University of Naples Federico II

Department of Biology

Ph.D. IN BIOLOGY
XXIX CYCLE

Genomics, transcriptomics and CRISPR/CAS9 gene editing approaches to investigate embryonic sex determination in insect pest *Ceratitis capitata*.

Coordinator:
Ch.mo Prof. Salvatore Cozzolino

Tutor:
Prof. Giuseppe Saccone

Ph.D. Student:
Angela Meccariello

Riassunto

Durante il mio dottorato di ricerca ho condotto studi di genomica e trascrittomica in un organismo non modello *Ceratitis capitata*. *Ceratitis capitata* (Diptera, Tephritidae), anche detta "medfly", è un dittero fitofago diffuso ormai in tutti i continenti e altamente dannoso per la coltivazione di molte specie vegetali che forniscono frutta per alimentazione umana. La sua vastissima diffusione, unita al fatto che si tratta di una specie polifaga, nonché le difficoltà riscontrate nel suo controllo ed eradicazione locale, hanno fatto di *C. capitata* una delle specie più dannose all'agricoltura a livello globale e quindi ad una significativa parte delle economie nazionali e di quelle internazionali. La riproduzione e la determinazione del sesso sono aspetti cruciali per capire la biologia degli insetti e bersagli ottimali per limitare la loro crescita e diffusione. Nel laboratorio in cui ho svolto la mia tesi di dottorato, da oltre vent'anni si è cercato di capire la cascata dei geni della determinazione del sesso e di caratterizzarli a livello molecolare in *Ceratitis capitata*. Uno studio evolutivo del pathway genetico che controlla la determinazione del sesso di *Ceratitis*, usando il sistema modello *Drosophila*, ha permesso di isolare quattro omologhi dei principali geni che determinano la scelta del sesso in *D. melanogaster: Sex-lethal* (Sxl), *transformer (tra)*, *transformer-2 (tra-2)* e *doublesex (dsx)*. Anni di studi hanno portato a dimostrare che i segnali primari della determinazione del sesso noti in *Drosophila* sono differenti in *C. capitata* e che il gene *Sex-lethal* non è coinvolto nel determinare la femminilità in *Ceratitis*. Il Male Determining Factor (Fattore M) localizzato sull'Y di Medfly ha un ruolo chiave nella determinazione del sesso, ma non è stato ancora isolato. Il progetto generale del mio Dottorato consiste in due linee di ricerca parallele e sinergiche: 1) approcci trascrittomici (Illumina) e genomici (PacBio) alla identificazione di nuovi geni di Medfly legati al cromosoma Y potenzialmente coinvolti nella determinazione del sesso maschile e di geni del cromosoma X; 2) sviluppo di una potente tecnica di gene editing, la CRISPR/Cas9, per studiare queste nuove funzioni geniche.

In questa tesi di dottorato mi sono focalizzata su due aspetti della ricerca genetica in *Ceratitis capitata* che hanno ramificazioni anche in campo applicativo:

- Il problema dell'identificazione di geni embrionali ad espressione maschio-specifica, tra i quali ci si aspetta di identificare il segnale primario mascolinizzante della determinazione del sesso, cioè il fattore M legato al cromosoma Y. A tal fine, abbiamo pensato di mettere a punto una tecnica poco utilizzata ma molto interessante: il "metabolic labelling" di trascritti di nuova sintesi in embrioni XX ed XY, per poterli purificare distinguendoli dai trascritti materni.
- La caratterizzazione del cluster di geni per le ceratotossine sul cromosoma X, molto eterocromatico. Il cromosoma X di *Ceratitis* è ritenuto essere povero di geni da precedenti studi genetici e citogenetici e ne è stata data conferma dall'analisi del suo genoma (Papanicolaou et al., 2016).
- Il problema della difficoltà di condurre studi genetici funzionali (inclusi i geni della determinazione del sesso) in un insetto non modello come la *Ceratitis*. La RNAi sia transiente che mediata da transgeni funziona bene in *Ceratitis*, ma è applicabile in genere su geni ad espressione embrionale (in transiente), e presenta difficoltà quando applicata con transgenesi perché la produzione di ceppi

transgenici è molto più laboriosa che in *Drosophila*, anche per il suo ampio genoma, che risulta ricco di zone eterocromatiche, le quali reprimono l'espressione del transgene. Abbiamo pensato di utilizzare la CRISPR/Cas9 per avere a disposizione un innovativo metodo di "genome editing" che permetta di indurre mutazioni ereditabili in specifici geni. Ho inoltre stabilizzato in laboratorio l'espressione batterica e la purificazione di Cas9 con his-tag. Infine, ho bersagliato con successo due geni in *Ceratitis* e ho contribuito allo sviluppo della CRISPR/Cas9 anche nell'altro dittero *Musca domestica*.

Summary

During my PhD, I have carried out studies of genomic and transcriptomic in the non-model system *Ceratitis capitata*. *C. capitata* (Diptera, Tephritidae), also called "medfly", is an insect pest that spread over almost every continent and represents a great danger to the cultivation of fruits for human alimentation. Its vast diffusion, together with its ability to parasitize many fruit hosts and the difficulties humans encountered when trying to control and eradicate it from local areas, made *C. capitata* one of the most dangerous species to agriculture on a global scale, and therefore, to a significant portion of national and international economies. The reproduction and the sex determination are two crucial aspects to understand the biology of insects and represent optimal targets to limit their growth and diffusion.

In the laboratory where I've conducted my PhD thesis, the identification and molecular characterization of the genes involved in the sex determination of *Ceratitis capitata* has been one of the main topics of research for more than 20 years. An evolutionary study of the sex determination pathway in *C. capitata* using the model system *Drosophila melanogaster* allowed to isolate four orthologues of the main genes involved in the sex determination pathway in *D. melanogaster*: *Sex-lethal* (*Sxl*), *transformer (tra)*, *transformer-2 (tra-2)* and *doublesex (dsx)*. Years of research have brought to demonstrate that the primary signals of sex determination known in *D. melanogaster* are different in *C. capitata* and that the *Sex-lethal* gene is not involved in the female sex determination in *C. capitata*. The male determining factor (M factor) localized on the Y chromosome in "medfly" has a key role in male sex determination, but still awaits to be molecularly isolated yet. The project of my PhD consists into two parallel and synergic research lines:

1) transcriptomic (Illumina®) and genomic (PacBio®) approaches to the identification of novel Y-linked genes potentially involved in male sex determination, and X-linked genes in "medfly";
2) the development and perfecting of a powerful technique of gene editing, the CRISPR/Cas9, to study these novel gene functions.

In this PhD thesis, I focused on three aspects of genetic research in *Ceratitis capitata*, which also have ramifications in the applicative field:

- I focused on the identification of embryonic genes with male-specific expression, amongst which it is expected to be found the primary masculinization signal of sex determination, namely the Y-linked M factor. For this purpose, I employed a novel and, I believe, promising technique: the

metabolic labelling of newly transcribed RNA in XX and XY embryos to purify and distinguish them from maternal transcripts. I have identified some hundreds of *Ceratitis* genes which seem to have zygotic expression during 5-6 hours of embryogenesis.

- I defined the genomic characterization of a gene cluster: *ceratotoxins,* on the X chromosome. Interestingly, in the medfly genome project, it has been shown from genetic and cytogenetic analyses that the X chromosome is highly heterochromatic and has only few genes (Papanicolaou et al., 2016).
- I approached the problem of conducting functional genetic studies, including on genes involved in the sex determination pathway, in a non-model system such as *C. capitata*. Transient as well as transgene-mediated RNA interference works in *C. capitata*, but it's generally applicable on genes with embryonic (transient) expression, and has some difficulties also when applied by transgenesis, because the production of transgenic flies in *C. capitata* is more laborious than in *D. melanogaster*, also for its larger genome, which is rich of heterochromatic regions, often repressing the expression of the integrated transgenes. I therefore applied CRISPR/Cas9 an innovative technique of gene editing that can allow to induce heritable mutations in specific genes and also, I established in the lab a Cas9-his tagged bacterial expression and purification method. I successfully targeted 2 *Ceratitis* genes and I contributed to CRISPR development also in the other dipteran *Musca domestica*,

INDEX

1 Introduction

1.1 Premise: Why study Insect pests?

1.2 Insect of interest: the agricultural pest *Ceratitis capitata*.

1.3 *Ceratitis capitata* life cycle

1.4 Sex determination in *Drosophila melanogaster* and *Ceratitis capitata*

1.5 Sex determination in other insects

1.6 Aim

2 Technique "metabolic labelling"

2.1 New molecular approach to isolate zygotic newly transcribed RNA

2.2 Results

2.3 Method

2.4 References

3 Technique CRISPR-CAS9

3.1 Highly efficient DNA-free gene disruption in the agricultural pest *Ceratitis capitata* by CRISPR-Cas9 RNPs. (paper in submission)

3.2 CRISPR-Cas9 targeted disruption of the *yellow* ortholog in the housefly identifies the *brown body* locus. (paper in submission)

4 Conclusions

1 Introduction

1.1 Premise: Why study Insect pests?

Insects are among the largest animal taxonomic group on Earth. Insects, as the majority of invertebrates, have a life strategy based on reproduction. They manage to completely colonize an ecosystem in a relatively short time. Insects have a critical role in both ecology and human society. In some cases, their interactions with humans can be harmful, and knowledge of the life cycle of insects that are vectors of disease and agricultural pests opens to new possibilities for developing eco-friendly alternatives to pesticide control strategies. Reproduction and sex determination are crucial aspects to understand the biology of insects and optimal targets for limiting their growth and dissemination at the same time. I decided to focus my PhD project on a major agricultural pest, *Ceratitis capitata* (Wiedemann, 1824), also called Mediterranean fruit fly ("medfly"), a widespread Diptera pest present in all continents and highly detrimental to crops of many species (Saccone et al., 2011). In particular, I have tried to fulfil some major gaps in the understanding of the genetics of sex determination in *Ceratitis* and in the reverse genetics techniques available to study *in vivo* gene functions.

1.2 Insect of interest: the agricultural pest *Ceratitis capitata*.

Ceratitis capitata is a highly invasive species to a wide range of cultivated and wild fruit, and has a huge economic impact. It has managed to adapt itself to vastly different conditions, and therefore occupy many ecological niches (to see distribution http://www.cabi.org/isc/datasheet/12367 Fig.1). The harmfulness of this species lies in the behaviour of females during reproduction. After the mating, they perforate the skin of many fruits with their extensible ovipositor. They prefer fruit with soft pulp, high sugar content and low acidity, to lay their eggs.

The eggs are slender, curved and shiny white (approximately 1.0 x 0.2 mm), and develop into larvae within 2-4 days in optimal conditions of temperature and humidity (Fig.2A). Larval development is divided into three stages and lasts for 4-6 days. The first stage larvae are less than half a millimetre long (Fig.2B); the last stage larvae (the third) are 7-9 mm long, and are also called "jumping" larvae (Fig.2C). The ability to jump allows the larvae to get out of the host fruit and pupate into the soil (Fig.2D). The pupae stage lasts for 6-11 days, and after that, the adults emerge (Fig.2E). *C. capitata* completes its life cycle in 21-30 days and the adults of *C. capitata* live for about 30 days.

The adult is 4-6 mm long and its body is coloured of shades ranging between grey and yellow. The head is yellow with red and iridescent eyes. The thorax is black with white spots features, while the abdomen is of a yellow-brown colour with dark streaks in the third and fifth abdominal segments. The wings show a characteristic colour in some areas, with one longitudinal and two transverse

bands of brown colour. Sexual maturity develops about two days after the eclosion, and mating of mature individuals follows a circadian cycle. The adults of *C. capitata* show an evident sexual dimorphism (Fig.3). Males and females are different at both morphological and physiological levels. The ovipositor is a peculiar characteristic of adult females, which are also bigger in size than the males. The males, instead, can be distinguished for the presence of a pair of supra-front-orbital bristles.

1.3 *Ceratitis capitata* life cycle

Ceratitis capitata is a holometabolous insect, with four life stages: embryo, larva, pupa and adult. The life cycle is the period of time from the oviposition to the adult stage, and in *Ceratitis capitata,* it lasts from 21 to 30 days in optimal temperature and climate conditions. The female lays from 1 to 10 eggs at a time in a 1 mm deep cavity, and can lay up to 22 eggs per day and 800 eggs in its entire life cycle. Females do not lay when the temperature is lower than 16 °C and the development at embryonic, larval, and pupal stages stops when temperature is lower than 10 °C.

The eggs hatch in 1.5-3 days in optimal temperature conditions. The length of the embryonic stage increases considerably with the lowering of the temperature. After the hatching, there are three larval stages, which can last up to 6-10 days when temperature ranges from 25-26.1 °C. Once matured at the 3rd instar larval stage, most of the larvae leave the fruit at dawn or just after dawn, and pupate into the ground, forming a brownish capsule in which metamorphosis will occur. The pupal stage lasts from 6-13 days when the mean temperature is between 24.4-26.1 °C. Most of the adults emerge from pupae in early morning, when temperature starts rising. They can fly for short distances, but can be transported by wind for kilometers of distance. Copulation can occur at any time of the day. The adults are not sexually mature at the moment of emersion from pupae; the males start showing sexual activity after 4 days from the emersion, whilst the females are ready to mate after 6-8 days from the emersion. The adults die in 4 days if they can't find any food. Usually, up to 50% of the flies dies in the first two months after the emersion. However, some adults may survive up until 6 months in conditions of optimal availability of food and water, and at cool temperatures.

1.5 Sex determination in *Drosophila melanogaster* and *Ceratitis capitata*

Sexual dimorphism in insects is a developmental output of two distinct processes: sex determination and sexual differentiation. Sex determination is controlled by genes that establish either the male or female sex during embryonic stages. These key genes, which represent the primary signals or respond to them, regulate a cascade of other genes composing the so-called sex determination

pathway, that acts through the expression of the sex-specific cytodifferentiation genes promoting the formation of male or female morphological structures and the development of sex-specific behaviours.

The sex determination pathway in *Ceratitis capitata* has been elucidated after evolutive and comparative studies conducted on *Drosophila melanogaster*, a model system for this kind of studies. In *D. melanogaster*, the master gene in the sex determination pathway is *Sex lethal* (*Sxl*), which codes for a splicing regulator protein. Transcription of *Sxl* is regulated in a sex-specific manner, dependent on the X:A ratio. Females, which have a double copy of the X chromosome, are homozygotes for four X-linked genes, called X-linked signalling elements (XSEs), whilst males are hemizygote for such genes. *Sxl* has two promoters: an early promoter, activated in females, and a late promoter, activated in males (Salz et al., 1989). This allows for the production of an inactive form in males and an active form in females of the SXL protein, respectively. At the blastodermal stage, the primary signal is no longer needed and the activity of SXL is maintained by autoregulation (Cline, 1984). The male transcript includes a male-specific exon, which contains an internal stop codon, resulting in the production of a truncated, non-functional SXL protein. The female transcript undergoes an event of alternative splicing, which splices out the male-specific exon, resulting in a functional SXL protein (Bell et al., 1988; Bopp et al., 1991) (Fig.4). In females, the functional SXL protein regulates the splicing of the downstream *transformer* gene (*tra*), allowing the production of a functional TRA protein. The *tra* pre-mRNA has two 3' splice sites: a female-specific splice site, and a non-sex-specific splice site. About 50% of the *tra* pre-mRNA undergoes a female-specific splicing event in the females. The female-specific splicing of the pre-mRNA splices out an exon containing a stop codon, thus allowing the production of a functional, full length TRA protein. In males, only the non-sex-specific splicing event occurs, producing a truncated, non-functional TRA protein (Boggs et al., 1987; Belote et al., 1989; Válcarcel et al., 1993) (Fig.5).

The TRA protein, expressed only in XX *Drosophila* embryos, continues the sex-determination pathway in a female-specific way. In a complex with the TRA-2 protein, coded by the autosomal gene *transformer-2*, TRA determines the activation of the downstream genes in the pathway: *doublesex* and *fruitless*, which are expressed in two different isoforms in the two sexes due to alternative splicing events. TRA and TRA-2 recruit general splicing factors, such as the splicing regulator proteins (SR), onto regulative elements, named *dsx* repeat elements (DsxRE). The DsxRE are TRA/TRA-2 binding elements repeated 6 times in the *dsx* pre-mRNA; they are highly conserved, 13 nt sequences located downstream of a *dsx* female-specific 3' splice site (Tian and Maniatis, 1993). In the absence of TRA, the splicing produces a male-specific DSX isoform

(DSXM) (Burtis and Baker, 1989), which represses the expression of the genes that direct sexual development in a female-specific way, both in morphology and behaviour. At the same time, DSXM activates the expression of the male-specific downstream genes, which direct the sexual development in a male-specific way. *Vice versa*, the female-specific DSX isoform (DSXF) represses the expression of the genes that induce male-specific sexual determination and activates the expression of the genes that induce female-specific sexual determination. *Fruitless* (*fru*) is another gene that contains a TRA/TRA-2 binding element, which is regulated by the TRA/TRA-2 complex in a sex-specific way during splicing. The male isoform of *fruitless* is responsible in conjunction with DSXM for the sexual development of the central nervous system (Rideout et al., 2007) (Fig.6).

A series of studies showed that sex determination in medfly is different from sex determination in *Drosophila*. The identification of fertile XXX females and fertile XXY males in a wild population of medfly showed that the presence of the Y chromosome is determining for male sex determination (Saccone et al., 2002). The analysis of the progeny generated from an autosome-Y reciprocal translocation T(Y;2) has allowed the prediction of the existence of one or more than one male determining factors on the long arm of the Y chromosome (Robinson, Franz and Fisher, 1999). Using a series of deletions on the Y chromosome, the male determining factor has been located in a region of the long arm, near the centromere, representing about 15% of the entire Y chromosome. Moreover, this study showed that the remaining 85% of the Y chromosome does not contain genes involved in sex determination, testis development and male fertility (Willhoeft and Franz, 1996). The Y chromosome of medfly also contains a repetitive element, which was identified by techniques of molecular subtraction (Anleitner and Haymer, 1992). The positional information obtained by deletion analysis of the Y chromosome could be a useful and necessary preliminary step to clone the male determining factor(s) by microdissection (Willhoeft & Franz, 1996).

A comparative molecular study has been carried out to identify genes involved in sex determination in *Ceratitis* using *Drosophila* genes as probes for cDNA and genomic libraries of medfly. This approach has brought to the isolation of orthologues of *Sxl* and *dsx* in medfly (Furia et al., 1992; Saccone et al., 1996; Saccone et al., 1998). In contrast with *Drosophila*, the orthologue of *Sxl* (*CcSxl*) expresses the same mRNA and the same protein isoforms in both XX and XY individuals, independently from the primary sex determination signal (Saccone, 1997; Saccone et al., 1998). Moreover, experiments involving two inducible transgenes showed that the product of *CcSxl* does not have significant effects on sex determination when expressed in *Drosophila* (Saccone et al., 1998) (Fig.7).

The *dsx* gene in medfly, *Ccdsx*, produces sex-specific transcripts through alternative splicing,

similarly to *Drosophila*, suggesting its functional conservation as regulator in sex determination (Saccone et al., 1996). Indeed, the ectopic expression of the male-specific isoform of CcDsx (CcDsxM) in *Drosophila* transgenic strands has induced a partial masculinization of XX flies, supporting the idea that *dsx* controls the sex determination in medfly, similarly to *Drosophila* (Saccone et al., 2008). Moreover, the sex-specific splicing regulation of the *Ccdsx* pre-mRNA seems to be homologous to that of *dsx* in *Drosophila*.

The orthologue of *dsx* in medfly shows two conserved characteristics:

1. the 3'-UTR of the female-specific exon in medfly has conserved short sequences made up by five fragments of 13 nucleotides each, corresponding to the TRA/TRA-2 repeated elements (dsxRE; also known as Tra/Tra-2 binding sites), identified in *Drosophila*. The evolutionary conservation of these elements in medfly inside the female-specific non-coding 3' region suggests the existence of a positive selection for the specific binding of trans-agent factors homologous to TRA and TRA-2 (Saccone, 1997; Saccone et al., 2002);
2. as in *Drosophila*, the female-specific exon in medfly is preceded by a 3' splice site, which has the characteristics of a suboptimal splice site (Saccone et al., 1996). The conservation of these cis-agent regulatory elements suggests that the regulation of the sex-specific splicing of the *dsx* gene in medfly is homologous to that in *Drosophila* and that, as in *Drosophila*, the male-specific splicing is the default splicing and the female-specific *dsx* splicing needs TRA/TRA-2 bound to the cis regulatory elements to enhance it.

Considering these data, a model for sex determination in medfly can be proposed, in which only a part of the sex determination pathway of *Drosophila* is evolutionarily conserved. Possibly, *CcSxl* can play only, if any, auxiliary roles in sex determination in medfly, as TRA-2 does in *Drosophila*, which, although expressed constitutively in a non-sex-specific way, works in conjunction with the female-specific isoform of TRA. On the other hand, *Ccdsx* shows evolutionary conservation not only in its nucleotide and amino acid sequences, but also in its sex-specific regulation, indicating its involvement in the control of sex differentiation. The female-specific splicing control of the *Ccdsx* pre-mRNA is exerted by homologous TRA/TRA-2 protein in medfly (CcTRA/CcTRA-2), which bind to dsxRE, a conserved splicing enhancer region. The discovery of dsxRE elements inside and in proximity of male-specific exons of *Cctra* has showed that the *CcTra* gene, in contrast to *tra* in *Drosophila*, is capable of auto-regulation (Pane et al., 2002).

