kheradpour, P., Kirkness, E.F., Koerich, L.B., Kristiansen, K., Kudrna, D., Kulathinal, R.J., Kumar, S., Kwok, R., Lander, E., Langley, C.H., Lapoint, R., Lazzaro, B.P., Lee, S.-J., Levesque, L., Li, R., Lin, C.-F., Lin, M.F., Lindblad-Toh, K., Llopart, A., Long, M., Low, L., Lozovsky, E., Lu, J., Luo, M., Machado, C.A., Makalowski, W., Marzo, M., Matsuda, M., Matzkin, L., McAllister, B., McBride, C.S., McKernan, B., McKernan, K., Mendez-Lago, M., Minx, P., Mollenhauer, M.U., Montooth, K., Mount, S.M., Mu, X., Myers, E., Negre, B., Newfeld, S., Nielsen, R., Noor, M.A.F., O'Grady, P., Pachter, L., Papaceit, M., Parisi, M.J., Parisi, M., Parts, L., Pedersen, J.S., Pesole, G., Phillippy, A.M., Ponting, C.P., Pop, M., Porcelli, D., Powell, J.R., Prohaska, S., Pruitt, K., Puig, M., Quesneville, H., Ravi Ram, K., Rand, D., Rasmussen, M.D., Reed, L.K., Reenan, R., Reily, A., Remington, K.A., Rieger, T.T., Ritchie, M.G., Robin, C., Rogers, Y.-H., Rohde, C., Rozas, J., Rubenfield, M.J., Ruiz, A., Russo, S., Salzberg, S.L., Sanchez-Gracia, A., Saranga, D.J., Sato, H., Schaeffer, S.W., Schatz, M.C., Schlenke, T., Schwartz, R., Segarra, C., Singh, R.S., Sirot, L., Sirota, M., Sisneros, N.B., Smith, C.D., Smith, T.F., Spieth, J., Stage, D.E., Stark, A., Stephan, W., Strausberg, R.L., Strempel, S., Sturgill, D., Sutton, G., Sutton, G.G., Tao, W., Teichmann, S., Tobari, Y.N., Tomimura, Y., Tsolas, J.M., Valente, V.L.S., Venter, E., Craig Venter, J., Vicario, S., Vieira, F.G., Vilella, A.J., Villasante, A., Walenz, B., Wang, J., Wasserman, M., Watts, T., Wilson, D., Wilson, R.K.,