It's known that in *Ceratitis*, the male-specific splicing of *tra* starts to occur at about 6-7 hours after oviposition in XY embryos. Its pre-mRNA comprises 5 exons (while only 3 exons in *Drosophila*), three of which (the first, the fourth and the fifth) are included in the mature transcript in both sexes, whilst the other two are only included in the male-specific transcript. In *Ceratitis*, its regulation is

based on the alternative splicing by exon skipping (of male exons in females) and on a differential and combined use of alternative regulative sites, 5' donor and 3' acceptor splice sites, which are present in the initial part of the gene. In XY medfly embryos a male-specific splicing produces two main transcripts in which two exons containing stop codons are retained, thus generating truncated proteins. Instead, in XX individuals, the alternative splicing excludes the second and the third exons, which are male-specific, from the mature transcript, allowing the production of a full-length protein, which is 429 amino acid long and called CcTRAF (Pane et al., 2002). The sex determination model in medfly, based on the key function of *Cctra*, supports that the female-specific CcTRA protein, in conjunction with CcTRA-2 (expressed constitutively in both sexes), acts as a splicing factor binding the *Cctra-dsx*RE, determining the inhibition of the male-specific splicing of *Cctra* and the regulation of the downstream genes beginning with *Ccdsx* (Pane et al., 2002). CcTRA acts on the pre-mRNA of *Ccdsx*, activating the female-specific splicing and generating two sex-specific protein isoforms (DSXF and DSXM), structurally and functionally conserved in comparison with *Drosophila melanogaster* (Saccone et al., 2010). Ultimately, it is possible to propose that in embryonic stages sex determination in *Ceratitis* is directed at the male sex through the action of the male determining factor, which promotes or simply allows the inclusion of the male-specific exons, probably by blocking the action of the CcTRA maternal protein (and possibly of CcTRA-2). Female sex determination relies on the action of a splicing complex which promotes an exon-skipping and is maintained by an epigenetic auto-regulative mechanism sustained by CcTRA activity. This mechanism can be suppressed by removing, even temporarily, mRNA coding CcTRA or CcTRA-2 protein with transient embryonic RNAi (Pane et al., 2002; Salvemini et al., 2009). A transient interference in the expression of these two genes in XX embryos determines the complete sexual transformation (at both the germinal and somatic levels), obtaining pseudo-male fertile XX adult flies. The positive outcome of the utilization of RNA interference against the *tra* gene in *Ceratitis capitata* obtained by microinjection of double-stranded RNA molecules in embryos at the very first developmental stages has given the clear confirmation of the role of *Cctra* as key gene in sex determination and has also sustained the model based on positive feedback (Pane et al., 2002). Indeed, an RNAi-induced temporary absence (for about 24 hours) of the *tra* transcripts both in XY male embryos and XX female embryos, causes an irreversible change in the splicing of *Cctra*, which becomes male-specific in the latter. This shift, that happens only in XX individuals, brings to the development of XX male adult flies, which have also happened to be fertile (Pane et al., 2002). The male determining factor could inhibit the activity of CcTRA and/or CcTRA-2 in males both at the level of transcription, or at the post-transcriptional level, determining the male-specific splicing of *Ccdsx* indirectly. In this case, we would expect the

sex-specific regulation of *Cctra* to be exerted by the male determining factor, and not by *CcSxl*, as a unique property of medfly. Alternatively, the male determining factor could also bind the *Ccdsx* pre-mRNA and directly influence its splicing, competing with TRA/TRA-2 for the binding with the splicing enhancer dsxRE. Ultimately, other genes could be interposed between the male determining factor and *Ccra*/*CcTra-2*, or even in parallel, increasing the complexity of the sex determination pathway in medfly. The isolation in medfly of the Y-linked male determining factor could help elucidate the validity of this model and localize the points of molecular divergence between the sex determination pathways in *Drosophila* and medfly.

As a novel method of sexing, alternatively to the female-specific lethality of the TSL strain, a method based on the transformation of females into males at the embryonic stage has been obtained with the utilization of RNAi targeted against the *Cctra* gene and a transgene stably inserted into the medfly genome (Saccone et al., 2007). In this strain, a transgene allows the production of dsRNA molecules specific for the *Cctra* gene during oogenesis (enhancer effect in proximity of the site of integration of the transforming vector; Petrella, 2014), so that in female transgenic embryos the repression of the *transformer* gene occurs, thus provoking a male-specific *Cctra* splicing in XX embryos, which is irreversible until adulthood, and inducing masculinization of XX individuals. Interestingly, one transgene copy of the construct producing *Cctra*-specific dsRNA in the mothers crossed with non-transgenic fathers is sufficient for masculinization of all embryos, including also the non-transgenic ones.

It's interesting to note that the XX male obtained with RNAi in this genetic sexing strain of *Ceratitis* are able to compete with wild males and their progeny is fertile when they are crossed with XX females (Petrella, 2014). The progeny of the cross between XX males and females is exclusively made up of XX embryos, which will develop into adult females. The fertility of XX males has been previously utilized to produce a progeny of only XX female embryos in order to carry out: 1) studies of molecular subtraction of male-specific transcripts using RNA from XX embryos alone and RNA from mixed XX/XY embryos (Salvemini et al., 2014); 2) transcriptome analysis on XX and XX/XY embryos (Salvemini, Nagarju, Arunkumar, Robinson, and Saccone, 2010-2017, unp. Res.).

1.5 Sex determination in other insects

Sex determination in insects is classified into three main categories depending on the different primary sex determining signals: 1) zygotic, 2) maternal, or 3) environmental signals (Sánchez, 2008). In *Drosophila melanogaster*, the primary signal is based upon the quantity of the product of four X-linked genes: the X-linked signalling elements (XSEs) (Erickson & Quintero, 2007). In

some species (apparently most of them), in which the female is the homomorphic sex (XX) and the male is the heteromorphic sex (XY), the primary signal is a male-determining factor, coded by a gene located on the Y chromosome. These species include the tephridites (*Ceratitis*, *Bactrocera* and *Anastrepha*). However, in *Musca domestica* a candidate M-factor has recently been located on an autosome. In species where the homomorphic sex is the male (ZZ) and the heteromorphic sex is the female (ZW), the primary signal for sex determination varies. In some Lepidoptera species (butterflies and moths), the primary signal depends on a Z-counting mechanism, similarly to the X-counting mechanism of *Drosophila*. In the silkworm, the primary signal is a Fem factor located on the W chromosome (Traut et al., 2007). In *Apis*, instead, the sex determination relies on a different mechanism: the females are diploid and the males are haploid. Finally, in some insect species sex determination depends on environmental factors, such as the temperature during the embryonic development. This mechanism has been found in some *Sciara* species (Ruiz et al., 2005). It's interesting to note that in the last few years, primary signals of sex determination have been identified in other species of insects: in *Drosophila*, the XSE (X-linked signaling elements, transcriptional factors which are present in XX embryos and able to activate the early transcription of *Sxl* from an alternative promoter, thus allowing its female-specific splicing); in *Apis mellifera*, CSD (Complementary Sex Determination, a gene with many alleles, which codes for a serine and arginine rich protein, similar to CcTRA protein (but not DmTRA protein); in *Bombyx mori*, the *fem* gene codes for a W-linked piRNA, which acts on the *Masculinizer* (*Masc*) gene, to promote femaleness in ZW individuals. In *Aedes aegypti*, the *nix* gene (*Nix* codes for a potential splicing factor related to *tra-2*, and the absence of *Nix* shifts the alternative splicing of *dsx* and *fru* towards female-specific isoforms); *in Anopheles gambiae*, the Y-linked *Yob* gene codes for a novel short protein with no conserved domains; *Yob* is expressed exclusively in male embryos within 2 hours from oviposition, and its transcription is maintained for the rest of the life. If *Yob* is injected in early embryos before two hours, only XY male mosquitos will be born, because the presence of *Yob* is lethal for female embryos, most likely for dosage compensation unbalance.

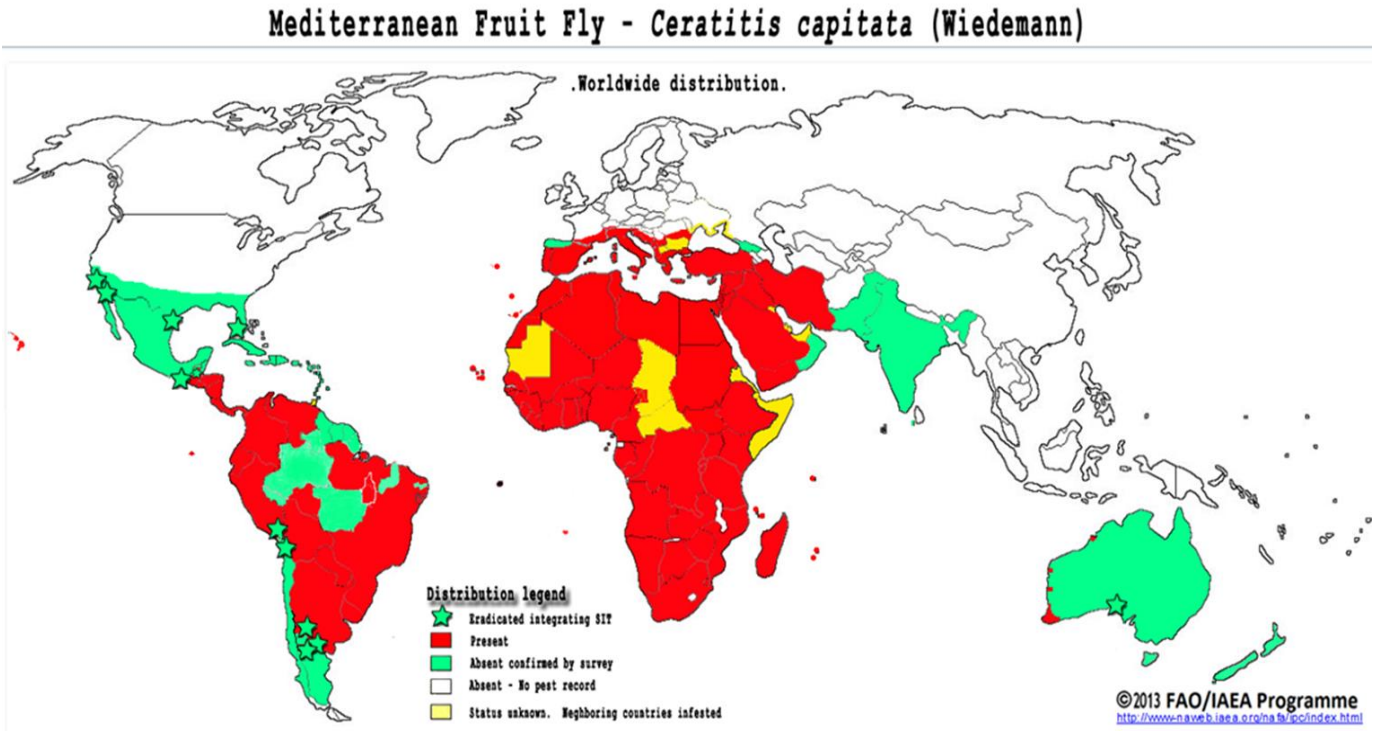


Fig.1 Distribution map of the Mediterranean fruit fly *Ceratitis capitata* (Wied.) (Designed by FAO/IAEA, December 2013)

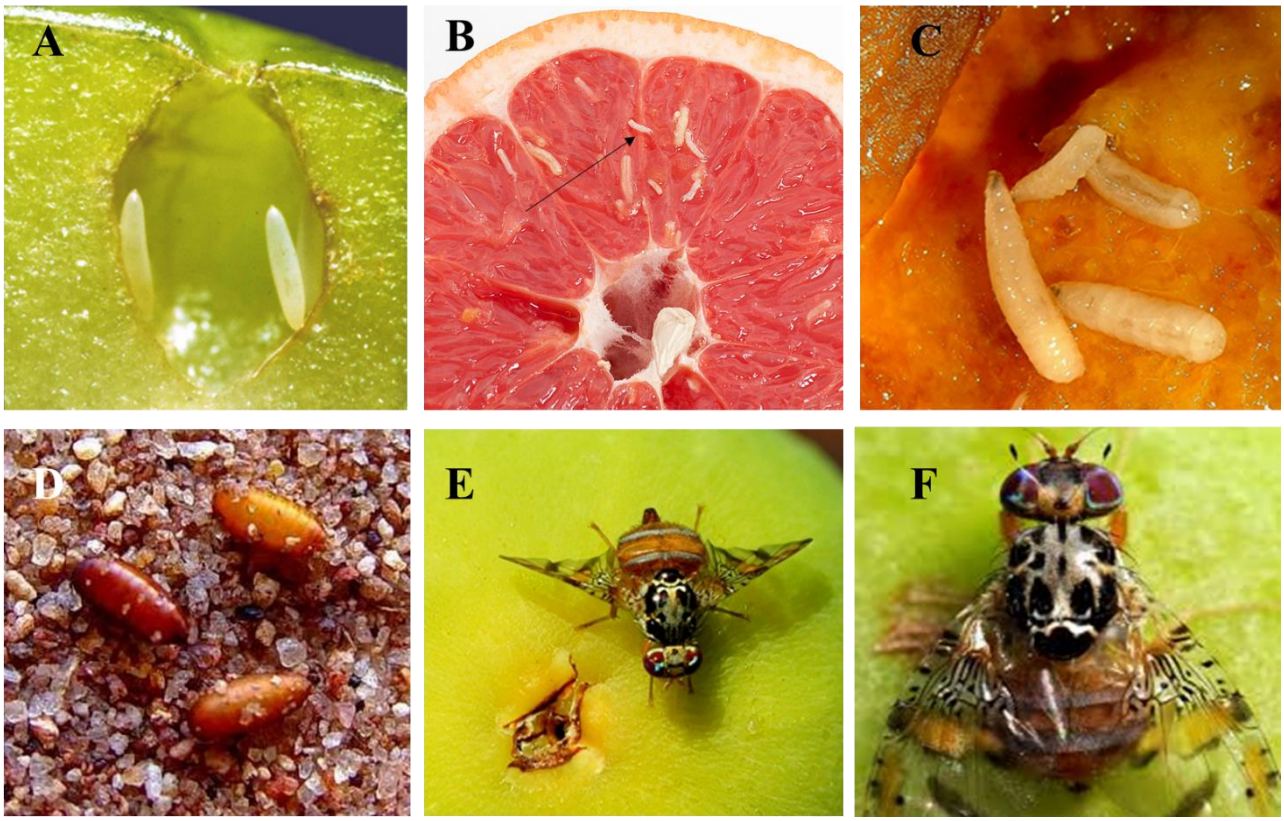


Fig.2 Eggs of *Ceratitis capitata* **(A)**; larvae of the first stage **(B)**; "jumping" larvae **(C)**; the pupal stage **(D)**; adult female **(E)**; adult male **(F)**.

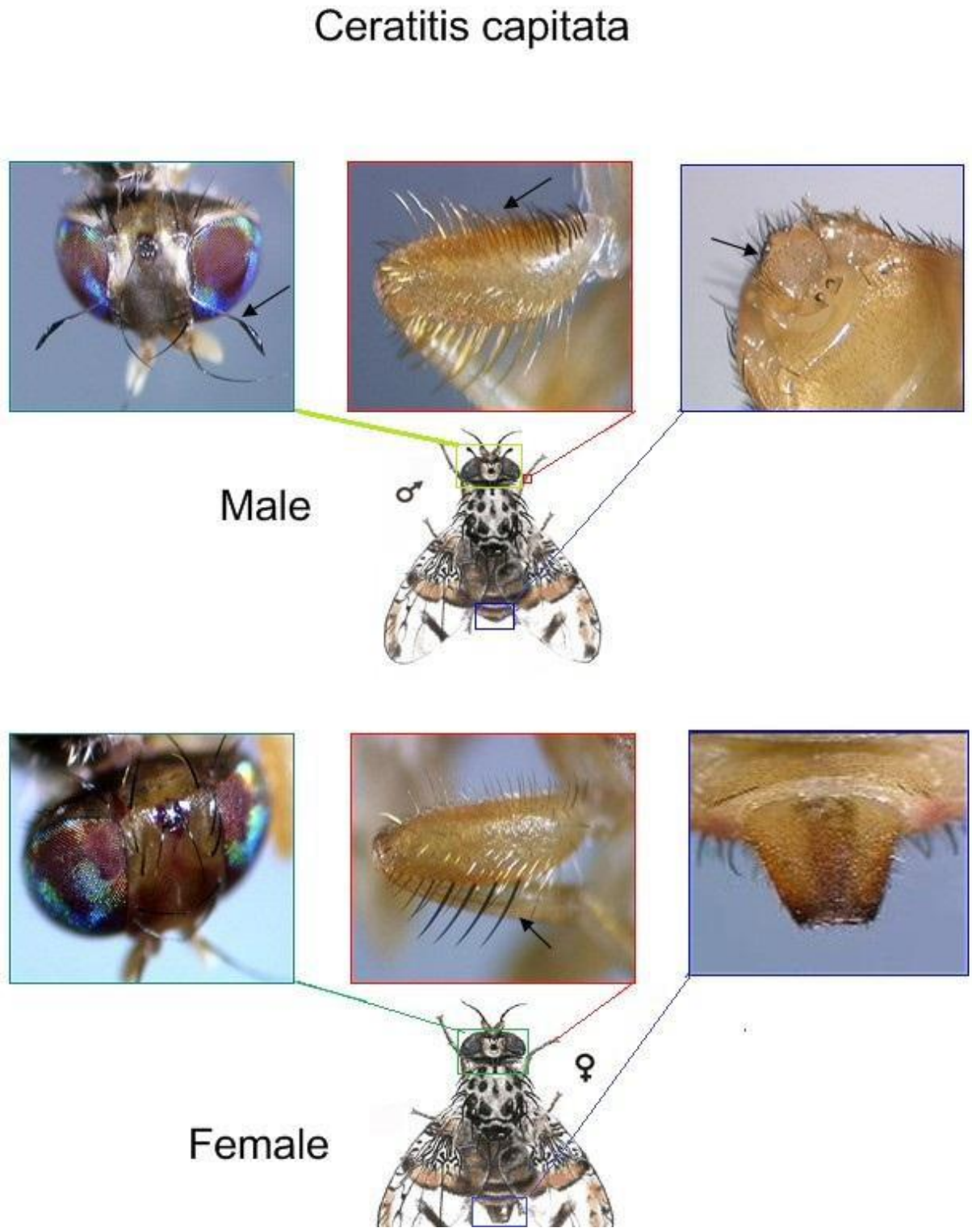


Fig.3 Sexual Dimorphism of *Ceratitis capitata*

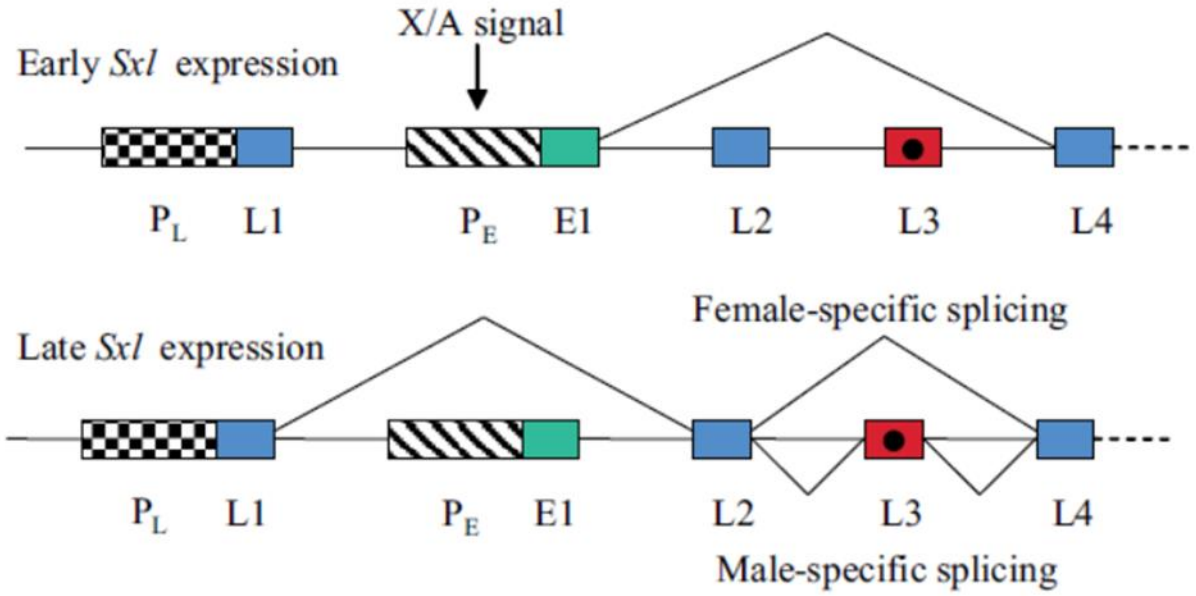


Fig.4 Gene *Sex lethal* (*Sxl*). The image shows the two sex-specific splicing events that occur in the *Sxl* gene, following the activation of the early promoter in the female and the late promoter in the male. Adapted from Sex-determining mechanisms in insects (Sánchez, 2008).

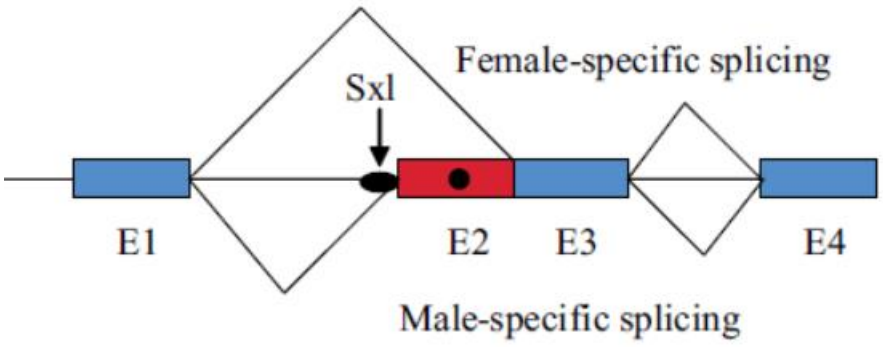


Fig.5 Gene *Transformer* (*Tra*). The image shows the female-specific splicing, which allows the production of a full length TRA protein, and the non-sex-specific splicing, which produces a truncated protein. The non-sex-specific splicing is the only splicing event that occurs in males (Sánchez, 2008).

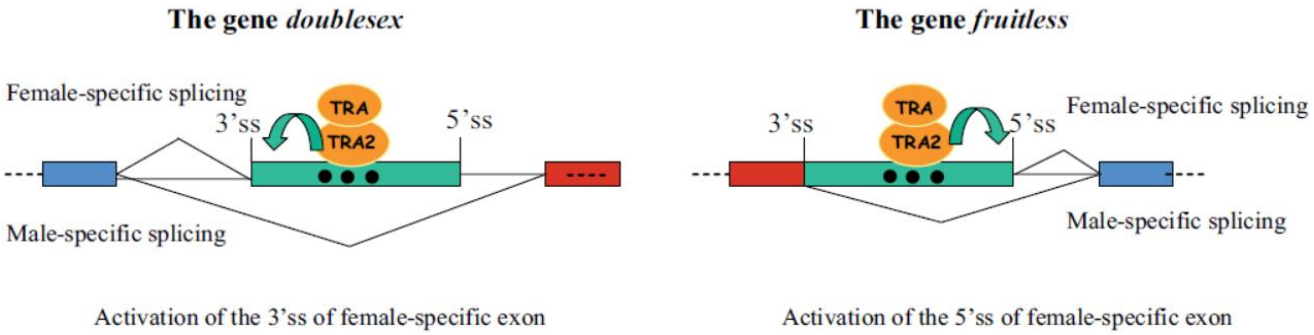


Fig.6 On the left: the gene *doublesex.* On the right: the gene *fruitless.* The image shows the female-specific splicing and male-specific splicing, respectively (Sánchez, 2008).

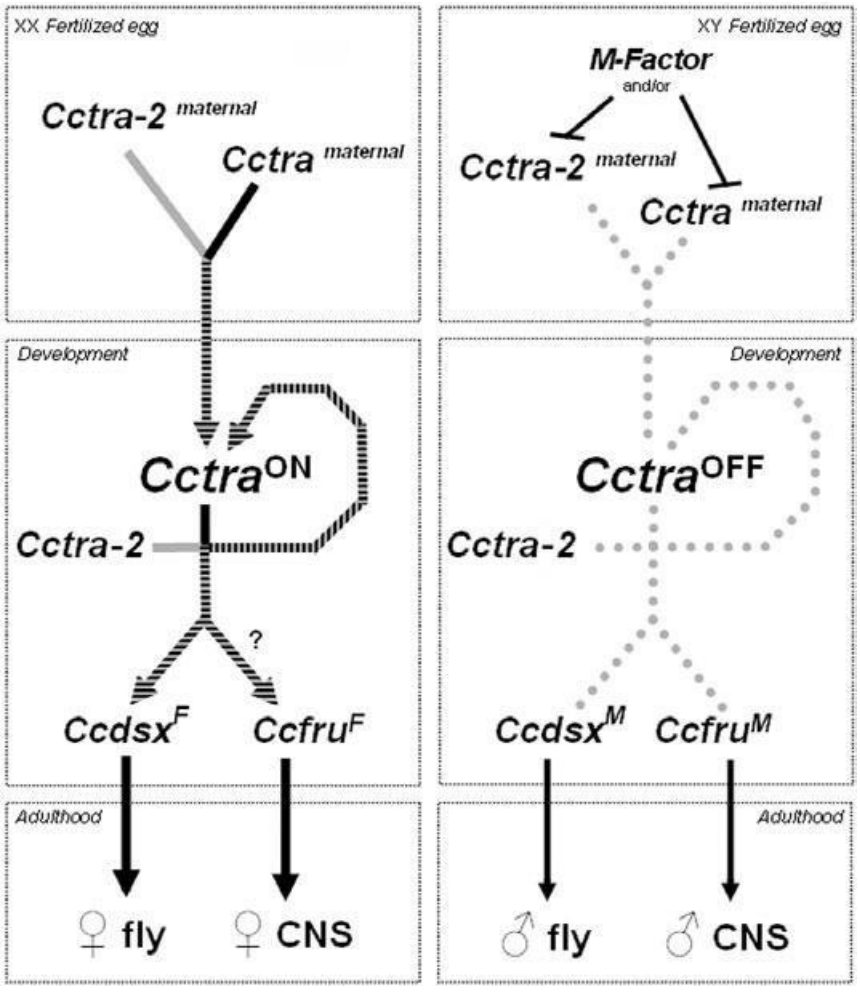


Fig.7 Sex determination mechanism in *Ceratitis capitata*.

1.6 A premise to my Aims

In the laboratory where I've conducted my PhD thesis, the genes of the regulative pathway of sex determination in insects have been studied for more than twenty years; they've been characterized at a molecular level in *Ceratitis capitata* (Medfly), an insect pest dangerous to agriculture. Years of studies have brought to demonstrate that the primary signals of sex determination known in *Drosophila* are different in *C. capitata* and that the *Sex lethal* gene is not involved in the female sex determination in *Ceratitis*. The male determining factor (M factor) located on the Y chromosome in Medfly has a key role in sex determination, but it hasn't been isolated yet. The vast majority of the analysis carried out in this direction, over the last twenty years, hasn't brought to definitive results. Only the functional investigations carried out in three fly strains, bearing translocations of parts of the Y chromosome on an autosome, have shown that a region of the long arm of the Y chromosome in *Ceratitis* is essential for the determination of the male sex (Zapater & Robinson, 1986), and have subsequently defined a region, representing the 15% of the entire chromosome, in which is located the M factor in Medfly (Willhoeft & Franz, 1996). Both sex chromosomes of *Ceratitis* (X and Y) are highly heterochromatic and seem or are expected to contain only few genes. Novel DNA sequencing technologies such as "PacBio", which lead to obtain single long reads of 10-20 Kb, are opening novel opportunities to assemble and analyze such difficult genomic regions.

To better understand medfly sex chromosome and sex determination, I have studied few known X-linked and I have isolated novel Y-linked genes by genomics/transcriptomics approaches. I have structurally characterized a cluster of X-linked paralogous genes, the *ceratotoxins* (*ctx*) as my contribution to the study of the medfly genome, which has been recently published (Papanicolaou et al., 2016). I have also developed a cost-effective CRISPR/Cas9 protocol, based on the expression/purification of His-tagged Cas9 protein, *in vitro* transcription of small guide RNAs (sgRNAs), and microinjections of pre-loaded Cas9 ribonucleoproteins with sgRNAs into insect embryos, including *Musca domestica* and *Ceratitis capitata*. One study has been submitted (Heinze et al., 2017) and the second one is going to be submitted within the next few weeks (Meccariello et al., 2017).

In my PhD project, I've used the transgenic *Ceratitis* sexing strain (see introduction; *in vivo* RNAi against *Cctransformer* to masculinize XX individuals) to produce a progeny of only XX female embryos in order to obtain female embryo transcriptomes and compare them with XX and XY mixed embryos. In particular, I've focused on an earlier period of *Ceratitis* development (0-6 hours from oviposition) with the respect of previously produced transcriptomes in our lab (4-8 h; 8-10h; 23-25h; Salvemini, Arunkumar, Robinson, Petrella, Saccone, unp. Res.); furthermore, I've prepared

total RNA from embryos instead of polyadenylated RNA, and I've tried to enrich the samples with transcripts produced in the zygote and developing embryo, at the expense of maternal transcripts. The reason why to enrich genes with zygotic transcription is that Y-linked genes can be expressed in embryos only at a zygotic level (obviously with no maternal contribution), whereas X-linked and autosomal genes can contribute maternally in the developing egg. For example, in *Drosophila* over 50% of the genes (about 7,000 of 15,000) contributes with maternal transcripts, whereas newly formed zygotic transcripts are produced of some hundreds within the first hours AEL (after egg laying). Therefore, it's not easy to discriminate zygotic transcripts from maternal ones, if not with molecular techniques that enrich the former.

It would be desirable to isolate the M factor, not only to understand the molecular identity of a primary Y-linked signal of male sex determination and deepen our understanding of its evolution, but also to develop a transgenic strain in which is possible to masculinize females expressing the M factor conditionally (eg. with a heat-shock promoter). The isolation of the M factor has been attempted in medfly using molecular subtraction analysis applied to RNA samples extracted from XX only, and XX/XY mixed embryos, respectively (Salvemini et al., 2014). From the comparison, a few number of putative Y-linked genes, which are transcribed early in XY embryos, have been identified, although their function is still in course of definition (Tu, J, Brantley, Salvemini and Saccone, unpub. Res.). Transcriptome profiling using RNAseq and differential expression analysis have been carried out and are still in course on XX and XX/XY embryos, to conduct other experiments of molecular subtraction, this time *in silico* (Salvemini et al., 2014).

My aims

The project of my PhD consists in the genetic study of sex determination in *Ceratitis* at embryonic stages, using innovative transcriptomic techniques (such as Illumina sequencing), genomic techniques (PacBio sequencing), as well as gene editing techniques (CRISPR/Cas9). I've developed my project within a web of international collaborations which includes for the most part the work the Zurich University (Prof. Daniel Bopp and Prof. C. Mosimann for CRISPR/Cas9 in medfly and *Musca domestica*; Prof. Mark Robinson for *PacBio genome sequencing and Illumina sequencing*), and also the FAO/IAEA Pest Control laboratory (Dr. Kostas Bourtzis; Seiberdorf, Austria) and the CDFD (Centre for DNA Fingerprinting and Diagnostics; Dr. Arunkumar) in Hyberabad (India).

My thesis project is focused on two main aspects of genetic research in *Ceratitis capitata*, which also have ramifications in the applicative field:

1) the problem of identification of embryonic genes with male-specific expression, among which it is expected to find the Y-linked M factor. For this purpose, we started to set up in the lab a relatively novel but complex technique: the metabolic labelling of newly transcribed RNAs in XX and XY embryos to distinguish them from maternal transcripts.
2) the problem of conducting functional genetic studies (including genes involved in the sex determination) in a non-model insect system such as *Ceratitis*. Transient embryonic RNAi worked efficiently in *Ceratitis* but only for few genes with embryonic expression and involved in autoregulatory mechanisms. Transgenesis is possible in the medfly but the production of transgenic strains is much more elaborated than in *Drosophila*, also because of the larger genome, which is enriched in heterochromatic regions and hence because there is the need to screen large number of transgenic lines to find expected expression. The insertions of transgenes in the medfly genome often present problems with transgene expression. I proposed to establish CRISPR/Cas9 as an innovative method of gene editing that allowed us to induce gene-specific and heritable mutations.

2 Technique: “metabolic labelling”

2.1 New molecular approach to isolate zygotic newly transcribed RNA

The development of whole-transcriptome by next-generation sequencing is revolutionizing our understanding of the complexity of developmental gene expression during early embryogenesis of insects (Jiménez-Guri et al., 2013). Such complexity could be partially reduced by distinguishing newly zygotic transcripts from maternally deposited mRNAs (Paris et al., 2015). This could help to identify key master genes for sex determination such as the Y-linked M factor of medfly.

It’s been known for more than 30 years that nucleotides containing thiolic groups, such as **4-thiouridine (4sU),** can be introduced in salvage pathways in eukaryotic cells, allowing the metabolic labeling of newly transcribed RNA (Melvin et al. 1978). These can be used to separate newly transcribed RNA from total RNA using mercury affinity chromatography (Melvin et al. 1978; Woodford et al. 1988) or thiol-specific biotinylation followed by purification with streptavidin-coated magnetic beads (Cleary et al. 2005) as eukaryotic mRNA do not normally contain thiolic groups.

Cleary et al. 2005 used **thio-substituted compound 2,4-dithiouracil (DTU)** which is a substrate for *Toxoplasma gondii* salvage enzyme uracil phosphoribosyl transferase (UPRT) in assays using cell lysates: indeed in *Toxoplasma gondii* cells expressing uracil-phosphoribosyltransferase (UPRT), newly transcribed RNA can be metabolically labeled using **4-thiouracil (4tU). This** 4TU/UPRT-based biosynthetic labeling was defined as “**TU**-tagging” by the authors.

Although some methods to purify tissue-specific RNA are already available (Roy et al., 2002; Tanke et al., 1993; Doyle et al. 2008; Heiman et al., 2008), each of them has its own restrictions and researchers are often confronted with these limitations when purifying RNA from cell types of interest. UPRT can be used to label biosynthetically newly transcribed RNA *in vivo* (Cleary et al., 2005). In natural conditions, UPRT couples the ribose 5-phosphate to the N1 nitrogen of uracil to form uridine monophosphate (UMP), which is subsequently incorporated in RNA. When given a uracil analog such as 4tU as substrate, UPRT incorporates it in the RNA, and this has little effects on cell physiology (Cleary et al., 2005). RNA resulting from the labeling with thiol nucleotides can be labeled and purified using commercially available reagents. For its ability to isolate newly synthetized RNAs from stable cell RNAs, this method has been used to measure the rate of synthesis and decay of RNA (Cleary et al., 2005; Dolken et al., 2008). Restricting the spatial expression of UPRT in select tissues, 4tU is modified and subsequently incorporated into newly synthetized RNA only in cells that express UPRT. This way, RNA from cells expressing UPRT can

be recovered purifying labelled RNA from a total RNA extraction of the whole organism. This method could be used to isolate RNA from cell types that are difficult to separate by dissection or dissociation, such as neurons o glial cells in the central nervous system (Miller et al., 2009). Mammals lack this enzyme activity and thus only their protozoan parasite incorporates chemically modified uracil into its nascent RNAs. Very interestingly, transgene expression of UPRT under the control of a tissue-specific promoter could allow purification of RNA from specific cells or tissues in transgenic animals, either vertebrates or invertebrates, fed with the DTU. An important characteristic of **TU labelling** is that the percentage of labelled RNA over total RNA depends on labelling time. Short labelling times can be used to detect changes in genetic expression over time in specific cell types. The control of spatial expression of UPRT in vertebrates can be obtained utilizing transgenes, through retroviral activity, electroporation, or mRNA injection. The tissue-specific expression of UPRT combined with 4tU administration can be utilized to purify tissue-specific RNA from complex intact tissues in *Drosophila melanogaster*.

To evaluate the capacity to label RNA biosynthetically in *Drosophila*, 4tU has been supplied and its incorporation into RNA has been monitored by total RNA extraction, binding of the thiol nucleotide to biotin, and streptavidin to detect labelled RNA. Wild-type or UAS-UPRT adults and larvae fed upon 4tU showed very low levels of labelled RNA. In contrast, larvae and adults fed upon 4tU and containing both transgenes Gal4 and UAS-UPRT expressed UPRT in select cell types and showed high levels of labelled RNA. Similarly, embryos immerged in a medium containing 4tU showed high levels of labelling only when the transgenes UAS-UPRT and Gal4 were both present (Miller et al., 2009). In conclusion, the combination of UPRT and 4tU can be utilized to label RNA biosynthetically in embryos, larvae and adults of *Drosophila.* Once the newly transcribed RNAs are separated from non-labeled pre-existing mRNA, it is possible to conduct comparative analysis with microarray or sequencing of newly transcribed RNA, pre-existing RNA or total RNA.

Metabolic labelling with **4tU** (**4-thiouracile)** has been utilized to observe the alterations of synthesis and decay of RNA following physiologic and pathologic events, such as cellular response to treatment with type I and type II interferons (Platanias, 2005). The mRNA levels of specific genes in a precise instant are the result of a balance regulated by *de novo* transcription and transcript decay (Guhaniyogi and Brewer 2001; Fan et al. 2002; Jing et al. 2005). Microarray analysis on total RNA can provide a measuring of the abundance of mRNA, but cannot determine whether the changes over time are caused by alterations in synthesis or decay. A series of attempts have been conducted to elude this problem. Decay rates have been determined by blocking transcription (eg. using actinomycin D), assuming that decay rates remain unchanged (Frevel et al., 2003; Yang et al.,

2003; Bernstein et al., 2004; Raghavan and Bohjanen, 2004). However, this method is invasive for cells and cannot be combined with essays for the measuring of *de novo* transcription. The low temporal resolution is another limitation of genic expression profiling using total RNA. This is particularly true for mammalian cells, due to the long half-life of their mRNA (Yang et al. 2003).

As a very relevant alternative, metabolic labeling with **4-thiouridine (4sU)** doesn't need the expression of UPRT and hence it can be used for example directly into developing non transgenic embryos. The incorporation of 4sU into newly transcribed RNA has been studied cultivating different cytotypes in presence of concentrations of 4sU ranging from 100 mM to 5 mM for an hour. Following total RNA extraction and thiol-specific biotinylation, the RNA labeled with 4sU has been specifically detected and quantified through dot blot assay. The labeling of newly transcribed RNA with 4sU can be applied to a vast diversity of cell types and organisms, including human, mice and plants, as well as *Drosophila*. It can be both used for *in vitro* and *in vivo* studies, as it is well tolerated by mice following intravenous injection (Kenzelmann et al. 2007). Metabolic labeling of newly transcribed RNA with 4sU has little to no side effects on genetic expression, RNA decay, protein stability and viability of cells (Melvin et al. 1978; Woodford et al. 1988; Ussuf et al. 1995; Kenzelmann et al. 2007). The transcriptional profile of three biologic replicas of murine cells treated with 200 mM 4sU for an hour has been compared with that of untreated cells and no significant transcriptional level alteration attributable to treatment with 4sU has been detected (Dolken et al., 2008). The efficiency of isolation can be confirmed combining the 4sU labeling with the cytidine-^{3}H labelling of newly transcribed RNA for 15, 30 and 60 minutes. Following thiol-specific biotinylation, up to 90% of RNA labelled with cytidine-^{3}H has been co-purified with the newly transcribed RNA. (Dolken et al., 2008).

Another application of newly transcribed RNA labelling with **4sU** (4-thiouridine) was aimed at the isolation of the zygotic genes of *Danio rerio* (*zebrafish*) and the subsequent characterization of the zygotic transcripts (Heyn et al., 2014). The maternal to zygotic transition of the control over development is fundamental for the cell cycle of every multicellular organism. Zygotic genomes are transcriptionally inactive up until zygotic genomic activation (ZGA), which starts after a fixed number of cell divisions (Baroux et al., 2008; Tadros and Lipshitz, 2009). The early expressed genes most probably support the rapid cell divisions that precede morphogenesis, and take place into the mechanisms of sex determination. In *zebrafish*, transcription of early genes begins at the 64 cell stage (Giraldez et al., 2006; Lindeman et al., 2011; Mathavan et al., 2005; O'Boyle et al., 2007). Through 4-thiouridine metabolic labelling of newly transcribed RNA in zebrafish embryos, it has

been possible to isolate the early zygotic transcripts. The results showed little to no contamination from maternal RNA and an efficient recovery of newly transcribed RNA, verified on previously known zygotic genes (Heyn et al., 2014). Embryos injected with 4-thiouridine triphosphate (4-sUTP) at the single cell stage have been cultivated until 128, 256 and 512 cell stages prior to the extraction, biotinylation and purification of labelled RNA, which has been subsequently sequenced. These experiments showed that the pool of genes expressed solely in the zygote does not contain genes essential to life (Heyn et al., 2014). Moreover, no orthologues of *Drosophila* have been detected among zygotic genes in zebrafish, indicating that early arthropod and vertebrate zygotic genes are completely different and can serve to species-specific functions, in contrast with maternal genes, which are highly conserved and serve to essential functions (Heyn et al., 2014).

For my PhD project, I adapted the technique from Dolken et al. (2008), which was based on the injection of 4-thiouridine (4sU) in murine cells and also from Heyn et al., (2014). This allowed me to avoid the need of the UPRT enzyme, for the labelling of newly transcribed RNA.

2.2 Results

Ceratitis male sex determination of medfly occurs at 5-6 hours from oviposition in XY developing embryos, by a Y-linked M factor still to be identified (Gabrieli et al., 2010). While in *Drosophila* cellularization occurs at 2-3 hours, in *Ceratitis* it occurs after 9 hours. So, male sex determination occurs during mitotic divisions of nuclei and prior their migration to the periphery of the egg.

The Y chromosome of *Ceratitis* is highly heterochromatic and full of repetitive elements. Little is known about Y-linked genes and their early transcription in *Ceratitis* as, in general, in many other invertebrate and vertebrate species. However novel DNA sequencing technologies such as PacBio are offering novel opportunities to obtain useful sequence information from heterochromatic chromosomes full of repetitive sequences.

One of the aims of my PhD was the isolation of Y-linked genes of *Ceratitis* expressed during the first hours of embryogenesis. These Y-derived transcripts are expected to have 2 characteristics: 1) expressed only in the zygotes, as no maternal contribution is possible from Y-linked genes, and 2) obviously expressed as RNA in XY but not in XX embryos. In order to isolate them, I focused on the study of a novel molecular approach, called metabolic labelling, aimed to isolate and identify newly transcribed zygotic transcripts (after the fecundation and during the first stages of development) at the embryonic stages of *Ceratitis* and discriminate those present only in XY embryos. I expected to find within the male-specific zygotic transcripts, one or few potentially candidates corresponding to the M masculinizing factor. I focused on the 5-6 hours embryonic developmental window.

I have conducted bibliographical research on studies concerning metabolic labeling of newly synthetized RNA through the incorporation of a thiol nucleotide. Particularly, I have found two very interesting studies, in which the authors have used 4-thiouridine (4sU) in murine cells (Dolken et al., 2008) and 'zebrafish' embryos (Heyn et al., 2014). Following injections of 4sU, the newly transcribed RNAs have incorporated this modified nucleotide and they can be isolated from the maternal RNAs, by affinity chromatography. We then adapted this technique, applying it for the first time in embryos of a non-Drosophilidae species, such as *Ceratitis*.

As preliminary step, I injected 150 embryos (Benakeion strain; XX/XY mixed embryos from a normal cross between XY and XX flies) with 10 mM 4sU and 150 embryos with injection buffer as control (see table 1). I collected for 1 h the embryos and I let them develop for 5 hours, to cover a developmental window of 5-6 hours.

These data from the 2 experiments showed similar survival rate of larvae from injected embryos in both parallel experiments (48% and 51%, respectively). Also the 2 survival rates of adults were

similar (23% and 28%, respectively). I then concluded that 4sU is not particularly toxic for *Ceratitis capitata*.

I carried out another test to evaluate if the micromanipulations and the injection of the chemical 4sU into the embryos could alter transcriptional metabolism, for example slowing down the developmental rate, and related processes such as including the sex-specific splicing of *Cctra*. For example, it has been previously observed that metabolic labeling of RNA by 100 mM 4sU triggers a nucleolar stress response in cell lines (Burger et al., 2013).

In particular, the embryos manipulation and injection consisted of 1) manual dechorionation, 2) sticking with double-sided tape on slides, 3) addition of mineral oil to avoid dehydration, 4) microinjection with 4sU after some hours (at least 6) from oviposition, 5) washes with heptane to detach from the double-sided tape, thus from the slide, 6) washes with PBS 1X, 7) total RNA extraction.

For this reason, we verified if there was an eventual delay in the male-specific splicing of *Cctra* in pools of XX female only embryos and XX/XY mixed embryos as well as we verified if a known Y-linked transcribed pseudogene is also expressed in treated embryos. The special XX female only unisexual progeny was obtained by crossing XX males (which are fertile in medfly) with XX females, taking advantage of the *Ceratitis* transgenic line provoking masculinization of XX individuals by a maternal RNAi against *Cctra*. The fertile XX males have been obtained by a transgenic line in which maternally driven *in vivo* RNAi against the *Cctra* gene leads to depletion of maternal *Cctra* mRNAs in eggs from transgenic mothers and hence *Cctra* default male-specific splicing in XX embryos which develop as fertile males (see introduction).

The XX female embryos and XX/XY mixed embryos were treated in the following way leading to 10 different samples (5 for XX and 5 for XX/XY):

1) UE = Untreated Embryos: Pool of 50 embryos harvested after an hour and left in H_2O for 5 hours (H);
2) EG = Embryos + Glue; Pool of 50 embryos harvested after an hour and left on a slide with glue (G), in mineral oil, from which have been then eluted with heptane after 5 hours, and washed with PBS 1X (ND);
3) EGD = Embryos + Glue + dechorionation; Pool of 50 embryos harvested after an hour, dechorionated (D), and left on a slide with glue, in mineral oil, eluted from the slide with heptane after 5 hours, and washed with PBS 1X (D);

4) EGDM = Embryos + Glue + dechorionation + microinjection of buffer; Pool of 50 embryos harvested after an hour, dechorionated (D), and left on a slide with glue, in mineral oil, microinjected (M) with injection buffer, eluted from the slide with heptane after 5 hours, and washed with PBS 1X;
5) EGDM-4sU = Embryos + GLUE + dechorionation + microinjection of 4SU; Pool of 50 embryos harvested after an hour, dechorionated (D), and left on a slide with glue, in mineral oil, microinjected (M) with 4sU, eluted from the slide with heptane after 5 hours, and washed with PBS 1X.

Total RNA has been extracted from each of the 10 samples, with Trizol® and an RT-PCR has been conducted. I have found that sex-specific splicing pattern of the C*ctra* gene, was not affected by the treatments and male-specific splicing of *Cctra* as expected, started to appear in XX/XY embryos, but not in XX only embryos.

After carrying out these preliminary tests, I injected a 10 mM 4sU solution in embryos derived from the following 2 crosses, to produce respectively XX female only embryos and XX/XY mixed embryos.