Wing, R.A., Wolfner, M.F., Wong, A., Ka-Shu Wong, G., Wu, C.-I., Wu, G., Yamamoto, D., Yang, H.-P., Yang, S.-P., Yorke, J.A., Yoshida, K., Zdobnov, E., Zhang, P., Zhang, Y., Zimin, A.V., Baldwin, J., Abdouelleil, A., Abdulkadir, J., Abebe, A., Abera, B., Abreu, J., Christophe Acer, S., Aftuck, L., Alexander, A., An, P., Anderson, E., Anderson, S., Arachi, H., Azer, M., Bachantsang, P., Barry, A., Bayul, T., Berlin, A., Bessette, D., Bloom, T., Blye, J., Boguslavskiy, L., Bonnet, C., Boukhgalter, B., Bourzgui, I., Brown, A., Cahill, P., Channer, S., Cheshatsang, Y., Chuda, L., Citroen, M., Collymore, A., Cooke, P., Costello, M., D'Aco, K., Daza, R., De Haan, G., DeGray, S., DeMaso, C., Dhargay, N., Dooley, K., Dooley, E., Doricent, M., Dorje, P., Dorjee, K., Dupes, A., Elong, R., Falk, J., Farina, A., Faro, S., Ferguson, D., Fisher, S., Foley, C.D., Franke, A., Friedrich, D., Gadbois, L., Gearin, G., Gearin, C.R., Giannoukos, G., Goode, T., Graham, J., Grandbois, E., Grewal, S., Gyaltsen, K., Hafez, N., Hagos, B., Hall, J., Henson, C., Hollinger, A., Honan, T., Huard, M.D., Hughes, L., Hurhula, B., Erii Husby, M., Kamat, A., Kanga, B., Kashin, S., Khazanovich, D., Kisner, P., Lance, K., Lara, M., Lee, W., Lennon, N., Letendre, F., LeVine, R., Lipovsky, A., Liu, X., Liu, J., Liu, S., Lokyitsang, T., Lokyitsang, Y., Lubonja, R., Lui, A., MacDonald, P., Magnisalis, V., Maru, K., Matthews, C., McCusker, W., McDonough, S., Mehta, T., Meldrim, J., Meneus, L., Mihai, O., Mihalev, A., Mihova, T., Mittelman, R., Mlenga, V., Montmayeur, A., Mulrain, L., Navidi, A., Naylor, J., Negash, T., Nguyen, T., Nguyen, N., Nicol, R., Norbu, C., Norbu, N., Novod, N., O'Neill, B., Osman, S., Markiewicz, E., Oyono, O.L., Patti, C., Phunkhang, P., Pierre, F., Priest, M., Raghuraman, S., Rege, F., Reyes, R., Rise, C., Rogov, P., Ross, K., Ryan, E., Settipalli, S., Shea, T., Sherpa, N., Shi, L., Shih, D., Sparrow, T., Spaulding, J., Stalker, J., Stange-Thomann, N., Stavropoulos, S., Stone, C., Strader, C., Tesfaye, S., Thomson, T., Thoulutsang, Y., Thoulutsang, D., Topham, K., Topping, I., Tsamla, T., Vassiliev, H., Vo, A., Wangchuk, T., Wangdi, T., Weiand, M., Wilkinson, J., Wilson, A., Yadav, S., Young, G., Yu, Q., Zembek, L., Zhong, D., Zimmer, A., Zwirko, Z., Jaffe, D.B., Alvarez, P., Brockman, W., Butler, J., Chin, C., Gnerre, S., Grabherr, M., Kleber, M., Mauceli, E. & MacCallum, I. (2007) Evolution of genes and genomes on the Drosophila phylogeny. Nature 450, 203–218.

Cook, O., Biehs, B. & Bier, E. (2004) brinker and optomotor-blind act coordinately to initiate development of the L5 wing vein primordium in Drosophila. Development (Cambridge, England) 131, 2113–2124.

Davidson, E.H. (2006) Gene Regulatory Networks and the Evolution of Animal Body Plans. Science (New York, NY) 311, 796–800.

de Celis, J.F. (2003) Pattern formation in theDrosophila wing: The development of the veins. BioEssays 25, 443–451.

de Celis, J.F., Barrio, R. & Kafatos, F.C. (1999) Regulation of the spalt/spalt-related gene complex and its function during sensory organ development in the Drosophila thorax. Development (Cambridge, England) 126, 2653–2662.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J. & Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607.

Deredec, A., Godfray, H.C.J. & Burt, A. (2011) Requirements for effective malaria control with homing endonuclease genes. Proceedings of the National Academy of Sciences of the United States of America 108, E874–80.

DiCarlo, J.E., Chavez, A., Dietz, S.L., Esvelt, K.M. & Church, G.M. (2015) RNA-guided gene drives can efficiently bias inheritance in wild yeast. bioRxiv, 1–13.

Erwin, D.H. & Davidson, E.H. (2009) The evolution of hierarchical gene regulatory networks. Nature reviews Genetics 10, 141–148.

Esvelt, K.M., Smidler, A.L., Catteruccia, F. & Church, G.M. (2014) Concerning RNA-guided gene drives for the alteration of wild populations. eLife, e03401.

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

Garcia-Bellido, A. & de Celis, J.F. (1992) Developmental genetics of the venation pattern of Drosophila. Annual review of genetics 26, 277–304.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J. & O'connor-Giles, K.M. (2013a) Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease. Genetics 194, 1029–1035.

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, 961–.

Gratz, S.J., Wildonger, J., Harrison, M. & O'Connor-Giles, K. (2013b) CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. Fly 7, 249.