Cross 1	**Cross 2**
♂XX + ♀XX	♂XY + ♀XX

2,783 XX (future XX females), and 2,617 XX/XY embryos have been injected over the course of 4-5 days. In total, the embryos were grouped in 9 samples, 5 samples for XX and 4 samples for XX/XY (table 2). The embryos were injected one hour after deposition (in a room at 18 °C) and have been left to develop for 5 hours, (in total 6 hours from deposition), have been detached from the slide and homogenized in Trizol® to conduct the following total RNA extraction. The 9 RNA samples (each sample: 70-130 micrograms of total RNA) were analyzed by gel electrophoresis (Fig.8) and by measuring quantity with Nanodrop® (table 3).

The 9 samples have been processed in parallel as following: the RNA has been subjected to thiol-specific biotinylation, with following purification using streptavidin coated magnetic beads (Cleary et al. 2005). Since eukaryotic RNA do not contain thiol groups, only newly synthetized zygotic RNA that incorporated 4sU are subject to biotinylation.

Our main concern was that the various invasive procedures (dechorionation, injections, removal of embryos with chemicals) could delay the development of the treated embryos. Male-specific splicing of *Cctra* was detected in all tested XX/XY samples, but absent in XX samples (Fig. 10);

similarly, I detected transcripts of the Y-linked *CcLap* pseudogene only in XX/XY but not in XX samples (Fig. 11), indicating that Y-linked transcriptional units are active in the treated embryos.

The 9 total RNA samples were incubated with a thio-reactive biotinylation reagent and streptavidin-magnetic beads were used to isolate 4sU-labeled RNAs by affinity chromatography. Each RNA labelled sample consisted of 0.300-0.200 micrograms extracted from a starting 70-130 micrograms total RNA.

From the eight samples I finally obtained by pooling, 4 samples, named as zygo: XX-only zygo embryos samples in 2 replicas (A and B); XY/XX mixed zygo embryos in 2 replicas (C and D). The XX sample n. 5 has been used only as a control experiment to evaluate the presence of nuclease activity during the procedure (for a schematic representation, see Fig.12).

The 4 zygo RNA samples has been sequenced using Illumina-HiSeq® technology at the C-CAMP institute (in collaboration with Dr. Arun Kumar K. P, head of Laboratory of Molecular Genetics Centre for DNA Fingerprinting & Diagnostics, Nampally, Hyderabad, India). 66 million paired-end 100 nucleotides-long reads of XY/XX sample and 60 million reads of XX-only sample were produced. The data obtained from the sequencing have been processed *in silico* (in collaboration with Dr. Marco Salvemini) to produce 3 in silico assembled zygo transcriptomes: XX-only and XX/XY corresponding zygo transcriptomes and a third zygo assembly using all reads from XX and XX/XY. The assembly of reads for each sample was done, using Trinity® software (Haas et al., 2013; Petrella et al., 2015), and resulted in 243,943 transcripts and 239,914 transcripts from XY/XX and XX-only samples, respectively. The unique zygo transcriptome resulted into 326,178 transcripts. The large number of transcripts suggested that the purification of only zygotically expressed genes was ineffective, as we expected a much lower transcript number and much less complexity. Indeed, we expected to assemble few hundreds of genes based on the observation of Heyn et al., in zebrafish (2014). However, we asked if an enrichment of zygotically expressed genes was present in the zygo samples. A comparative differential expression analysis was performed by Salvemini using my zygo samples as well as other mRNAseq unpublished data from XX and XX/XY embryos at different time points (4-8 hours; 8-10 hours; 22-24 hours). The 5-6 hours zygo samples were clearly enriched for hundreds of transcripts, which are currently under additional bio-informatic analysis. Within the first 10 most expressed zygo transcripts we have found for example the *Ceratitis* orthologue of the *Drosophila* zyvgotic gene *slam* which is expressed in all somatic cells of the blastoderm embryo, but is excluded from germline cells (Acharya et al., 2014).

In *Drosophila*, *slam* transcript is first detected at embryonic cycle 13 and remains detectable through the slow phase of cellularization and fades away during the fast phase of cellularization.

A new RNA sequencing using Illumina-HiSeq® on unfertilized eggs to conduct an *in silico* subtraction from the zygotic transcriptome is currently also in progress to further enrich the zygo samples with transcripts exclusively produced during first hours of Ceratits embryogenesis. Furthermore, we expect that the Y-linked M factor transcript should be absent also in the unfertilized eggs.

A differential expression analysis of my zygo samples comparing XY/XX versus XX by was performed in collaboration with Mark Robinson (UNIZH) and 200 putative male-specific transcripts were identified using EdgeR analysis (Robinson et al., 2010).

I took advantage of transcriptomics and genomics medfly database containing other published and unpublished data which is accessible in our lab (Salvemini, unpub. res) and I performed an extensive BLASTn analysis to filter and select the best male-specific candidates from the 200 ones. A first idea to filter them, was to select those transcripts present in the "zygo" and other XX/XY embryonic transcriptomes but completely absent in XX ones; a second idea was to use novel PacBio genome sequencing data on XY individuals, performed in collaboration with Prof. Mark Robinson on a special medfly strain, FAM18, which bears a shorter Y chromosome, which is still able to determine maleness (Saccone, Robinson, Salvemini, pub. Res). I used the putative male-specific transcripts by BLASTn to search for the presence of putative Y-linked genes in PacBio genomic reads and in a first Canu Assembly (Robinson, M. and Schmeing, S.; unpub. Res.), both available in our lab restricted database. I have also searched by FlyBLASTx protein homology in *Drosophila* database.

However, I performed a "one by one" BLAST analysis of the first 100 transcripts to find those present also in other embryonic transcriptomes at later stages of development (8-10 hours; 0-48 hours) and possibly also in adult flies transcriptomes. I have confirmed by BLAST that only 6 of them are exclusively present in XX/XY transcriptomes and hence absent in XX-only transcriptomes. In particular, 2 transcripts (*CczygoY1 CcYzygo2*) showed by BLASTx in Flybase some scattered and weak similarity to *H2B* histone and to *zero population growth* proteins, respectively. Interestingly, these transcripts seem to correspond to truncated Y-linked paralogous sequences derived by partial duplication and mobilization from *Ceratitis* orthologues, localized on autosomes. However, it is presently unclear whether they can play any function, and functional studies are required and are underway to investigate the problem.

I have also used *CczygoY1 CczygoY2* sequences in BLASTn to search the published medfly genome (Baylor college Genome project) and I have found no corresponding scores. This is expected for genomic sequences derived from highly repetitive regions such as the Y chromosome, which are very difficult to be assembled *in silico* from short Illumina reads. The novel PacBio DNA

sequencing technology helps to overcome this problem, because it can lead to long single DNA molecules reads up to 20 Kb and then use such long reads for a more accurate assembly especially for highly repetitive regions. A special strain of medfly, called Fam18 (bearing a shorter Y chromosome), has been used to extract genomic DNA and perform both PacBio (only from males) and Illumina sequencing (from males and females) (Mark Robinson, Saccone, G., and Salvemini, M., UNIZH, unpub. Data). I prepared genomic DNA from the FAM18 strain (2 samples: from males and from females), while I was in the Bopp lab in 2015, and the samples were used for Illumina sequencing and PacBio. The PacBio sequencing technology has a limitation of 10-15% DNA sequencing error rate; however, the introduced errors are usually insertions or deletions o single nucleotides which can be corrected in various ways. Our collaborator Mark Robinson (UNIZH) has used Illumina reads from males and from females to perform a correction of the PacBio reads. Hence, I have used a local database to search by BLASTn the database including the "corrected" PacBio reads, using *CczygoY1* and *CcYzygo2* sequences and I have found various corresponding sequences, confirming their presence in the FAM18 strain. Hence the 2 genes are present on the deleted Y chromosome, which contains also a functional M factor.

I am presently developing a Cas9 and an embryonic RNAi approaches to target the 2 Y-linked genes and evaluate by RT-PCR on injected XX/XY embryos a change in the *Cctra* sex-specific splicing pattern. We expect to observed in case of a reduction of the male determining M activity caused by RNAi in XY, a concomitant reduction of the male-specific spliced products in favour to the female-specific ones.

4sU 10 mM	Buffer	Embryos	Larvae/embryos	Pupae/larvae	Adults/pupae	Adults/embryos
0	+	150	51% (76/150)	59% (45/76)	95% 43/46	28% (43/150)
+	0	150	48% (72/150)	50% (36/72)	97% (35/36)	23% (35/150)

Table 1: Survival rates of the embryos injected with 4sU 10 mM and embryos injected with buffer only.

Samples	Microinjected embryos XX
1	487
2	442
3	574
4	565
5	715
Total	2,783

Samples	Microinjected embryos XX/XY
6	602
7	712
8	624
9	679
Total	2,617

Table 2: Embryos injected with 4sU.

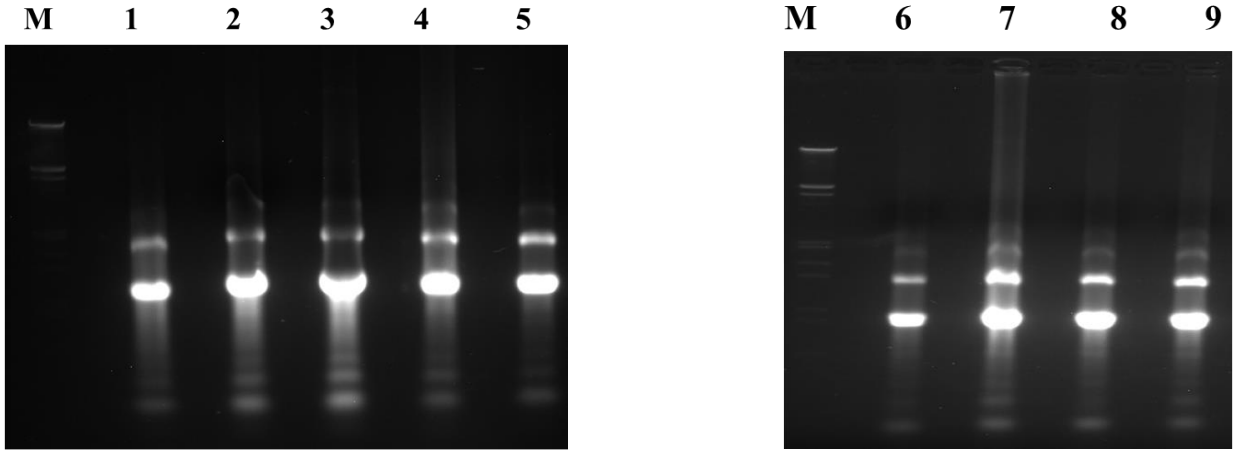


Fig. 9 total RNA extraction from the 9 samples. Samples 1-5 are XX only embryos; samples 6-9 are XY/XX embryos; M is the Lambda DNA/EcoRI plus HindIII Marker (*Thermo Fisher Scientific*®).

Samples	ng/µL	260/280	260/230	Total RNA in V_F
1	3863	2.26	1.70	96 µg in 25 µL
2	4729.7	2.23	1.90	118 µg in 25 µL
3	6323.1	2.24	1.90	128 µg in 25 µL
4	3557.2	2.22	1.62	89 µg in 25 µL
5	3523.8	2.29	1.90	88 µg in 25 µL
6	2281.6	2.20	1.76	57 µg in 25 µL
7	4385.8	1.90	1.19	109 µg in 25 µL
8	3457.7	2.25	2.18	86 µg in 25 µL
9	4261.9	2.21	2.02	106 µg in 25 µL

Table 3: Spectrophotometrical analysis of the 9 RNA samples conducted with Nanodrop® 2000 (Thermo Scientific).

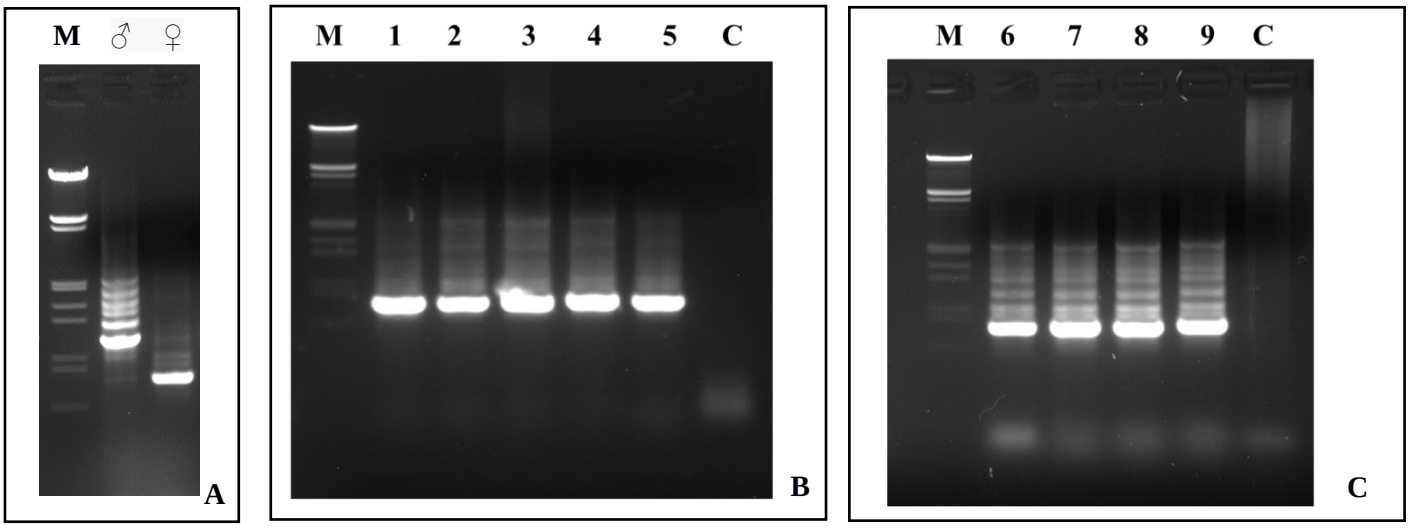


Fig. 10: Gel electrophoresis analysis of the RT-PCR of the *Transformer* gene. A: adult flies; B: XX only embryos; C: XY/XX embryos.

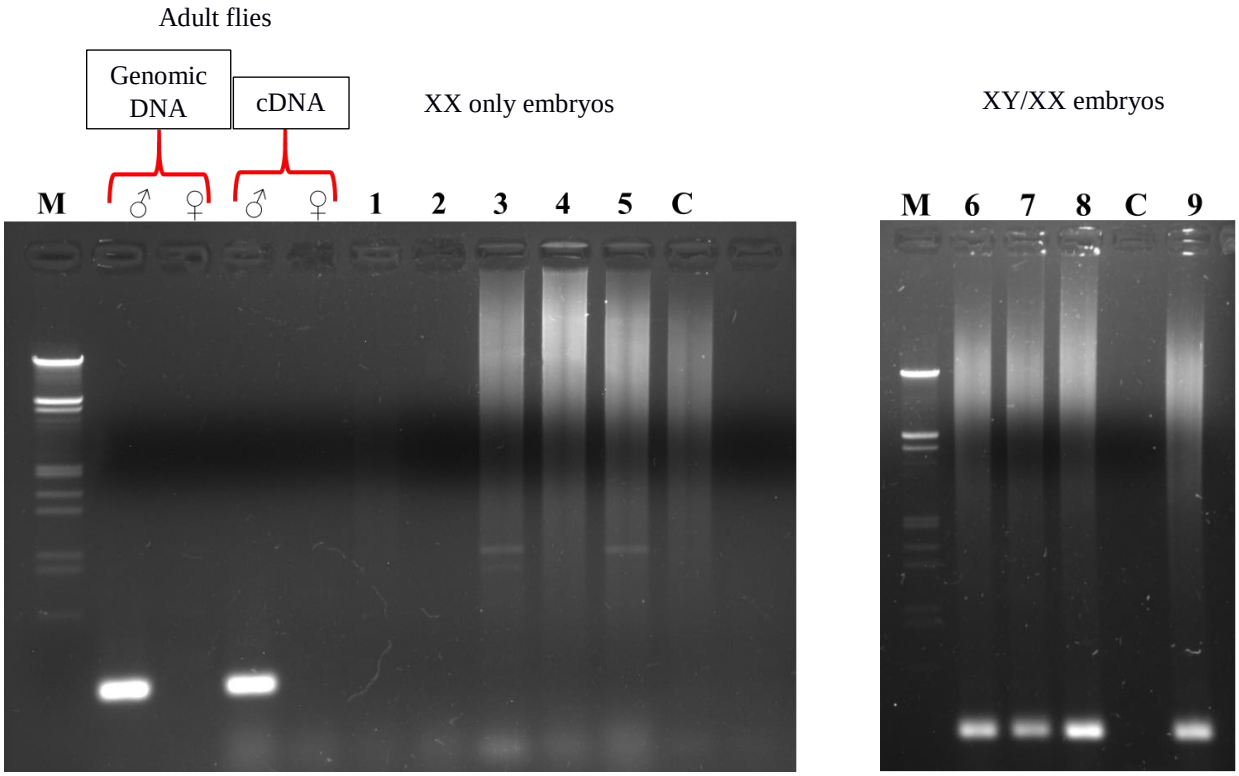


Fig. 11: RT-PCR of *CcLap* Y-linked pseudogene.

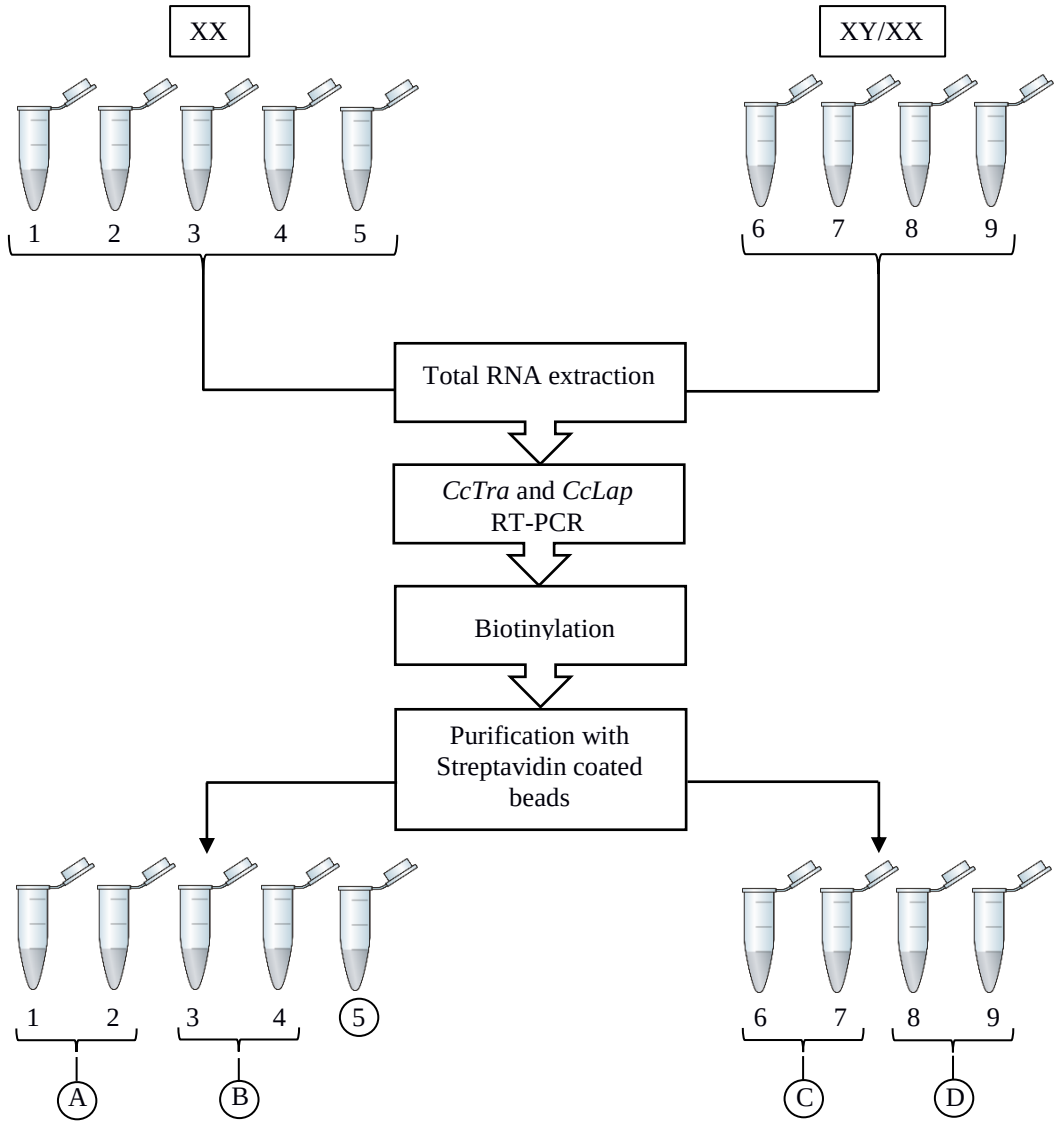


Fig. 12: Schematic of the treatments carried out on the 9 embryos samples. XX only samples 1-2 and 3-4 were grouped into the biological replicates A and B, respectively. XY/XX samples 6-7 and 8-9 were grouped into the biological replicates C and D, respectively. Sample 5 was used as control.

2.3 Method

Rearing of the *C. capitata* strain

The *C. capitata Benakeion* strain are reared in standard laboratory conditions at 25°C and 12:12 h light-dark regimen. Adult flies are fed yeast/sucrose powder (1:2). Eggs are collected in water dishes and transferred to larval food (30 g soft tissue paper, 30 g sugar, 30 g yeast extract, 10 ml cholesterol stock, 2 ml HCl stock, 8.5 ml benzoic stock, water 400 ml). Pupae are collected and stored in Petri dishes until eclosion.

Microinjection of 4sU into *Ceratitis capitata* embryos

Embryos were collected 1 hour after egg laying (AEL), hand dechorionated and injected with 10 mM TriLink® 4-Thio-UTP. In parallel, I also injected a pool of embryos with injection buffer 1X (5 mM KCl, 0,1 mM PO_4) as control. The microinjected embryos were left to develop at room temperature for 5 hours before being detached from the slide with 750 µL Heptane PESTANAL® (Fluka). After the detachment, the embryos were washed five times with PBS 1X before being homogenized in Trizol® (Invitrogen).

RNA and genomic DNA extraction

Total RNA from embryos and single adult of the *C. capitata* was extracted using Trizol reagent (Invitrogen) following the modified protocol described by Chomczynski and Mackey (1995). Genomic DNA extraction from adult of *C. capitata* (male and female) was performed with "Holmes-Bonner" buffer according to Maniatis et al. (1982).

Reverse Transcription PCR

RT-PCR was performed using the EuroScript M-MLV Reverse Transcriptase (Euroclone, Pero, IT). Starting from a maximum concentration of 4 ug of RNA from each sample, add 1 µl of oligo (dT), incubate the sample at 65°C for 5 min. then the sample is mixed with 5X Reaction Buffer, dNTP mix, RNase inhibitor, MMLV reverse transcriptase in the total volume of 20 µl. The mixture was incubated in a thermal cycle at 42 °C for 1h and 70 °C for 10 min.

A PCR was conducted the primers Sod+/Sod- as positive control. The following PCR cycles were performed: 5 min at 94 °C, 35 cycles with 30 secs at 94 °C, 30 secs at 60 °C, 45 secs at 72 °C, 10 min at 72°C. RT-PCR products were analysed by 1% agarose gel electrophoresis. *Cc*SOD-TGCTCCGAGAACGTTCACG; *Cc*SOD- TCATCGGTCAATTCGTGCAC.

The RT-PCR expression analysis on *Cclap-ps* transcripts was performed with the Y2 primer pairs (Y2+: 5'-AAGGACTTGTGATTGGATTG-3'; Y2-: 5'-ATGCCGTCGTCCAACATC-3'). The RT-

PCR analysis of the *Cctra* clone was performed with the following primers: *Cctra* 180+ (located in *Cctra* exon 1), 5'- ATAGAGCGCAGTGTCAATCC- and 3'; *Cctra* 920- (located in *Cctra* exon 2) 5' -TCCTGTTCTTCCGATCTGTG- 3'.

Biotinylation and purification of 4sU-labeled RNA

I adapted the biotinylation and purification of 4sU-labeled RNA techniques from Dölken et al., 2008. For the biotinylation of 4sU-labeled RNA EZ-Link Biotin-HPDP (Pierce®) was dissolved in dimethylformamide (DMF) at a concentration of 1 mg/mL and stored at 4°C. Biotinylation was carried out in 10 mM Tris (pH 7.4), 1 mM EDTA, and 0.2 mg/mL Biotin-HPDP at a final RNA concentration of 100 ng/µL for 1.5 h at room temperature. A quantity of total RNA ranging from 70 µg to 130 µg was used for the biotinylation reaction. Unbound Biotin-HPDP was removed by chloroform/isoamylalcohol (24:1) extraction using Phase-lock-gel (Heavy®) tubes (Eppendorf®). A 1/10 volume of 5 M NaCl and an equal volume of isopropanol were added and RNA was precipitated at 20,000 g for 20 min. The pellet was washed with an equal volume of 75% ethanol and precipitated again at 20,000 g for 10 min. The pellet was resuspended in 20 µL RNase-free water. After denaturation of RNA samples at 65° C for 10 min followed by rapid cooling on ice for 5 min, biotinylated RNA was captured using µMACS streptavidin beads and columns (Miltenyi®). Up to 20 µg of biotinylated RNA were incubated with 20 µL of µMACS streptavidin beads with rotation for 15 min at room temperature. The beads were transferred and magnetically fixed to the columns. Columns were washed three times with 1 mL 65° C washing buffer (100 mM Tris-HCl, pH 7.4, 10 mM EDTA, 1 M NaCl, 0.1% Tween20) followed by five washes with room temperature washing buffer. Labelled RNA was eluted by the addition of 100 µL of freshly prepared 100 mM dithiothreitol (DTT) followed by a second elution round 5 min later. RNA was recovered from the washing fractions and eluates using the RNeasy MinElute Spin columns (Qiagen®) following its protocol.

2.4 References

- Acharya S, Laupsien P, Wenzl C, Yan S, Großhans J. Function and dynamics of slam in furrow formation in early *Drosophila* embryo. Dev Biol. 2014 Feb 15;386(2):371-84. doi: 10.1016/j.ydbio.2013.12.022. Epub 2013 Dec 22.
- Anleitner JE, Haymer DS. Chromosoma. Y enriched and Y specific DNA sequences from the genome of the Mediterranean fruit fly, *Ceratitis capitata*. 1992 Mar;101(5-6):271-8.
- Baroux, C., Autran, D., Gillmor, C.S., Grimanelli, D., and Grossniklaus, U. (2008). The maternal to zygotic transition in animals and plants. Cold Spring Harb. Symp. Quant. Biol. 73, 89–100.
- Bell, L.R., Maine, E.M., Schedl, P. and Cline, T.W. (1988). *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similar to RNA binding proteins. Cell 55: 1037–1046.
- Belote, J.M., Mckeown, M., Boggs, R.T., Ohkawa, R., and Sosnowski, B.A. (1989). The molecular genetics of *transformer*, a genetic switch-controlling sexual differentiation in *Drosophila.* Devel. Genet. 10: 143-154.
- Bernstein, J.A., Lin, P.H., Cohen, S.N., and Lin-Chao, S. 2004. Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. Proc. Natl. Acad. Sci. 101: 2758–2763.
- Boggs, R.T., Gregor, P., Idriss, S., Belote, J.M. and Mckeown, M. (1987). Regulation of sexual differntiation in *Drosophila melanogaster* via alternative splicing of RNA from the *transformer* gene. Cell 50: 739–747.
- Bopp, D., Bell, L.R., Cline, T.W. and Schedl, P. (1991). Developmental distribution of female-specific *Sex-lethal* proteins in *Drosophila melanogaster*. Genes Devel. 5: 403–415.
- Burger K, Mühl B, Kellner M, Rohrmoser M, Gruber-Eber A, Windhager L, Friedel CC, Dölken L, Eick D. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol. 2013 Oct;10(10):1623-30. doi: 10.4161/rna.26214. Epub 2013 Sep 4.
- Burtis, K.C. and Baker, B.S.,1989. *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell 56: 997-1010.
- Chomczynski P1, Mackey K. Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. Biotechniques. 1995 Dec;19(6):942-5.

- Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R., and Boothroyd, J.C. 2005. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat. Biotechnol. 23: 232–237.
- Cline, T.W. (1984). Autoregulation functioning of a *Drosophila* gene product that establishes and maintains the sexually determined state. Genetics 107:231-277.
- Dolken L., Ruzsics Z., Radle B. 2008. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA. 14(9):1959-72.
- Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML, Gong S, Greengard P, Heintz N. Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell. 2008 Nov 14;135(4):749-62. doi: 10.1016/j.cell.2008.10.029.
- Erickson JW, Quintero JJ. Indirect effects of ploidy suggest X chromosome dose, not the X:A ratio, signals sex in Drosophila. PLoS Biol. 2007 Dec;5(12):e332.
- Fan, J., Yang, X., Wang, W., Wood III., W.H., Becker, K.G., and Gorospe, M. 2002. Global analysis of stress-regulated mRNA turnover by using cDNA arrays. Proc. Natl. Acad. Sci. 99: 10611–10616.
- Frevel, M.A., Bakheet, T., Silva, A.M., Hissong, J.G., Khabar, K.S., and Williams, B.R. 2003. p38 Mitogen-activated protein kinase-dependent and -independent signaling of mRNA stability of AU-rich element-containing transcripts. Mol. Cell. Biol. 23: 425–436.
- Furia M, Giordano E, Digilio A, Polito LC. A cDNA clone representative of a novel splicing pattern of the *D. melanogaster dunce* gene. Genetica. 1992;87(3):159-63.
- Gabrieli P, Falaguerra A, Siciliano P, Gomulski LM, Scolari F, Zacharopoulou A, Franz G, Malacrida AR, Gasperi G. Sex and the single embryo: early deveiopment in the Mediterranean fruit fly, *Ceratitis capitata*. BMC Dev Biol. 2010 Jan 26;10:12. doi: 10.1186/1471-213X-10-12.
- Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science 312, 75–79.
- Guhaniyogi, J. and Brewer, G. 2001. Regulation of mRNA stability in mammalian cells. Gene 265: 11–23
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for

reference generation and analysis. Nat Protoc. 2013 Aug;8(8):1494-512. doi: 10.1038/nprot.2013.084. Epub 2013 Jul 11.

- Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, Suárez-Fariñas M, Schwarz C, Stephan DA, Surmeier DJ, Greengard P, Heintz N. A translational profiling approach for the molecular characterization of CNS cell types. Cell. 2008 Nov 14;135(4):738-48. doi: 10.1016/j.cell.2008.10.028.
- Heyn P., Kircher M., Dahl A., Kelso J., Tomancak P., Kalinka A.T., Neugebauer K.M. 2014. The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species. Cell Reports.
- Heinze S.D., Kohlbrenner T, Ippolito D.,Meccariello A, Burger A,Mosimann C, Saccone G, and Daniel Bopp. (2017). CRISPR-Cas9 targeted disruption of the yellow ortholog in the housefly identifies the brown body locus. (IN SUBMISSION)
- Jiménez-Guri E, Huerta-Cepas J, Cozzuto L, Wotton KR, Kang H, Himmelbauer H, Roma G, Gabaldón T, Jaeger J. Comparative transcriptomics of early dipteran development. BMC Genomics. 2013 Feb 24;14:123. doi: 10.1186/1471-2164-14-123.
- Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S.C., Gram, H., and Han, J. 2005. Involvement of microRNA in AU-rich element-mediated mRNA instability. Cell 120: 623–634.
- Kenzelmann, M., Maertens, S., Hergenhahn, M., Kueffer, S., Hotz-Wagenblatt, A., Li, L., Wang, S., Ittrich, C., Lemberger, T., Arribas, R., et al. 2007. Microarray analysis of newly synthesized RNA in cells and animals. Proc. Natl. Acad. Sci. 104: 6164–6169.
- Lindeman, L.C., Andersen, I.S., Reiner, A.H., Li, N., Aanes, H., Østrup, O., Winata, C., Mathavan, S., Mu¨ ller, F., Alestro¨ m, P., and Collas, P. (2011). Prepatterning of developmental gene expression by modified histones before zygotic genome activation. Dev. Cell 21, 993–1004.
- Maniatis, T. Sambrook, J. Fritsch, E.F. Molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y. Cold Spring Harbor Laboratory, 1982
- María Fernanda Ruiz, Mercedes Alvarez, José M. Eirín-López, Francesca Sarno, Leonor Kremer, José L. Barbero, and Lucas Sánchez. An Unusual Role for *doublesex* in Sex Determination in the Dipteran *Sciara* Genetics. 2015 Aug; 200(4): 1181–1199. Published online 2015 Jun 10. doi: 10.1534/genetics.115.177972. PMCID: PMC4574253.
- Mathavan, S., Lee, S.G.P., Mak, A., Miller, L.D., Murthy, K.R.K., Govindarajan, K.R., Tong, Y., Wu, Y.L., Lam, S.H., Yang, H., et al. (2005). Transcriptome analysis of zebrafish embryogenesis using microarrays. PLoS Genet. 1, 260–276.

- Meccariello A, Monti S, Romanelli A, Colonna R, Primo P, Inghilterra MG, Del Corsano G, Heinze S.D.,Salvemini M, Robinson M, Chiavacci E, Mosimann C, Daniel Bopp and Giuseppe Saccone (2017). Efficient DNA-free gene targeting of the major agricultural pest Ceratitis capitata using CRISPR/Cas9 ribonucleoprotein complexes. (IN SUBMISSION)
- Melvin, W.T., Milne, H.B., Slater, A.A., Allen, H.J., and Keir, H.M. 1978. Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. Eur. J. Biochem. 92: 373–379.
- Miller, M.R., Robinson, K.J., Cleary, M.D., Doe, C.Q. (2009). TU-tagging: cell type-specific RNA isolation from intact complex tissues. Nat. Methods 6(6): 439--441.
- O'Boyle, S., Bree, R.T., McLoughlin, S., Grealy, M., and Byrnes, L. (2007). Identification of zygotic genes expressed at the midblastula transition in zebrafish. Biochem. Biophys. Res. Commun. 358, 462–468.
- Pane,A., Salvemini, M., Delli Bovi, P., Polito, L.C. And Saccone,G.. 2002. The transformer gene in Ceratitis capitata provides a genetic basis for selecting and remembering the sexual fate. Development 129:3715-3725.
- Alexie Papanicolaou, Marc F. Schetelig, Peter Arensburger, Peter W. Atkinson, Joshua B. Benoit, Kostas Bourtzis, Pedro Castañera, John P. Cavanaugh, Hsu Chao, Christopher Childers, Ingrid Curril, Huyen Dinh, HarshaVardhan Doddapaneni, Amanda Dolan, Shannon Dugan, Markus Friedrich, Giuliano Gasperi, Scott Geib, Georgios Georgakilas, Richard A. Gibbs, Sarah D. Giers, Ludvik M. Gomulski, Miguel González-Guzmán, Ana Guillem-Amat, Yi Han, Artemis G. Hatzigeorgiou, Pedro Hernández-Crespo, Daniel S. T. Hughes, Jeffery W. Jones, Dimitra Karagkouni, Panagiota Koskinioti, Sandra L. Lee, Anna R. Malacrida, Mosè Manni, Kostas Mathiopoulos, Angela Meccariello, Shwetha C. Murali, Terence D. Murphy, Donna M. Muzny, Georg Oberhofer, Félix Ortego, Maria D. Paraskevopoulou, Monica Poelchau, Jiaxin Qu, Martin Reczko, Hugh M. Robertson, Andrew J. Rosendale, Andrew E. Rosselot, Giuseppe Saccone, Marco Salvemini, Grazia Savini, Patrick Schreiner, Francesca Scolari, Paolo Siciliano, Sheina B. Sim, George Tsiamis, Enric Ureña, Ioannis S. Vlachos, John H. Werren, Ernst A. Wimmer, Kim C. Worley, Antigone Zacharopoulou, Stephen Richards, Alfred M. Handler. The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. Genome Biol. 2016; 17: 192. Published online 2016 Sep 22. doi: 10.1186/s13059-016-1049-2.
- Petrella V, PhD THESIS(2014) : HOW TO PRODUCE MALE-ONLY PROGENY IN PEST INSECTS FOR SIT: A BIOTECH APPROACH.

- Petrella V, Aceto S, Musacchia F, Colonna V, Robinson M, Benes V, Cicotti G, Bongiorno G, Gradoni L, Volf P, Salvemini M. (2015). De novo assembly and sex-specific transcriptome profiling in the sand fly *Phlebotomus perniciosus* (Diptera, Phlebotominae), a major Old World vector of *Leishmania infantum. BMC Genomics*. Oct 23;16:847. doi: 10.1186/s12864-015-2088-x.
- Platanias, L.C. 2005. Mechanisms of type-I- and type-II-interferonmediated signalling. Nat. Rev. Immunol. 5: 375–386.
- Raghavan, A. and Bohjanen, P.R. 2004. Microarray-based analyses of mRNA decay in the regulation of mammalian gene expression. Brief. Funct. Genomic. Proteomic. 3: 112–124.
- Rideout, E.J., Billeter, J.C. and Goodwin, S.F. (2007). The sex-determination gene *fruitless* and *doublesex* specify a neural substrate required for courtship song. Current Biology 17: 1473-1478.
- Robinson, A.S., Riva, M.E. & Zapater, M. Theoret. Appl. Genetics (1986) 72: 455. doi:10.1007/BF00289526
- Robinson, A.S., G. Franz & K. Fisher, 1999. Genetic sexing strains in the medfly, *Ceratitis capitata*: development, mass rearing and field application. Trends Entomol. 2: 81–104
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139-40. doi: 10.1093/bioinformatics/btp616. Epub 2009 Nov 11.
- Roy, P.J., Stuart, J.M., Lund, J., Kim, S.K. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. Nature, 2002 Aug 29;418(6901):975-9.
- Saccone, G, M Salvemini, and LC Polito (2011).. "The *transformer* gene of *Ceratitis capitata*: a paradigm for a conserved epigenetic master regulator of sex determination in insects." Genetica 139.1 : 99-111.
- Saccone, G., Pane,A., Simone,A., Salvemini,M., Milano,A., Annunziata,L., Mauro,U., Polito,L.C. 2007. New sexing strains for Mediterranean fruit fly Ceratitis capitata: transforming females into males. In: Vreysen MJB, Robinson AS, Hendrichs J (eds). Area-wide control of insect pests, pp 95–102.
- Saccone,G. 1997. L'omologo del gene double-sex di *Drosophila melanogaster* in *Ceratitis capitata*: evidenza di una parziale conservazione evolutiva nei due ditteri di una gerarchia di regolazione genica nel differenziamento sessuale. PfD. Thesis, Università degli studi di Napoli, Federico II.
- Saccone,G. And Polito,L.C. 2002. Sex determination, flies, fruit flies and butterflies. Genetica 116: 15-23.

- Saccone,G., Peluso,I., Artiaco,D., Giordano,E., Boop,D. And Polito,L.-1998. The Ceratitis capitata omologo of the Drosophila sex determinig gene sex-lethal is structurally conserved, but non sex specifically regulated. Devolopment 125:1495-1500.
- Saccone,G., Peluso,L., Testa, G., Di Paola, F., Pane,A. And Polito, L.C.. 1996. Drosophila sex-lethal and double-sex homologous genes in Ceratitis capitata sercing for sex-specific genes to develop a Medfly transgenic sexing strain, in Enhancement of the sterile insect technique trout Genetic Transformetion using nuclear techniques IEA/FAO,Vienna
- Saccone,G., Salvemini,M., Polito,L.C., 2010. The transformer gene of Ceratitis capitata: a paradigm for a conserved epigenetic master regulator of sex determination in insects. Genetica. Oct.2 [Epub ahead of print
- Salvemini M, Polito C, Saccone G.-2010- Fruitless alternative splicing and sex behaviour in insects: an ancient and unforgettable love story? J Genet. 2010 Sep; 89(3): 287-99.
- Salvemini, M.; U. Mauro, S. Velaeti, C. Polito And G. Saccone. (2006) - A new Minos vector for eye-specific expression of white+ marker in Ceratitis capitata and in distantly related dipteran species. Insect Molecular Biology.15(3), 341–349
- Salvemini,M., Robertson,M., Arison,B., Atkinson,P., Polito,L.C. And Saccone,G.. 2009. Ceratitis capitata transformer-2 gene is required to establish and maintain the autoregulation of Cctra, the master gene for female sex determination. Int.j.Dev.Biol,53:109-120.
- Salvemini M, D'Amato R, Petrella V, Ippolito D, Ventre G, Zhang Y, Saccone G. (2014). Subtractive and differential hybridization molecular analyses of Ceratitis capitata XX/XY versus XX embryos to search for male-specific early transcribed genes. *BMC Genet*;15 Suppl 2:S5. doi: 10.1186/1471-2156-15-S2-S5. Epub 2014 Dec 1
- Salz HK, Maine EM, Keyes LN, Samuels ME, Cline TW, Schedl P. (1989). The *Drosophila* female specific sex determination gene, *Sex-lethal*, has stage, tissue, and sexspecific RNAs suggesting multiple modes of regulation. Genes Dev. 3(5):70819
- Sanchez L. (2008). Sex-determining mechanisms in insects. *Int. J. Dev. Biol*. 52: 837-856 (2008)
- Tadros W., Lipshitz H.D. 2009. The maternal-to-zygotic transition: a play in two acts. Development. 136(18):3033-42. doi: 10.1242/dev.033183.
- Tanke HJ, van der Keur M. Selection of defined cell types by flow-cytometric cell sorting. Trends Biotechnol. 1993 Feb;11(2):55-62. Review.
- Tian, M. and Maniatis,T. (1993) A splicing enhancer control alternative splicing of *doublesex* pre-mRNA. Cell 16:105-114

- Traut W., Sahara K., Marec F. (2007). Sex chromosomes and sex determination in Lepidoptera. Sex Dev. 1(6):332-46. doi: 10.1159/000111765
- Ussuf, K.K., Anikumar, G., and Nair, P.M. 1995. Newly synthesised mRNA as a probe for identification of wound responsive genes from potatoes. Indian J. Biochem. Biophys. 32: 78–83
- Valcárcel, J., Singh, R., Zamore, P.D. and Green, M.R. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of *transformer* pre-mRNA. Nature 362: 171–175.
- Wiedemann (1824).
- Willhoeft U, Franz G. Identification of the sex-determining region of the Ceratitis capitata Y chromosome by deletion mapping. Genetics. 1996 Oct;144(2):737-45.
- Woodford, T.A., Schlegel, R., and Pardee, A.B. 1988. Selective isolation of newly synthesized mammalian mRNA after *in vivo* labeling with 4-thiouridine or 6-thioguanosine. Anal. Biochem. 171: 166–172.
- Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., and Darnell Jr., J.E. 2003. Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes. Genome Res. 13: 1863–1872.

3 Technique CRISPR-CAS9

3.1 Highly efficient DNA-free gene disruption in the agricultural pest *Ceratitis capitata* by CRISPR-Cas9 RNPs. (paper in submission)

3.2 CRISPR-Cas9 targeted disruption of the yellow ortholog in the housefly identifies the brown body locus. (paper in submission)

Highly efficient DNA-free gene disruption in the agricultural pest *Ceratitis capitata* by CRISPR-Cas9 RNPs

Angela Meccariello[1], Simona Maria Monti[2], Alessandra Romanelli[3], Rita Colonna[1], Pasquale Primo[1], Maria Grazia Inghilterra[1], Giuseppe Del Corsano[1], Antonio Ramaglia[4], Giovanni Iazzetti[1], Francesco Patti[5], Svenia D. Heinze[6], Marco Salvemini[1], Helen Lindsay[6,7], Elena Chiavacci[6], Mark D. Robinson[6,7], Christian Mosimann[6], Daniel Bopp[6] and Giuseppe Saccone[1]

[1]Department of Biology, University of Naples Federico II, 80126, Napoli, Italy

[2]Institute of Biostructure and Bioimaging (IBB-CNR), Naples, Italy

[3]Department of Pharmacy, University of Naples Federico II, 80134, Napoli, Italy

[4]Department of Physics "E. Pancini", University of Naples Federico II, 80126, Napoli, Italy

[5]Stazione Zoologica Anton Dohrn, Center Villa Dohrn for Benthic Ecology, Punta San Pietro, 80077, Ischia, Italy.

[6]Institute of Molecular Life Sciences, University of Zürich, Zürich 8057, Switzerland

[7]SIB Swiss Institute of Bioinformatics, University of Zürich, Zürich 8057, Switzerland

Author for correspondence: Giuseppe Saccone, Via Cinthia, n. 26, 80126, Napoli, Italy, giuseppe.saccone@unina.it +39-0812535187

29

30 running head: efficient Cas9 gene disruption in the medfly

31 Key words: Cas9, *white eye*, *paired*, *Ceratitis capitata*, *Tephritidae*, *agricultural pest*,

32 *genetics, gene editing,*

33

ABSTRACT

The Mediterranean fruitfly *Ceratitis capitata (medfly)* is an invasive agricultural pest of high economical impact and has become an emerging model for developing new genetic control strategies as alternative to insecticides. Here, we report the successful adaptation of CRISPR-Cas9-based gene disruption in the *medfly* by injecting *in vitro* pre-assembled Cas9 ribonucleoparticles (RNP) loaded with gene-specific sgRNAs into early embryos. When targeting the eye pigmentation *white eye (we),* a high rate of somatic mosaicismwas observed in surviving G0 adults. Germline transmission of mutated *we* alleles was found in the progeny of more than 70% of G0 flies. Large deletions were recovered in the *we* gene when two sites were simultaneously targeted by duplex sgRNAs. CRISPR-Cas9 targeting of the Ceratitis ortholog of the Drosophila segmentation *paired* gene (*Ccprd*) gene caused segmental malformations in late embryos and irregular movements in first instar larvae. The observed mutant phenotypes are consistent with presence of non-homologous end joining (NHEJ) lesions in the targeted genes. This simple, fast and highly effective method for gene-specific mutagenesis together with the availability of genome sequences and transcriptomes, will significantly facilitate future studies of gene functions and gene regulation in *Ceratitis* as well as advance the design and development of new effective strategies for pest control management.