Han, Q., Calvo, E., Marinotti, O., Fang, J., Rizzi, M., James, A.A. & Li, J. (2003) Analysis of the wild-type and mutant genes encoding the enzyme kynurenine monooxygenase of the yellow fever mosquito, Aedes aegypti. Insect molecular biology 12, 483–490.

Hastings, I.M. (1994) Selfish DNA as a Method of Pest Control. Philosophical Transactions: Biological Sciences 344, 313–324.

Hsu, P.D., Lander, E.S. & Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262–1278.

Huang, Y., Magori, K., Lloyd, A.L. & Gould, F. (2007) Introducing transgenes into insect populations using combined gene-drive strategies: modeling and analysis. Insect biochemistry and molecular biology 37, 1054–1063.

Isaacs, A.T., Jasinskiene, N., Tretiakov, M., Thiery, I., Zettor, A., Bourgouin, C. & James, A.A. (2012) Transgenic Anopheles stephensi coexpressing single-chain antibodies resist Plasmodium falciparum development. Proceedings of the National Academy of Sciences of the United States of America 109, E1922–30.

Isaacs, A.T., Li, F., Jasinskiene, N., Chen, X., Nirmala, X., Marinotti, O., Vinetz, J.M. & James, A.A. (2011) Engineered resistance to Plasmodium falciparum development in transgenic Anopheles stephensi. N. J. Besansky (Ed). PLoS Pathog 7, e1002017.

Ivanov, E.L., Sugawara, N., Fishman-Lobell, J. & Haber, J.E. (1996) Genetic requirements for the single-strand annealing pathway of double-strand break repair in Saccharomyces cerevisiae. Genetics 142, 693–704.

James, A.A. (2005) Gene drive systems in mosquitoes: rules of the road. Trends Parasitol 21, 64–67.

Kantorovitz, M.R., Kazemian, M., Kinston, S., Miranda-Saavedra, D., Zhu, Q., Robinson, G.E., Göttgens, B., Halfon, M.S. & Sinha, S. (2009) Motif-blind, genome-wide discovery of cis-regulatory modules in Drosophila and mouse. Developmental Cell 17, 568–579.

Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. & Kim, J.-S. (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Research 19, 1279–1288.

Kistler, K.E., Vosshall, L.B. & Matthews, B.J. (2015) Genome engineering with CRISPR-Cas9 in the mosquito Aedes aegypti. Cell reports 11, 51–60.

Lunde, K., Biehs, B., Nauber, U. & Bier, E. (1998) The knirps and knirps-related genes organize development of the second wing vein in Drosophila. Development (Cambridge, England) 125, 4145–4154.

Lunde, K., Trimble, J.L., Guichard, A., Guss, K.A., Nauber, U. & Bier, E. (2003) Activation of the knirps locus links patterning to morphogenesis of the second wing vein in Drosophila. Development (Cambridge, England) 130, 235–248.

Mali, P., Esvelt, K.M. & Church, G.M. (2013) Cas9 as a versatile tool for engineering biology. Nature Methods 10, 957–963.

Marshall, J.M. (2009) The effect of gene drive on containment of transgenic mosquitoes. J Theor Biol 258, 250–265.

Moreira, L.A., Ito, J., Ghosh, A., Devenport, M., Zieler, H., Abraham, E.G., Crisanti, A., Nolan, T., Catteruccia, F. & Jacobs-Lorena, M. (2002) Bee venom phospholipase inhibits malaria parasite development in transgenic mosquitoes. The Journal of biological chemistry 277, 40839–40843.

Nijhout, H.F. (1991) The development and evolution of butterfly wing patterns. Smithsonian series in comparative evolutionary biology (USA).

Nimmo, D.D., Alphey, L., Meredith, J.M. & Eggleston, P. (2006) High efficiency site-specific genetic engineering of the mosquito genome. Insect molecular biology 15, 129–136.

Nolan, T., Bower, T.M., Brown, A.E., Crisanti, A. & Catteruccia, F. (2002) piggyBac-mediated germline transformation of the malaria mosquito Anopheles stephensi using the red fluorescent protein dsRED as a selectable marker. The Journal of biological chemistry 277, 8759–8762.