Introduction

The Mediterranean fruitfly *Ceratitis capitata* (medfly) is an economically relevant agricultural pest infesting more than 260 crop species, including fruits, vegetables, and nuts[1], Wild populations can be contained by the Sterile Insect Technique (SIT), an eradication strategy based on the repeated release of large numbers of laboratory-grown sterile males into infested areas[2,3]. *Ceratitis capitata* was the first non-Drosophilidae insect species in which transposon-mediated germline transformation was established[4,5]. Also embryonic RNA interference was successfully applied to study *in vivo* functions of *Ceratitis* genes controlling female sex determination during[6,7]. Various *Ceratitis* transgenic strains have been developed, aiming to improve SIT and other pest control strategies[8-16]. Nonetheless, a more comprehensive study of gene functions in *Ceratitis* will be needed to further improve existing control strategies. To generate long-lasting and hereditable changes in gene function, the CRISPR-Cas9 system with its modular and simple components provides a promising tool for insects[17,18]. Furthermore, transgene-based CRISPR-Cas9 can be used to produce homozygous loss-of-function mutations as well as a novel gene drive system for insect population control[19,20].

Various teams reported the successful use of the Cas9 system to mediate genome modifications in *Drosophila melanogaster* based on injecting different combinations of its components into embryos such as DNA plasmids expressing Cas9 protein plus single-guide RNA (sgRNA)[19], *in vitro*-transcribed *Cas9* mRNA plus sgRNA[21,22], and sgRNA into transgenic flies which express Cas9 in the germ line[23,24]. Lee et al., 2014[25] injected purified Cas9 protein preloaded with 2 RNAs, trRNA (transactivating RNA) and gene-specific crRNA (CRISPR RNA), into *Drosophila* embryos and observed a high rate of Cas9-induced genetic lesions. As Basu *et al*., 2015[26] reported that the use of a single sgRNA molecule, including both trRNA and crRNA, is more effective than the trRNA/ crRNA dual system, we opted for this technical improvement[27].

Altogether, we decided on the strategy of injecting Cas9 ribonucleoprotein (RNP) complexes into insect embryos for the following reasons: 1) preloaded Cas9 complexes should act immediately without delay following injection; 2) it has higher efficiencies compared to other approaches; 3) potentially less off-target events[28,29]; and, 4) Cas9 protein is more stable than Cas9 mRNA.

Here, we show that injecting *in vitro* generated Cas9-sgRNA RNPs into *Ceratitis capitata* embryos is very effective at inducing mono- and bi-allelic lesions of the targeted genes in both somatic and germline cells.

RESULTS

Cas9-induced somatic disruption of the *we* gene

To test the feasibility of Cas9-mediated gene disruption in *Ceratitis*, we targeted the *white eye (we*) gene, a locus required for eye pigmentation[4,5,30] . The *Ceratitis we* gene is the ortholog of the X-linked *white* (*w*) gene in *Drosophila*[5] and is located on the *Ceratitis* fifth chromosome[31,32]. The *we* gene is an ideal target to test Cas9-mediated disruption for the following reasons: 1) lack of eye pigmentation is an easily scored phenotype; 2) a *we*-mutant *Ceratitis* strain is available to test new loss-of-function alleles for complementation, and 3) *we* function is cell-autonomous and hence somatic mutant cells can be readily detected in the adult eye.

CHOPCHOP was used to identify potential Cas9 target sequences in *we*[33] and to design three sgRNAs, *we*-g1, *we*-g2, and *we*-g3 **(Fig. 1A**). Injections with unloaded recombinant Cas9 protein caused an almost twofold lower survival rate (15-19%) compared to buffer alone injections (30%), suggesting a measurable level of toxicity of Cas9 protein in *Ceratitis* embryos **(Tab. 1)**.

We next loaded recombinant Cas9 protein with individual sgRNAs *in vitro* and injected the RNP complexes into early syncytial embryos of the wildtype Benakeion strain. We aimed at targeting syncytial nuclei to maximize the efficiency of inducing NHEJ lesions. A

biallelic hit at this early stage is expected to produce large clones of mutant tissue in injected individuals[19,28]. In this first round we injected the Cas9/*we*-g1 in a buffer containing 300 KCl mM to improve solubilization of Cas9[28,34] . Of 240 injected embryos, 64 larvae hatched and 6 survived to adulthood **(Tab. 2)**. Three displayed a mosaic pattern of *white* unpigmented ommatidia surrounded by wild-type pigmented ommatidia **(Fig. 1B**, **C** and **D)**. One of the two eyes of one individual was completely white, suggesting that a biallelic gene disruption event occurred at a very early stage in the primordial lineage **(Fig. 1C)**. Two additional rounds of embryos injections were performed with *we*-g2 and *we*-g2/*we*-g3 RNPs, respectively **(Tab. 2)**. In these experiments the KCl concentration in the buffer was reduced to 150 mM[25] and a lower percentage of adults displayed eye pigmentation mosaicism (4% for *we*-g2 and 23% for *we*-g1/*we*-g2; Table 2).

The presence of mutations was confirmed by sequencing of PCR products spanning the cleavage sites. Genomic *we* PCR products were obtained from pools of injected larvae or single adult flies (see Supp data). Indels - mostly deletions- of variable length, were detected, consistent with previous studies[18].

Sequencing of amplified genomic DNA showed heterogeneity in nucleotide calls around the cleavage site close to the protospacer-adjacent motif (PAM) site, consistent with a range of different NJEH-induced alterations. Sequencing of cloned PCR products showed NHEJ deletions, ranging from 2-21 bp, relative to the PAM site of *we*-g1 and *we*-g2. However, duplex Cas9 targeting did not produce any deletions between the two targeted sites (**Fig. 1E**, *we*-g2: 4-6). On the other hand, when injecting *we*-g2 and *we*-g3 RNPs in early embryos, we were able to recover deletions spanning the 2 targeted sites ranging from 355 bp to 673 bp **(Fig. 1F)**.

Cas9-induced germ line disruption of the *we* gene

To test for germline transmission of Cas9-induced *we* alleles, 9 injected G0 red eyed flies (3 *we*-g1/RNP flies and 6 *we*-g2 RNP flies; **Tab. 3**) were individually crossed with *we* mutant partners (carrying a frameshift mutation in the 6th exon[30]). Seven injected flies sired G1 progeny, six of which gave rise to mutant *white eye* flies **(Fig. 2A)**, with a highly variable transmission rate (1.5% - 100% **(Tab. 3)**. Non-complementation of the CRISPR-Cas9-induced mutations confirms that they are allelic to the original *we* mutation **(Fig. 2A)**. Of the three *we*-g1 injected individuals, two males sired small batches of progeny in which 100% (*we-g1#2*; 6 flies) and 45% (*we-g1#3; 10 out of 22*), respectively, displayed the mutant phenotype (**Tab. 3).** Of the six *we*-g2 injected individuals, 4 males produced various proportions of G1 white-eyed mutant progeny. Remarkably, the *we*-g1#2 and *we*-g2#4 lines gave rise to 100% and 99% G1 white-eyed offspring, respectively **(Tab. 3)**. Thus, our results demonstrate that Cas9 activity is highly effective in the germ line, producing mostly mutant primordial germ cells.

We have randomly chose 4 mutant *we-g1* targeted (two from *we-g1#*2 cross and two from *we*-g1#3 cross) and 2 mutant *we-g2* targeted G1 flies (one from we-g2#3 line and one from we-g2#4 line)**(Tab. 3)**. The 4 mutant G1 flies bore both one common allele from the *we* strain and 3 novel Cas9-induced *we* alleles inherited from the injected fathers. An identical *we* deletion of 14 bp (*we-g1/1* allele) was found in 2 *white eye* flies from the line *we-g1#2*, suggesting that both flies inherited the same mutation from their common male founder **(Fig. 2B).** Mutant G1 flies of the *we-g1#3* line bore two different alleles (4 bp and 10 bp deletions; we-g1/2 and we-g1/3, respectively)**.** All three CRISPR-induced alleles are small deletions, causing frame-shifts in the exon 2 *we* coding region, and did not complement the original *we* mutation.

Two *we-g2* targeted G1 mutant flies bore two novel alleles, one with a 4 bp deletion and one with an unusually long 84 bp deletion **(Fig. 2B)**.

To study genes for which no mutant alleles are available, it would be useful to screen for mutant phenotypes by *inter se* crossings of G0 individuals in which the germ line has been targeted by CRISPR-Cas9. To test this possibility, we injected *we*-g3 or *we*-g2+g3 RNPs All of the surviving G0 flies, 20 and 38 respectively, did not show somatic mosaicism in the eyes **(Tab. 4)**. When crossed *inter se* **(Tab. 5),** we recovered two mutant individuals out of 26 G1 flies (8%) from the *we*-g3 cross and three out of 184 flies (2%) from the *we*-g2+g3 cross, demonstrating the feasibility of the crossing strategy.

Sequencing of the targeted regions in 2 flies of the *we-g3* cross identified, as expected, 4 novel *we* alleles. 2 out ot 3 tested flies of the *we-g2+we-g3* cross bore all an identical 9 deletion *we* allele at *we*-g3, most likely derived from the same G0 injected parent **(Fig. 2C)**. One of these 3 flies carried a large deletion of 650 bp provoked by duplex targeting with *we*-g2+*we*-g3.

This experiment showed that a cross of 20 and 38 adult G0 flies from injected CRISPR/Cas9 embryos can lead to 2-8% of mutant G1 flies bearing heteroallelic loss of function combinations within two medfly generations (less than 2 months).

Cas9-induced somatic disruption of the *Ceratitis paired* gene

The *Drosophila paired* gene (*prd*) is zygotically expressed and required for proper segmentation of the developing embryo[35]. *Drosophila* embryos homozygous for loss-of-function alleles lack every other segment and die before hatching (pair-rule phenotype). As early segmentation events are well conserved amongst higher dipterans[36], it can be assumed that an ortholog of *prd* is also present in *Ceratitis*. If we can induce mutant somatic clones of this ortholog by Cas9-sgRNA injections, we should observe segmental defects in *Ceratitis* embryos. BLAST searches identified a candidate *prd* gene in *Ceratitis* with 66% amino acid identity over the N-terminal half of the protein containing the PRD domain and the PRD homeobox and 37% identity over the C-terminal half (*Ccprd* XP_004524654.1; Bopp et al., 1986). CHOPCHOP was used to select two target sequences in this candidate

prd gene, *Ccprd* (*Ccprd*-g1 and *Ccprd*-g2), in a region encoding the conserved PRD domain (green box; **Fig. 3A**). Two injections series with sgRNA-*Ccprd1* and one with sgRNA-*Ccprd2* resulted in a survival rates of hatching embryos of 35%, 24%, and 26%, respectively **(Tab. 1** and **6).** Approximately 1-5% of the injected embryos showed either delayed development (late hatching rate) or arrested at late stages of embryogenesis. This was not observed when injecting Cas9 alone. Some lethality was observed at larval stages, with few individuals showing impaired locomotor activity and abnormal cuticular morphology. 5 to 11% of injected individuals developed to adulthood **(Tab. 6)** with a delay of 1-2 days compared to flies injected with buffer alone or with white targeting RNPs.

Almost all of the embryos that failed to hatch appeared to have disorganized cuticular structures**.** Some were up to 50% shorter in size compared to control embryos and displayed a reduced number of segments **(Fig. 3 B)**. This phenotype is reminiscent of the *pair rule* phenotype described for *Drosophila*[35,37]. Interestingly, we observed a late phenotype in larvae that were able to hatch. They showed irregular movements in different segmental compartiments suggesting aberrations affecting muscular contractions **(Supp. data videos)**. Similar lethality and abnormalities were observed in embryos and larvae following *Ccprd-2 sgRNA RNP injections.* These embryos are most likely mosaics with variable proportion of normal cells ($Ccprd^{+}/Ccprd^{+}$ or $Ccprd^{+}/Ccprd^{-}$) and mutant cells ($Ccprd^{-}/Ccprd^{-}$) Because of the mosaic nature of Cas9-induced mutagenesis events in a single individual and that *Ccprd* acts cell-autonomously, we did not expect a complete pair rule phenotype. The morphological and structural deviations in those regions are consistent with the occurrence of large mutant clones caused by biallelic somatic clones arising from gene targeting. Sequencing of the targeted regions from surviving larvae injected with *Ccprd1* led to the identification of lesions in the locus **(Fig. 3D)**. Surprisingly, efforts to identify gene targeting events caused by *Ccprd*-g2 failed. All 60 different plasmid clones from the corresponding region showed only wild type sequence. We conclude that *Ccprd1* targeting was effective in inducing indels which correlate with embryos/larval malformations most likely due to impaired segmentation.

DISCUSSION

Over the last two decades novel genetic strategies in pest insect management have been developed to improve their effectiveness in the field. Genetic technologies used thus far in the medfly have been based on the random integration of transposable elements into the genome[4], site-specific modification of the randomly integrated transgene[12] and embryonic or transgene-mediated RNA interference (RNAi)[6,15]. The disadvantage is that such genetically modified medfly must be continuously tested with respect to fitness and competitiveness as well as to stability and expression of the transgene[38]. The CRISPR-Cas9 technology offers the possibility to avoid random integration of exogenous DNA but instead provides a more robust and controlled approach to introduce new genetic features, which can be either addition of exogenous DNA or nucleotides changes in specific genes. Here we explored the use of Cas9-sgRNA RNP complexes, avoiding plasmid-based or DNA-mediated delivery of Cas9. This DNA-free (plasmid or integrated transgene) method may allow to circumvent existing regulatory restrictions which are in effect in most nations concerning the use of GM organisms in the field. This "green" editing technology may facilitate global acceptance not only for plants or fungi, but also for insect pest control[39].

The use of preloaded Cas9-sgRNA complexes is becoming a successful approach for targeted disruption of genes in a growing number of species, including *Drosophila melanogaster*[25], *Aedes aegypti*[27] and, recently the zebrafish *Danio rerio*[28,40]. To our knowledge, this is the first report showing that the same approach can be used to effectively mutate genes in a major agricultural pest insect such as *Ceratitis capitata*, with up to 100% mutagenesis rate in the germ line. As observed by Lee et al. (2014)[25], Cas9-sgRNA complexes act immediately but are rapidly degraded, often within few hours after administration; similarly Kim et al. (2014)[17] reported that Cas9 protein is degraded within 24 h after being applied to cultured human cell lines. As the short-lived activity of Cas9 prevents the induction of late mutational events, this may help to reduce off-target effects and mosaicism of mutant and wild type tissues in the injected individuals.

We report here that Cas9-mediated NJEH events generates lesions in the *Ceratitis we* and

Ccprd genes. Targeting the *we* gene caused red-white eye mosaicism in up to 50% of injected invididuals, indicating a very high rate of somatic bi-allelic hits. Transmission of *we* mutated alleles to G1 progeny was found to be highly effective being close to 100%. Lee et al., (2014)[25] observed a 5-time lower germline transmission rate (20%) when targeting *Drosophila* genes using RNPs. One possible explanation is that the higher concentration of KCl (300 mM) used in our study increases Cas9 stability or solubility[28]. Injecting RNP complexes in *Aedes aegypti*, Basu et al., (2015)[26] found a germline transmission rate of induced mutations up to 90%. Yet this rate was determined with high-resolution melt analysis (HRMA) rather than phenotypic analysis of G1 adults, as we used in our study. Gilles et al., (2015)[41] observed up to 100% of mutant progeny, by co-injecting Cas9 mRNA and sgRNA-producing DNA plasmid into the coleopteran *Tribolium* embryos. However, the author targeted a single copy dominant marker transgene, rather than 2 alleles of an endogenous gene.

Several studies have reported that the simultaneous use of 2 Cas9-sgRNAs is an effective means to generate deletions of sequences between two targeted sites[27,42]. While *we*-g1 and *we*-g2 RNPs were individually effective in gene editing, the absence of a 96bp long deletion spanning the two targeted sites suggests that steric hindrance between two adjacent Cas9-sgRNA complexes could have caused the cut of only one of the two. In contrast, the use of *we*-g2 and *we*-g3 RNPs, simultaneously targeting two sites more distant of each other (489bp) was effective in both somatic (G0) and germ-line cells (G1). On the other side, the use of the single *we*-g2 RNP provoked a 84bp long deletion in germ line cells (Fig. 2).

Hence it is feasible that a multiplex CRISPR-Cas9 system can be adapted to specifically remove exons also in the medfly. This feature could be used to test the function of protein domains encoded by single exons. We can envision many more uses for generating precise deletions such as the analysis of *in vivo* putative gene *cis*-regulatory elements, or to investigate potentially redundant functions of duplicated genes, or to remove long non-coding RNAs or miRNAs.

With 2 sgRNA/RNPs, we targeted the Ceratitis zygotic gene *Ccprd* potentially involved in embryos segmentation. We observed mutant phenotypic effects (embryos and larval malformations, impaired larval motility, delayed development) and mutated sequences at one of the 2 targeted sites. The lack of DNA sequence changes for the *Cc-prd*-g2 and the observed phenotypic effects seems to be inconsistent, but this could be due to different reasons: 1) the lower efficiency of this second sgRNA and the lower representation in the somatic larval clones of the mutated sequence with the respect of the wild type one; 2) *Cc-prd* RNPs could interfere also with early *prd* transcription delaying and/or lowering its zygotic gene expression; and, 3) off target effects impairing related gene(s) encoding PRD box domain. Additional investigations will be required to understand these potential problems.

CRISPR-Cas9 will be helpful in investigating natural traits of this major agricultural pest, for example invasiveness and host adaptation, reproduction, olfaction (fruit seeking behavior), chemoreception, toxin and insecticide metabolism[43]. The recent availability of a medfly genome draft combined with the successful implementation of CRISPR genome editing technology, as reported here, opens the road to transfer basic knowledge to applied research. Next challenges for the CRISPR-Cas9 technology in *Ceratitis* will be to exploit homology-directed recombination for genome editing, either to insert transgenes in specific regions or to replace DNA sequences with slightly mutated ones.

Methods

Rearing of *Ceratitis capitata.* The *C. capitata* wildtype strain were reared in standard laboratory conditions at 25 °C, 70% relative humidity and 12 : 12 h light–dark regimen. Adult flies were fed with yeast/sucrose powder (1 : 2). Eggs were collected in water dishes, and transferred to larval food (soft tissue paper 30 g, sugar 30 g, yeast extract 30 g, cholesterol stock 10 ml, HCl stock 2 ml, Benzoic stock 8,5 ml, water 400 ml). Pupae were collected and stored in Petri dishes until eclosion.

Strain of *Ceratitis capitata*

1. The wild-type eye color strain Benakeion was originally established in the laboratory by P. A. Mourikis (Benakeion Instirescue of Phytopathology, Athens, Greece).
2. The strain *we/we* Benakeion[30] was kindly provided by Prof. Kostas Bourtzis (*Pest Control of* FAO/IAEA,Seiberdorf, Austria)

Cas9 purification. We produced our own supply of Cas9 endonuclease by expressing HIS tagged protein in bacteria[25] and following the purification protocol described in Monti et al.[44] and Dathan et al.[45].The *pET* plasmid that encodes His-tagged Cas9 was transformed into BL21(DE3). The recombinant protein expression was induced in the presence of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 16 h at 22 °C. A 100 mL pellet was re-suspended in 10 mL of cold lysis buffer (50 mM Phosfate, 500 mM NaCl, 10 mM Imidazole, 1 mM DTT, pH 8) supplemented with protease inhibitor mixture (1 mM PMSF and 1.0 mg/mL of lysozyme) and incubated at room temperature for 30 min. Cells were disrupted by sonication on ice with 10 s on/10 s off cycles for a total of 10 min on. After centrifugation at 14,000 rpm for 30 min at 4 °C, the supernatant was purified on an ÅKTA FPLC chromatography system using a 1 mL HisTrap HP. The column was washed with lysis buffer and bound protein was eluted using a gradient of 10 mM–500 mM imidazole. Protein elution was monitored by measuring absorbance at 280 nm and the resulting fractions were analyzed by SDS-15% PAGE. The eluted fractions were dialyzed against buffer (20 mM HEPES, 150 mM KCl, 1 mM DTT, 10% glycerol, pH 7.5). Analysis of the protein by gel-filtration was compared with a calibration obtained with marker proteins run on the column under the same conditions. Gel-filtration analyses were carried out on a Superdex 200 GL column.

sgRNA Design and Synthesis. sgRNA were designed using CHOPCHOP https://chopchop.rc.fas.harvard.edu/[33]. CHOPCHOP lists the Target Sequence (including the PAM), the genomic location of the target, the strand (- or +), the GC content of the guide and the Off-targets. Following templates for sgRNA production were performed, with minor modifications, as described by Bassett et al (2013). We selected the target of interest, for *white eye* we used three target site sgRNAs_

g1: 5' GAGTAAGTGAGATTATCCG 3';

g2: 5' GCTGGTGAATCGTGTGAAGG 3' ;

g3: 5' CGGGTGAAGGGTTTATCGGG 3'

with respectively, protospacer-adjacent motif (PAM): CGG; GGG; TGG. For *paired* we had two target site sgRNA_Prd1: 5' GGTCGCGTCAATCAATTAGG 3' with PAM, TGG and sgRNA_Prd2: 5' AGAATCCCAGCATATTTTCG 3' with PAM , TGG. We added to the tagert sequence to the CRISPR_F sequence as shown below:5'-GAAATTAATACGACTCACTATAGG[20nt of CRISPR target same strand as PAM sequence]GTTTTAGAGCTAGAAATAGC-3'. We used seme sgReverse (PAGEpurified):5'AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATGGACT AGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC AAC. The PCR was performed using Q5® High-Fidelity DNA Polymerase (NEB). The DNA template was extract over 400 µL phenol:chloroform (1:1), followed by extraction over 400 µL chloroform. To the final upper layer (~ 200µL), add 20 µL 3M NaOAc (pH 5.2) and 400 µL 100% EtOH and precipitate at -20°C for at least 1 h. Pellet the DNA at 14,000 rpm 15 min 4°C; Wash the pellet in 70% EtOH (3x), allowing for each washing step to spend at least one hours at -20°C. Air dry and resuspend the pellet in ~40µL ddH2O. *In vitro* transcription of sgRNAs we followed the instructions of Megatranscript T7 kit (Ambion) using 400 ng of target template with reaction run at 37°C over night. After this step at reaction was added 2 uL of Turbo DNase and incubate for an additional 15 min at 37°C to remove the DNA template.

The assembly of complex RNP and the injection mix, The reaction of binding was prepared with 1.8 µg of Cas9 protein , 1 µg of sgRNA and was added KCl according to the protocol proposed by Burger et al.[28], and incubation for 10 min at 37°C. Embryos were collected 1 hour AEL (after egg laying), hand-dechorionated and microinjected.