Oye, K.A., Esvelt, K.M., Appleton, E., Catteruccia, F., Church, G.M., Kuiken, T., Lightfoot, S.B.-Y., McNamara, J., Smidler, A. & Collins, J.P. (2014) Biotechnology. Regulating gene drives. Science (New York, NY) 345, 626–628.

Patel, N.H., Martin-Blanco, E., Coleman, K.G., Poole, S.J., Ellis, M.C., Kornberg, T.B. & Goodman, C.S. (1989) Expression of engrailed proteins in arthropods, annelids, and chordates. Cell 58, 955–968.

Persson, U., Izumi, H., Souchelnytskyi, S., Itoh, S., Grimsby, S., Engström, U., Heldin, C.H., Funa, K. & Dijke, ten, P. (1998) The L45 loop in type I receptors for TGF-beta family members is a critical determinant in specifying Smad isoform activation. FEBS letters 434, 83–87.

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 of the United States of America 111, E2967–76.

Rafiqi, A.M. & Lemke, S. (2011) Megaselia abdita: Culturing and egg collection. Cold Spring Harbor ….

Rose, T.M., Henikoff, J.G. & Henikoff, S. (2003) CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. Nucleic acids research 31, 3763–3766.

Sander, J.D. & Joung, J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology 32, 347–355.

Shao, Y., Guan, Y., Wang, L., Qiu, Z., Liu, M., Chen, Y., Wu, L., Li, Y., Ma, X., Liu, M. & Li, D. (2014) CRISPR/Cas-mediated genome editing in the rat via direct injection of one-cell embryos. Nat Protoc 9, 2493–2512.

Sinkins, S.P. & Godfray, H.C.J. (2004) Use of Wolbachia to drive nuclear transgenes through insect populations. Proceedings Biological sciences / The Royal Society 271, 1421–1426.

Sinkins, S.P. & Gould, F. (2006) Gene drive systems for insect disease vectors. Nature reviews Genetics 7, 427–435.

Sivasankaran, R., Vigano, M.A., Müller, B., Affolter, M. & Basler, K. (2000) Direct transcriptional control of the Dpp target omb by the DNA binding protein Brinker. The EMBO journal 19, 6162–6172.

Skipper, K.A., Andersen, P.R., Sharma, N. & Mikkelsen, J.G. (2013) DNA transposon-based gene vehicles - scenes from an evolutionary drive. J Biomed Sci 20, 92.

Strigini, M. & Cohen, S.M. (1997) A Hedgehog activity gradient contributes to AP axial patterning of the Drosophila wing. Development (Cambridge, England) 124, 4697–4705.

Tomoyasu, Y., Arakane, Y., Kramer, K.J. & Denell, R.E. (2009) Repeated co-options of exoskeleton formation during wing-to-elytron evolution in beetles. Current biology : CB 19, 2057–2065.

Tomoyasu, Y., Wheeler, S. & Denell, R. (2005) Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum. Nature 433, 643–647.

Ward, C.M., Su, J.T., Huang, Y., Lloyd, A.L., Gould, F. & Hay, B.A. (2011) Medea selfish genetic elements as tools for altering traits of wild populations: a theoretical analysis. Evolution 65, 1149–1162.

Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. Nature.

Wiegmann, B., Trautwein, M., Winkler, I., Barr, N., Kim, J., Lambkin, C., Bertone, M., Cassel, B., Bayless, K. & Heimberg, A. (2011) Episodic radiations in the fly tree of life. Proceedings of the National Academy of Sciences 108, 5690–5695.

Windbichler, N., Menichelli, M., Papathanos, P.A., Thyme, S.B., Li, H., Ulge, U.Y., Hovde, B.T., Baker, D., Monnat, R.J., Burt, A. & Crisanti, A. (2011) A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature 473, 212–215.

Zhang, F., Wen, Y. & Guo, X. (2014) CRISPR/Cas9 for genome editing: progress, implications and challenges. Human molecular genetics 23, R40–R46.