Genomic DNA extraction and molecular analysis. DNA extraction was performed, with minor modifications, according to the protocol proposed by Holmes and Bonner [**1973**][46]. Each larvae and adult were placed in a 1.5 ml Eppendorf tube was crushed by a pestle in 200 µl Holmes Bonner buffer (urea 7 M, Tris-HCl 100 mM pH 8.0, EDTA 10 mM pH 8.0, NaCl 350 mM, SDS 2%). Subsequently, the DNA was purified with one phenolchloroform extraction, followed by one chloroform extractions and then an ethanol precipitation. The precipitated pellet was resuspended in 30 µl or 100 µl water containing RNaseA. Genomic DNA (from single adult fly or larae or poll of larvaes or G1 mutants) was used as the template for to amplify the sgRNA target sites. We used as follow for *gene white* (F, forward primer; R, reverse primer):

F_g1_g2_g3- 5'GCCCTACGAGCAATCCTCT 3' ;

R_g1_g2- 5'TCTGCAATGAGCGTCATATAC 3';

R_g2_g3- 5'TTCTGCGATAGCTTTTTCAACA 3'.

For *gene Paried* we used

F_Prd_1_2- 5'CTTCGACACACAACCGTGTG 3' and

R_Prd_1_2- 5'AGAATGCTTGTGGGAATGTTCT 3'.

We followed these PCR using DreamTaq (Life Technologies) according to the manufacturer's instructions. The PCR products were purified with StraPrep PCR Purificaton Kit (Agilent Techologies) and and sub-cloned using with StrataClone PCR cloning Kit (Agilent Techologies) and the positive clones were sequenced using the plasmid primers T7 and T3 and analyzed on an ABI 310 Automated Sequencer (Applied Biosystems).

Acknowledgments

We thank Prof. Jin-Soo Kim (Department of Chemistry at Seoul National University, Korea). (Korea) for providing the His-tagged Cas9 expression plasmid. We wish to acknowledge that the partecipation of AM to a short course on CAS9 technology organized by the Insect Genetic Technology Research Coordination Network was important to successfully develop this work (http://igtrcn.org; 17-21 August 2015). Hence we thank Dave O'Brochta, Max Scott and Bill Reid for the invaluable and kind help.

We wish to thank also Tessa G. Montague (Harvard Univ.) which added the medfly genome to CHOPCHOP. We wish to thank Claudia Brunner (Bopp lab) for technical assistance. GS wish to thank Jin-Soo Kim, Kate O'Connor-Gilles (USA) and Andrew R. Bassett (UK) for sharing information and reagents. GS wish to thank Marcelo Jacobs-Lorena for suggestions on the manuscript.

Contributions: A.M. and G.S. designed the experiments. A.M. and S.M.M expressed and purified the CAS9, sgRNAs design and selection was performed by A.M. and M.R., A.M. produced sgRNAs, M.R. and HL analyzed induced indels. D.B. hosted in his lab (Zurich Univ) A.M. and G.S. to establish CAS9 technology in the medfly and in *Musca domestica*; the purified CAS9 was tested in zebrafish by C.M., which together with AB provided also a lab protocol and technical tips concerning MOOPs gel analysis of sgRNAs, and KCl concentration to improve Cas9 stability/efficiency (Zurich Univ.); GS wrote the manuscript with major inputs by DB and contributions, by A.M. and C.M. A.M. supervised 4 master students (R.C., M.G.I., P.P., G.D.C) over a 2 years long work, which contributed to the this study, with medfly rearing, crossing flies, PCR analysis, plasmid cloning and DNA sequencing. E.C. contributed with sequencing and analysis of 20 plasmid clones from Cas9/sgRNA injected G0 individuals (*we* and *Ccprd*). MS contributed to the DNA sequence analysis of mutant alleles. A.Ra. and G.I. helped with the embryos micro-injections set up and larvae video recording. A.M. and S.M.M. contributed in the Cas9

purification and protocol set up. GS thanks Marcelo Jacobs Lorena and Philippos Papathanos for critical reading of the manuscript.

Competing financial interests statement

The authors declare no competing financial interests.

Additional information

Tables

Table 1 Survival rate after Cas9 alone injections

Table 2 Statistics of G0 *we* somatic CRISPR mutations

Table 3 Statistics of G1 Germline transmission of CRISPR-induced *we* mutations with crosses of G0 single flies.

Table 4 Injections data of *we*-g3 and *we*-g2+g3 RNPs in early embryos.

Table 5 G1 Germline transmission of CRISPR-induced *we* mutations from pooled G0 injected flies (*we*-g3 and *we*-g2+g3).

Table 6 Statistics of G0 *paired* CAS9 targeting

Figure legends

Figure 1 CRISPR-Cas9 targeting of *we* gene: (A) A scheme of the *we* gene and the 3 sgRNAs targeting conserved Drosophila WHITE regions in Ceratitis. (B) wild type and mutant strain with white eye flies. (C) and (D) examples of G0 somatic mosaics. € Sequences of *white eye* gene edited alleles in G0 somatic mutant individuals. (F) Sequences of *white eye* alleles in G0 somatic mutant individuals, targeted with 2 RNPs.

Figure 2 *we* mutants from G1 progeny were analyzed by DNA sequencing. Absence of complementation of CRISPR-Cas9 mutated *we* alleles in G1 mutant flies in heterozygotes, carrying one original allele from the *we* medfly strain (3 wg-1 and 2 wg-2 induced alleles in 5 flies; vertical blu bar). Heteroallelic combinations of novel Cas9-mutated *we* alleles in

5 G1 individuals (11 alleles; vertical green bar). In G1 alleles, the g2 in parenthesis indicate that a second RNP was coinjected but no corresponding mutations were found in each individual. In the fifth mutant fly, 3 different mutated alleles (*we*-g3(g2)/3a, b and c) were surprisingly found. This could be due to 2 sperm fertilization, following the first division of the haploid egg.

Figure 3 CRISP-Cas9 targeting of *Ccprd.* (A) A scheme of the *Ccprd* gene and 2 sgRNAs targeting conserved *Drosophila* PRD in *Ceratitis*. (B) Comparison of wt and injected embryos, showing shorted lenght and a reduction in the segment number. (C) 2 injected larvae: larvae 1 shows incomplete/malforrmed development with the respect of larvae appeaaring wild type. (D) Sequences of gene edited *Ccprd* alleles from G0 larvae injected with RNP containing *Cc-prd1* sgRNA.

1 Liquido NJ, S. L., Cunningham RT. Host plants of the Mediterranean fruit fly (Diptera, Tephritidae). An annotated world list. *Ann Entomol Soc Am.* **77:1–57.** (1991).

2 Dyck, V. A. H., J.; Robinson, A.S. . Sterile Insect Technique Principles and Practice in Area-Wide Integrated Pest Management. *Eds* **XIV, 787 p., ISBN: 1-4020-4050-4.** (2005).

3 Morrison NI, F. G., Koukidou M, Miller TA, Saccone G, Alphey LS, Beech CJ, Nagaraju J, Simmons GS, Polito LC. Genetic improvements to the sterile insect technique for agricultural pests. *Asia Pacific Journal of Molecular Biology and Biotechnology* **18(2):275-295** (2010).

4 Loukeris, T. G., Livadaras, I., Arca, B., Zabalou, S. & Savakis, C. Gene-Transfer into the Medfly, Ceratitis-Capitata, with a Drosophila-Hydei Transposable Element. *Science* **270**, 2002-2005, doi:DOI 10.1126/science.270.5244.2002 (1995).

5 Zwiebel, L. J. *et al.* The White Gene of Ceratitis-Capitata - a Phenotypic Marker for Germline Transformation. *Science* **270**, 2005-2008, doi:DOI 10.1126/science.270.5244.2005 (1995).

6 Pane, A., Salvemini, M., Bovi, P. D., Polito, C. & Saccone, G. The transformer gene in Ceratitis capitata provides a genetic basis for selecting and remembering the sexual fate. *Development* **129**, 3715-3725 (2002).

7 Salvemini, M. *et al.* Ceratitis capitata transformer-2 gene is required to establish and maintain the autoregulation of Cctra, the master gene for female sex determination. *Int J Dev Biol* **53**, 109-120, doi:10.1387/ijdb.082681ms (2009).

8 Dafa'alla, T. H. *et al.* Transposon-free insertions for insect genetic engineering. *Nat Biotechnol* **24**, 820-821, doi:10.1038/nbt1221 (2006).

9 Gong, P. *et al.* A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly. *Nat Biotechnol* **23**, 453-456, doi:10.1038/nbt1071 (2005).

10 Salvemini, M., Mauro, U., Velaeti, S., Polito, C. & Saccone, G. A new Minos vector for eye-specific expression of white plus marker in Ceratitis capitata and in distantly related dipteran species. *Insect Mol Biol* **15**, 341-349, doi:10.1111/j.1365-2583.2006.00651.x (2006).

11 Schetelig, M. F., Caceres, C., Zacharopoulou, A., Franz, G. & Wimmer, E. A. Conditional embryonic lethality to improve the sterile insect technique in Ceratitis capitata (Diptera: Tephritidae). *Bmc Biol* **7**, doi:Artn 4 10.1186/1741-7007-7-4 (2009).

12 Schetelig, M. F. *et al.* Site-specific recombination for the modification of transgenic strains of the Mediterranean fruit fly Ceratitis capitata. *P Natl Acad Sci USA* **106**, 18171-18176, doi:10.1073/pnas.0907264106 (2009).

13 Fu, G. L. *et al.* Female-specific insect lethality engineered using alternative splicing. *Nat Biotechnol* **25**, 353-357, doi:10.1038/nbt1283 (2007).

14 Ogaugwu, C. E., Schetelig, M. F. & Wimmer, E. A. Transgenic sexing system for Ceratitis capitata (Diptera: Tephritidae) based on female-specific embryonic lethality. *Insect Biochem Molec* **43**, 1-8, doi:10.1016/j.ibmb.2012.10.010 (2013).

15 Ruiz, M. F. *et al.* The Gene Transformer of Anastrepha Fruit Flies (Diptera, Tephritidae) and Its Evolution in Insects. *Plos One* **2**, doi:ARTN e1239 10.1371/journal.pone.0001239 (2007).
16 Scolari, F. *et al.* Fluorescent sperm marking to improve the fight against the pest insect Ceratitis capitata (Wiedemann; Diptera : Tephritidae). *New Biotechnol* **25**, 76-84, doi:DOI 10.1016/j.nbt.2008.02.001 (2008).
17 Kim, H. & Kim, J. S. A guide to genome engineering with programmable nucleases. *Nat Rev Genet* **15**, 321-334, doi:10.1038/nrg3686 (2014).
18 Reid, W. & O'Brochta, D. A. Applications of genome editing in insects. *Curr Opin Insect Sci* **13**, 43-54, doi:10.1016/j.cois.2015.11.001 (2016).
19 Gratz, S. J. *et al.* Genome Engineering of Drosophila with the CRISPR RNA-Guided Cas9 Nuclease. *Genetics* **194**, 1029-+, doi:10.1534/genetics.113.152710 (2013).
20 Hammond, A. *et al.* A CRISPR-Cas9 gene drive system-targeting female reproduction in the malaria mosquito vector Anopheles gambiae. *Nat Biotechnol* **34**, 78-83, doi:10.1038/nbt.3439 (2016).
21 Bassett, A. R., Tibbit, C., Ponting, C. P. & Liu, J. L. Highly Efficient Targeted Mutagenesis of Drosophila with the CRISPR/Cas9 System. *Cell Rep* **4**, 220-228, doi:10.1016/j.celrep.2013.06.020 (2013).
22 Yu, Z. S. *et al.* Highly Efficient Genome Modifications Mediated by CRISPR/Cas9 in Drosophila. *Genetics* **195**, 289-+, doi:10.1534/genetics.113.153825 (2013).
23 Kondo, S. & Ueda, R. Highly Improved Gene Targeting by Germline-Specific Cas9 Expression in Drosophila. *Genetics* **195**, 715-+, doi:10.1534/genetics.113.156737 (2013).
24 Ren, X. J. *et al.* Optimized gene editing technology for Drosophila melanogaster using germ line-specific Cas9. *P Natl Acad Sci USA* **110**, 19012-19017, doi:10.1073/pnas.1318481110 (2013).
25 Lee, J. S. *et al.* RNA-Guided Genome Editing in Drosophila with the Purified Cas9 Protein. *G3-Genes Genom Genet* **4**, 1291-1295, doi:10.1534/g3.114.012179 (2014).
26 Basu, S. *et al.* Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in Aedes aegypti. *P Natl Acad Sci USA* **112**, 4038-4043, doi:10.1073/pnas.1502370112 (2015).
27 Kistler, K. E., Vosshall, L. B. & Matthews, B. J. Genome Engineering with CRISPR-Cas9 in the Mosquito Aedes aegypti. *Cell Rep* **11**, 51-60, doi:10.1016/j.celrep.2015.03.009 (2015).
28 Burger, A. *et al.* Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. *Development* **143**, 2025-2037, doi:10.1242/dev.134809 (2016).
29 Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S. & Yang, S. H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther-Nucl Acids* **4**, doi:ARTN e264 10.1038/mtna.2015.37 (2015).
30 Gomulski, L. M. *et al.* Genomic organization and characterization of the white locus of the Mediterranean fruitfly, Ceratitis capitata. *Genetics* **157**, 1245-1255 (2001).
31 Zacharopoulou, A. Polytene Chromosome Maps in the Medfly Ceratitis-Capitata. *Genome* **33**, 184-197 (1990).

32 Zacharopoulou, A. *et al.* The Genome of the Mediterranean Fruit-Fly Ceratitis-Capitata - Localization of Molecular Markers by Insitu Hybridization to Salivary-Gland Polytene Chromosomes. *Chromosoma* **101**, 448-455, doi:Doi 10.1007/Bf00582839 (1992).

33 Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42**, W401-W407, doi:10.1093/nar/gku410 (2014).

34 Jinek, M. *et al.* Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science* **343**, 1215-+, doi:10.1126/science.1247997 (2014).

35 Nüsslein-Volhard C, W. E. Mutations affecting segment number and polarity in Drosophila. *Nature* **287** (1980).

36 Peel, A. D., Chipman, A. D. & Akam, M. Arthropod segmentation: Beyond the Drosophila paradigm. *Nat Rev Genet* **6**, 905-916, doi:10.1038/nrg1724 (2005).

37 Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. & Noll, M. Conservation of a Large Protein Domain in the Segmentation Gene Paired and in Functionally Related Genes of Drosophila. *Cell* **47**, 1033-1040, doi:Doi 10.1016/0092-8674(86)90818-4 (1986).

38 Harvey-Samuel, T., Ant, T., Gong, H. F., Morrison, N. I. & Alphey, L. Population-level effects of fitness costs associated with repressible female-lethal transgene insertions in two pest insects. *Evol Appl* **7**, 597-606, doi:10.1111/eva.12159 (2014).

39 Kanchiswamy, C. N., Malnoy, M., Velasco, R., Kim, J. S. & Viola, R. Non-GMO genetically edited crop plants. *Trends Biotechnol* **33**, 489-491, doi:10.1016/j.tibtech.2015.04.002 (2015).

40 Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**, 227-229, doi:10.1038/nbt.2501 (2013).

41 Gilles, A. F., Schinko, J. B. & Averof, M. Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle Tribolium castaneum. *Development* **142**, 2832-+, doi:10.1242/dev.125054 (2015).

42 Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *The Journal of biological chemistry* **289**, 21312-21324, doi:10.1074/jbc.M114.564625 (2014).

43 Papanicolaou, A. *et al.* The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome biology* **17**, 192, doi:10.1186/s13059-016-1049-2 (2016).

44 Monti, S. M., Supuran, C. T. & De Simone, G. Carbonic Anhydrase IX as a Target for Designing Novel Anticancer Drugs. *Curr Med Chem* **19**, 821-830 (2012).

45 Dathan, N. A. *et al.* Biochemical characterization of the chloroplastic beta-carbonic anhydrase from Flaveria bidentis (L.) "Kuntze". *Journal of enzyme inhibition and medicinal chemistry* **29**, 500-504, doi:10.3109/14756366.2013.813942 (2014).

46 Holmes, D. S. & Bonner, J. Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid. *Biochemistry* **12**, 2330-2338 (1973).

A

Cc *white eye* gene

1 Kb

we-g1 ←

GAGTAAGTGAGATTATCCG **CGG**

```
Cc we  PSPALTADNLTYSWYNLD
       ++++  A+NLTY+W N+D
Dm w   SGSGQLAENLTYAWHNMD
```

we-g2 →

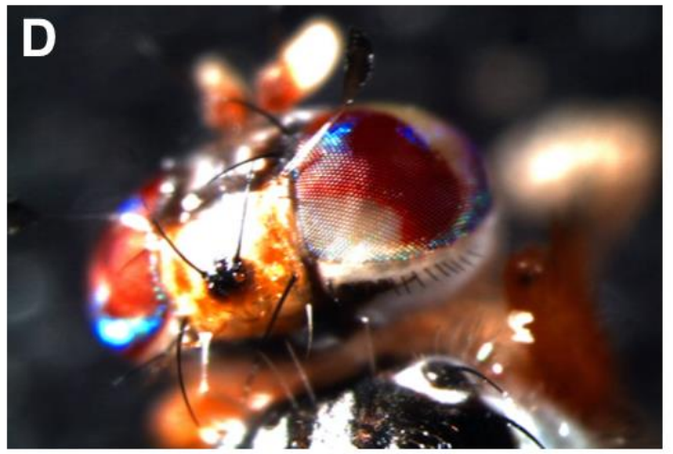
GCTGGTGAATCGTGTGAAGG **GGG**

```
Cc we  SSWKQLVNRVKGVFCNE   17
       S W QLVNR  G FCNE
Dm w   SGWRQLVNRTRGLFCNE  102
```

we-g3 →

CGGGTGAAGGGTTTATCGGG **TGG**

```
Cc we  VPGRVKGLSGGERKRL   16
       VPGRVKGLSGGERKRL
Dm w   VPGRVKGLSGGERKRL  240
```



E

we-g1

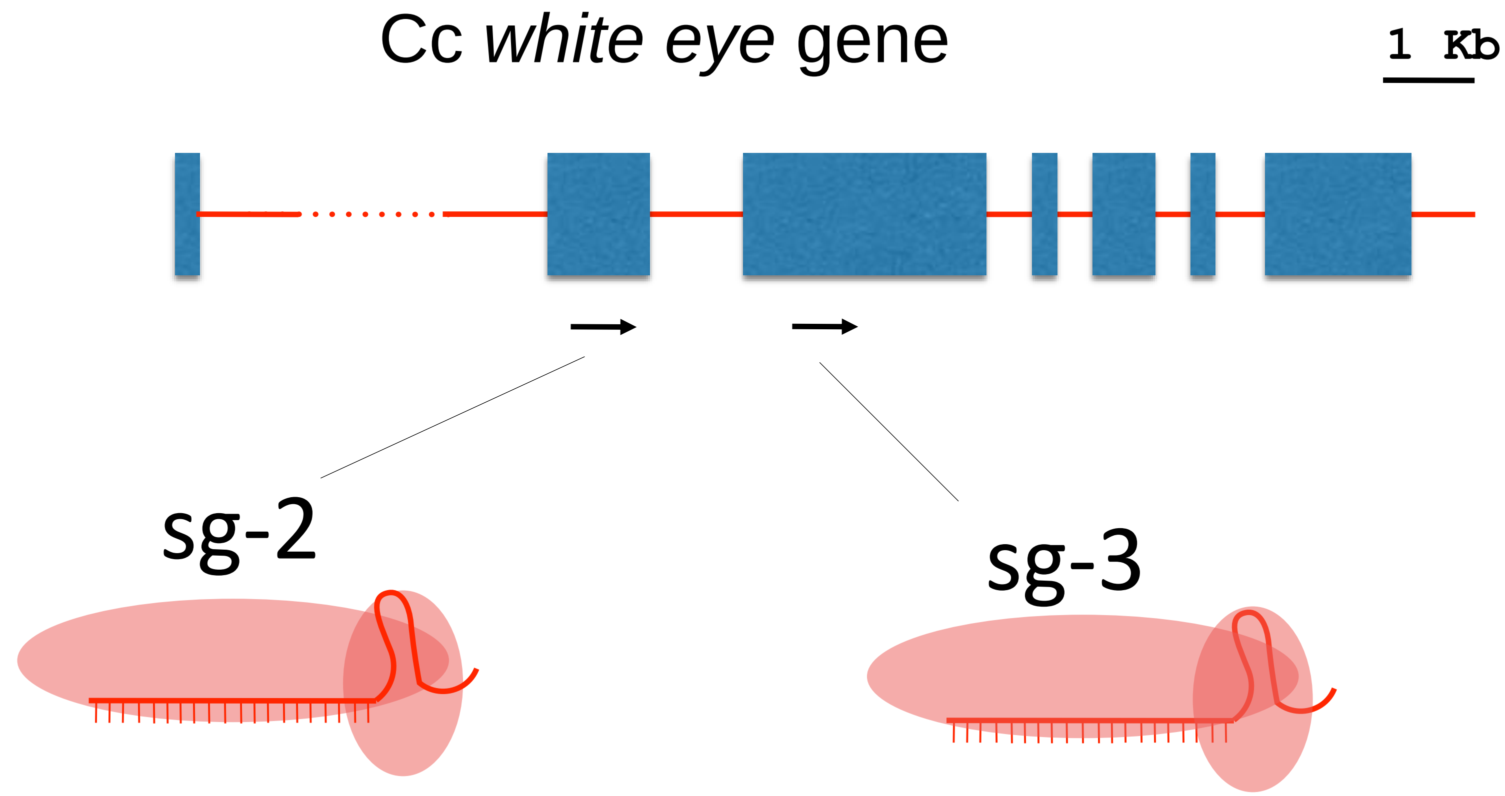
```
wt  CCATCGCCTGCACTCACCGCGGATAATCTCACTTACTCTTGGTACAATTTGGAC
1   CCATCGCCTGCACTCACCG---------------ACTCTTGGTACAATTTGTAC -15
2   CCATCGCCTGC-------------ATCTCACTTACTCTTGGTACAATTTGGAC -14
3   CCATCGCCTGCACTCACCGCG-------------A-TCTTGGTACAATTTGGAC -14
4   CCATCGCCTGCA---------------CTCACTTACTCTTGGTACAATTTGGAC -15
5   CCATCGCCTGCACTCACCGC--------------------GTACAATTTGGAC -21
```

we-g2

```
wt  GTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTGTGAAGGGGGTT
1   GTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTGTG--GGGGGTT -2
2   GTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTGTGAATCGGGTT 2S
3   GTACATCAGCCCGGCTCGAGTTG--------------------GAAGGGGGTT -21
4   GTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTGT---GGGGGTT -3
5   GTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCG------GGGGGTT -6
6   GTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGT-------------GGGTT -14
```

Fig. 1

F



		Del
wt	CGGCGCTGTACATCAGCCCGGCTCGAGTTGGAAGCA**GCTGGTGAATCGTGTGAAGG**GGG ——489—— -**CGGGTGAAGGGTTTATCGGG**TGGCGAGCGTAAGCGACTGGCATTCGCCTCGGAAG	
1	CGGCGCTGTACATCAGCCCGGCTCG -TCGGG**TGG**CGAGCGTAAGCGACTGGCATTCGCCTCGGAAG	-536
2	- **G**CGAGCGTAAGCGACTGGCATTCGCCTCGGAAG	-649
3	CGGCGCTGTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTG - GG**TGG**CGAGCGTAAGCGACTGGCATTCGCCTCGGAAG	-522
4	CGGCGCTGTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTGTGA - GAGCGTAAGCGACTGGCATTCGCCTCGGAAG	-516
5	CGGCGCTGTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTG - GGG**TGG**CGAGCGTAAGCGACTGGCATTCGCCTCGGAAG	-510
6	- -CGACTGGCATTCGCCTCGGAAG	-673
7	CGGCGCTGTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTGTG - ATCGGG**TGG**CGAGCGTAAGCGACTGGCATTCGCCTCGGAAG	-517
8	CGGCGCTGTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTGTG - - - - - - - - - - TGA**A**T**C**G**T**GTGCCGGGTCGGGTGAAGGGTTTATCGGG**TGG**CGAGCGTAAGCGACTGGCATTCGCCTCGGAAG	-489
9	CGGCGCTGTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTGTGA - GGG**TGG**CGAGCGTAAGCGACTGGCATTCGCCTCGGAAG	-511
10	CGGCGCTGTACATCAGCCCGGCTCGAGTTGGAAGCAGCTGGTGAATCGTGTGAAGGGGG(+40bp)- - - - - - - - - - - - - - - (160bp+)ATCGGG**TGG**CGAGCGTAAGCGACTGGCATTCGCCTCGGAAG	-355

Fig. 1: CRISPR/Cas9 targeting of *we* gene: A) A scheme of the *we* gene and the 3 sgRNAs targeting conserved Drosophila WHITE regions in Ceratitis. B) wild type and mutant strain with white eye flies. C and D) examples of G0 somatic mosaics. E) Sequences of *white eye* gene edited alleles in G0 somatic mutant individuals. F) Sequences of *white eye* alleles in G0 somatic mutant individuals, targeted with 2 RNPs.

Fig. 2

A

B

```
we+          AAGAGTAAGTGAGATTATCCGCGGTGAGTGCAGGCGATGGCGGTGAGAGTG  wt
we-g1/1      AAGAGTAAGTGAGAT--------------GCAGGCGATGGCGGTGAGAGTG -14
we-g1/2      AAGAGTAAGTGAGA----CCGCGGTGAGTGCAGGCGATGGCGGTGAGAGTG -4
we-g1/3      AAGAGTAAGTGAG----------GTGAGTGCAGGCGATGGCGGTGAGAGTG -10

we+          ACCGCGGATAATCTCAC(64bp)GCTGGTGAATCGTGTGAAGGGGGTTTTC  wt
we-g2/1      ACCGCGGATAATCTCAC----------(84bp)-----------GGTTTTC -84
we-g2/2      ACCGCGGATAATCTCAC(64bp)GCTGGTGAATCGTGT------GGTTTTC -4
```

G1

C

```
we+          ATACGTTGATTGGCGTGCCGGGTCGGGTGAAGGGTTTATCGGGTGGCGAGC wt
we-g3/1a     ATACGTTGATTGGCGTGCCGGGTCGG----------------GTGGCGAGC -16
we-g3/1b     ATACGTTGATTGGCGTGCCGGGTCGGGTGAAGGGTTTAGCGAGGGTGGCGAGC +2
we-g3/2a     ATACG-----------------------------------CTGGTGGCGAGC -34
we-g3/2b     ATACGTTGATTGGCGTGCCGGGTCGGGTGAA---------GGGTGGCGAGC -9
we-g3(g2)/1a ATACGTTGATTGGCGTGCCGGGTCGGGTGAA---------GGGTGGCGAGC -9
we-g3(g2)/1b ATACGTTGATT-----------------------------GGGTGGCGAGC -29
we-g3(g2)/2a ATACGTTGATTGGCGTGCCGGGTCGGGTGAA---------GGGTGGCGAGC -9
we-g3(g2)/2b ATACG-----------------------------------CTGGTGGCGAGC -34
we-g3(g2)/3a ATACGTTGATTGGCGTGCCGGGTCGGGTGAA---------GGGTGGCGAGC -9
we-g3(g2)/3b ATACGTTGATTGGCGTGCCGGGT----------------TCGGGTGGCGAGC -15
we-g3+g2/3c  ------------------(650 bp)--------------------GCGAGC -650
```

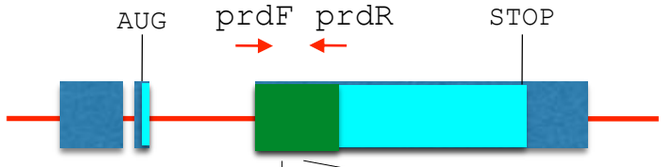
G1

Fig. 2: Absence of complementation of CRISPR/Cas9 mutated *we* alleles in G1 mutant flies in heterozygotes, carrying one original allele from the *we* medfly strain (3 wg-1 and 2 wg-2 induced alleles in 5 flies; vertical blu bar). Heteroallelic combinations of novel CRISPR/Cas9 mutated *we* alleles in 5 G1 individuals (11 alleles; vertical green bar). In G1 alleles, the g2 in parenthesis indicate that a second RNP was coinjected but no corresponding mutations were found in each individual. In the fifth mutant fly, 3 different mutated alleles (*we*-g3(g2)/3a, b and c) were surprisingly found. This could be due to 2 sperm fertilization, following the first division of the haploid egg.

A

Cc paired

1 Kb



sgRNA prd1

GGTCGCGTCAATCAATTAGGTGG

V N Q L G G V

Ccprd VNQLGGVFI
VNQLGGVFI
Dmprd VNQLGGVFI

sgRNA prd2

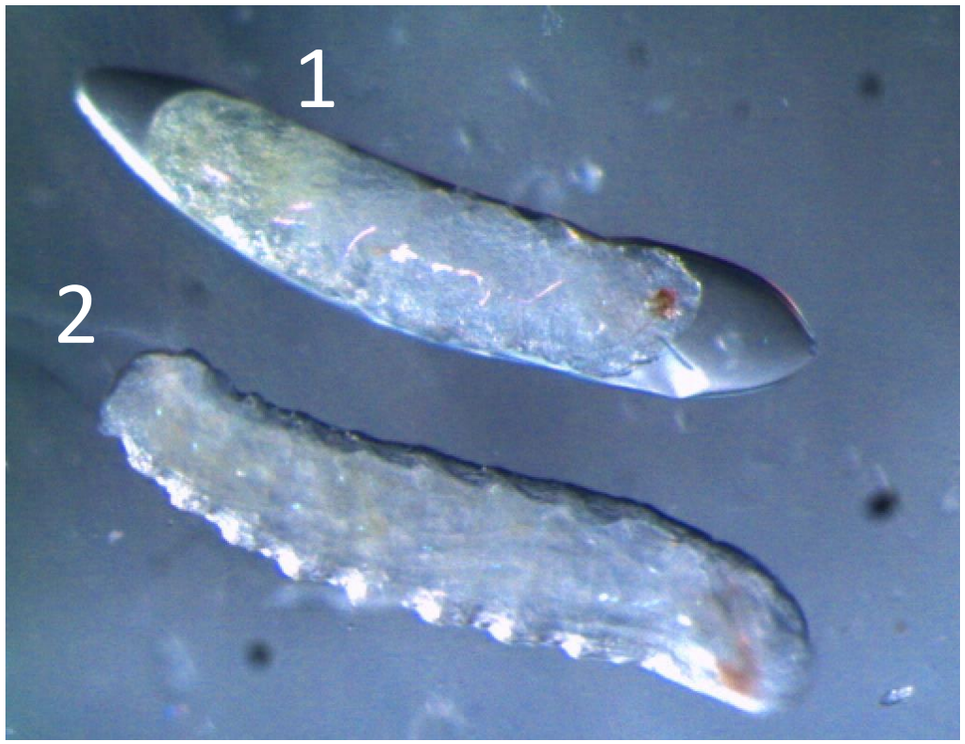
AGAATCCCAGCATATTTTCGTGG

N P S I F S W

Ccprd NPSIFSWE
Ccprd +P +FSWE
Ccprd SPGMFSWE

B

C



D

Cc-prd1

```
Wt AGGGTCGCGTCAATCAATTAGGTGGTGTGTTCATAAA wt
   AGGGTCGCGTCAAT-------GTGGTGTGTTCATAAA -7
   AGG------------------TGGTGTGTTCATAAA -19
   AGGGTCGCGTCAATCAATT--GTGGTGTGTTCATAAA -2
   AGGGTCGCGT----------------GTGTTCATAAA -16
```

Fig. 3 in green PRD-pax6 domain

Fig. 3: CRISP/Cas9 targeting of *Cc-paired.* A) A scheme of the Cc-prd gene and 2 sgRNAs targeting conserved Drosophila PRD in Ceratitis. B) Comparison of wt and injected embryos, showing shorted lenght and a reduction in the segment number. C) 2 injected larvae: larvae 1 shows incomplete/malformared development with the respect of larvae appeaaring wild type. D) Sequences of gene edited Cc-prd alleles from G0 larvae injected with RNP containing Cc-prd1 sgRNA.

Supplementary data: Tab 1

Survival rate after Cas9 alone injections

Cas9 µg/µl	Injected embryos	Larvae/ Embryos	Pupae/ Larvae	Adults/Pupae	Adults/ Embryos
0	114	0.56 (64/114)	0.84 (54/64)	0.66 (36/54)	0.31 (36/114)
0.9	150	0.59 (88/114)	0.37 (33/88)	0.82 (27/33)	0.18 (27/150)
1.8	110	0.57 (63/110)	0.49 (31/63)	0.68 (21/31)	0.19 (21/110)
3,6	134	0.51 (69/134)	0.41 (28/69)	0.71 (20/28)	0.15 (20/134)

Supplementary data: Tab 2

G0
we somatic CRISPR mutations

Cas9 µg/µl	sgRNA 0.2 µg/µl	Injected embryos	Larvae/embryos	Pupae/larvae	Adults/pupae	somatic mosaics	Adults/embryos surv
1.8	we-g1	240	56% (134/240)	33% (34/103*)	18% (6/34)	50% (**3**/6)	3% (6/209)
1.8	we-g2	180	44% (80/180)	36% (29/80)	83% (24/29)	4% (**1**/24)	13% (**24**/180)
1.8	we-g1/we-g2	240	43% (104/240)	44% (44/99*)	95% (42/44)	23% (10/42)	17% (42/240)

In bold adult flies, which were crossed to produce G1.

Supplementary data: Tab 3

G1 Germline transmission of CRISPR-induced *we* mutations with crosses of G0 single flies

Line	Sex of the founder	*we* flies	*we*$^{+}$ flies
we-g1#1	F$^{\$}$	0	0
we-g1#2	M$^{\$}$	6	0
we-g1#3	M$^{\$}$	10*	12
we-g2#1	F	0	55
we-g2#2	M	0	0
we-g2#3	M	53**	34
we-g2#4	M	113**	2
we-g2#5	M	1	68
we-g2#6	M$^{\$}$	14	10

$^{\$}$mosaics

*= 3 alleles sequenced

**= 2 alleles sequenced

Supplementary data: Tab 4

Injections data of *we*-g3 and *we*-g2+g3

Cas9 µg/µl	sgRNA 0.2 µg/µl	Injected embryos	Larvae/embryos	Pupae/larvae	Adults/pupae	somatic mosaics	Adults/embryos surv
1.8	we-g3	200	46% (92/200)	34% (31/92)	64% (20/31)	0	10% (**20**/200)
1.8	*we*-g2+g3	250	49% (123/250)	53% (58/110*)	65% (38/58)	0	16% (**38**/237)

Supplementary data: Tab 5

G1 Germline transmission of CRISPR-induced *we* mutations from pooled G0 injected flies

sgRNAs	G0 injected parents	we^{CRISPR} flies	we^{+} flies	Total adult G1 flies
we-g3	**8**M x **12**F	2*	24	26
we-g2+g3	**18**M x **20**F	3**	181	184

*= 4 *we*-g3 mutant alleles sequenced

**= 5 *we*-g2+g3 mutant alleles sequenced

Supplementary data: Tab 6

Lethality of paired CAS9 editing

Cas9 µg/µl	sgRNA µg/µl	embryos	larvae/embyos	pupae/larvae	adults/pupae	adults/embryos
0	0	172	46% (80/172)	74% (59/80)	97% (57/59)	33% (57/172)
1.8	prd1 0.2	164	35% (58/164)	17% (10/58)	80% (8/10)	5% (8/164)
1.8	prd1 0.2	244	24% (58/244)	50% (29/58)	96% (28/29)	11% (28/244)
1.8	prd2 0.2	150	26% (40/150)	43% (17/40)	88% (15/17)	10% (15/150)

CRISPR-Cas9 targeted disruption of the *yellow* ortholog in the housefly identifies the *brown body* locus

Svenia D. Heinze[1†], Tea Kohlbrenner[1†], Domenica Ippolito[1], Angela Meccariello[2], Alexa Burger[1], Christian Mosimann[1], Giuseppe Saccone[2], and Daniel Bopp[1*]

[1]Institute of Molecular Life Sciences, University of Zürich, Zürich 8057, Switzerland

[2]Department of Biology, University of Naples Federico II, 80126, Napoli, Italy

[†]these authors contributed equally to this work

Author for correspondence: Daniel Bopp
Institute of Molecular Life Sciences, University of Zürich
Winterthurerstrasse 190
8057 Zürich
Switzerland
Tel: +41 44 635 48 69
FAX: +41 44 635 68 23
email: daniel.bopp@imls.uzh.ch

running head: the *yellow* locus in the housefly

keywords: CRISPR-Cas9, *brown body*, *yellow*, *Musca domestica, genetics, genome editing,*

ABSTRACT

The classic *brown body* (*bwb*) mutation in the housefly *Musca domestica* impairs normal melanization of the adult cuticle. In *Drosophila melanogaster*, a reminiscent pigmentation defect results from mutations in the *yellow* gene that encodes the dopachrome conversion enzyme (DCE). Here, we demonstrate that the *bwb* locus structurally and functionally represents the *yellow* ortholog of *Musca domestica*, *MdY*. In Musca strains with *bwb* phenotype, we identified two mutant *MdY* alleles that contain reading frame-disrupting lesions predicted to result in premature truncation of the MdY open reading frame. To independently confirm that mutations in *MdY* correspond to the *bwb* locus, we targeted wildtype *MdY* by injecting preassembled sgRNA and Cas9 ribonucleoprotein complexes against the MdY locus into early syncytial embryos. We successfully recovered new mutant *MdY* alleles, and non-complementation of predicted frame-shift *MdY* alleles and the original *bwb* alleles confirmed that our Cas9-induced *MdY* mutations generated allelic variants of the same locus. We further found evidence for CRISPR-mediated interchromosomal recombination between wildtype and mutant *bwb* alleles. Our work resolves the molecular identity of the *bwb* mutation in *Musca domestica* as lesions in the *MdY* gene and establishes the conserved involvement of DCE activity in pigment formation in Musca. Our results further establish the feasibility and impact of Cas9-mediated genome editing in the *Musca* model.

Introduction

The so-far unidentified *brown body* (*bwb*) locus in the housefly *Musca domestica* was named after a recessive loss-of-function phenotype in which the adult cuticle manifests in a brown color rather than the wildtype black pigmentation **(Fig. 1a).** The absence of black coloration in mutant *Musca* has been proposed to result from impaired synthesis and incorporation of the black pigment melanin during pupal stages [1]. In insects, melanization of the cuticle is widely common and contributes to the diverse coloration patterns that are the most visible features of the outer morphology. Most insights into the pathway that produces and incorporates melanin into the insect cuticle comes from studies in *Drosophila melanogaster* [2] [3] [4]. The melanin pathway starts with conversion of tyrosine to Dihydroxyphenylalanine (DOPA) by *tyrosine hydroxylase* (TH). DOPA in turn is converted to dopamine by *dopa carboxylase (DDC)*. Both substrates are used as precursors for production of black melanin. In *Drosophila*, the dopachrome conversion enzyme (DCE) catalyzes steps downstream of TH and DCC in melanin production; however, whether DCE is involved in only in DOPA, only in dopamine conversion, or in both, remains unclear. Loss of function in the *yellow* gene (*y*) that encodes *Drosophila* DCE causes a lack of melanin incorporation and results in a yellowish overall appearance of the cuticle. Similar phenotypes have been observed in other insects such as the lepidopteran *Bombyx mori, Papilio xuthus,* and the coleopteran *Tribolium castaneum* [5] [6] [7] [8] [9]. Loss of activity of the corresponding *yellow* orthologs in these species drastically lowers the synthesis of melanin, causing regions of the body that are normally black to display a lighter coloration.

In recessive *bwb*-mutant houseflies, the loss of normal black coloration affects all body parts. This phenotype closely resembles the phenotype observed in *yellow*-mutant *Drosophila*. As no other gene in the network of melanin genes is known to manifest this phenotype, we sought to investigate whether the *bwb* gene in the housefly is the structural and functional homolog of the *yellow* (*DCE*) gene in *Drosophila*. In addition to the correspondence in phenotypes, the *bwb* locus has been mapped to Chromosome III in Musca which corresponds to Muller element A, the X chromosome in Drosophila that harbors the *y* locus [10,11].

Here, we identified the locus mutated in *bwb* as the *Musca domestica* ortholog of the *Drosophila* gene *yellow* and refer to the gene as *MdY.* The *MdY* gene shows a high degree of similarity at the level of both protein sequence and gene structure. We find three mutant alleles of the *MdY* gene in *bwb*-mutant flies: the alleles MdY^{a2} and MdY^{b} feature sequence disruptions of the coding sequence, while a third allele MdY^{a1} is a compound allele featuring the coding

lesion of MdY^{a2} and an additional 1.5 kb sequence insertion in the *5' UTR*. Using CRISPR-Cas9, we generated a series of *de novo* loss-of-function alleles of *MdY* that all fail to complement original *bwb* alleles; these experiments mark, to our knowledge, the first report of the successful application of Cas9-based mutagenesis in *Musca domestica*. Altogether, we conclude that the *bwb* phenotype in Musca is caused by a lack of DCE activity normally provided by the *MdY* gene. Our findings clarify the molecular lesions in the classic *bwb* mutation and further underline the notion that *yellow* plays a conserved role in the melanin production pathway in dipteran species.

Results

1. Characterization of the *yellow* ortholog in *Musca domestica*

Based on phenotype resemblance and mapping position, we hypothesized that *bwb* affects the so-far undescribed DCE ortholog in *Musca domestica*. To identify sequences homologous to the *Drosophila yellow* gene in Musca, we performed BLAST searches against the recently published genome of the multi-marked *aabys* strain that shows the *bwb* phenotype [12]. We recovered an annotated mRNA sequence (NCBI RefSeq *XM_011292650.1*) with a high degree of sequence similarity to *Drosophila yellow* (**Supplementary data Fig. 1**). Annotation of the *aabys* genome called this gene a pseudogene based on the lack of an intact open reading frame. Indeed, we detected a frame-shift starting 67 codons downstream of the first AUG start codon (**Fig. 1b**) [12]. Hence, the molecular nature of this allele already suggested that the *bwb*-mutant *aabys* strain carries a non-functional *yellow* variant. The mutated putative Musca *yellow* gene, which we named *MdY*, is located on *Scaffold18750* (502 kb) and is composed of two exons separated by a 35.6 kb-spanning intron (**Fig. 1b**). A *Musca* homolog of the *acheate* (*ac*) gene is present on the same scaffold separated by 156 kb, revealing a relatively close linkage of *yellow* and *ac* orthologs in *Musca.* This coupling is conserved in *Drosophila melanogaster* and sibling species, and previous work proposed that evolutionary variation of the *y-ac* region is reduced due to the selective fixation of one or more advantageous mutations in this region [13]. We next isolated *MdY* sequences from a wildtype strain *Siat* that displays a normal melanization pattern; the *MdY* sequence in *Siat* has an intact ORF of 522 amino acids, markedly lacking the 4 bp insertion that causes a frame-shift in the *aabys*-derived allele (**Fig. 1b and d**). The predicted MdY protein shares a high level of

similarity with YELLOW proteins of other dipterans (Supplementary Figure 2a): as expected, *MdY* shares the highest level of identity (88%) with the predicted YELLOW protein in *Stomoxys calcitrans,* a close relative of the same family *Muscidae* (Supplementary data Fig. 2b).

To extend our analysis of mutant *bwb* alleles, we included the M^{III} strain that carries the male determining factor on Chromosome III tightly linked to wildtype alleles of *bwb* and *pointed wings* (*pw*), while females of this strain are homozygous mutant for *bwb* and *pw* (**Fig. 1a**). Analysis of *MdY* sequences isolated from M^{III} females unveiled that again the *bwb* allele that contains a 4 bp insertion at amino acid position 67. Unexpectedly, we also reproducibly detected a second mutant *MdY* allele in this strain, a nonsense mutation at amino acid position 65 (**Fig. 1d**). Therefore, the M^{III} laboratory strain carries two different loss-of-function alleles of *MdY*.

When mapping wildtype genomic *MdY* fragments against the corresponding *aabys* genomic sequences that harbor the 4 bp insertion in the coding sequence, we additionally identified a structural difference in the *5′ UTR* region upstream of the putative start codon (**Fig 1b**). Sequence analysis revealed the presence of a 1.5 kb insertion in the *aabys*-derived allele that is absent in the wildtype *Siat* strain (**Fig 1b and c**). A BLAST search against the Musca genome shows that this *aabys*-specific insertion shares 77% of sequence identity to an incomplete gene complement of the nicotinic acetylcholine receptor subunit-encoding *Mdalpha2* [14]. This 5' UTR insertion in the mutant *MdY* allele of the *aabys* strain consequently represents a third, compound mutant allele of *MdY*.

Altogether, we identified two distinct coding-frame alleles in *bwb*-mutant *Musca domestica* strains, and a third compound lesion that introduced a 1.5kb insertion into the *5' UTR* of one of the putatively inactive *MdY* loci. We refer to the original *aabys*-derived compound allele as MdY^{a1}, and to the two M^{III}-derived alleles as MdY^{a2} and MdY^{b} (**Fig. 1d, Supplementary data Fig. 3**). From our analysis of these mutants, we conclude that the *bwb* phenotype is associated with nonsense mutations in the *MdY* gene. These observations support the notion that the housefly DCE homolog is involved in the melanin production pathway.

2. CRISPR-Cas9-mediated disruption of *MdY* confirms causative association with *bwb*.

We next performed targeted disruption of the wildtype *MdY* locus to confirm its predicted role in melanization of the housefly cuticle and to corroborate whether the *bwb* mutations are indeed loss-of-function alleles of *MdY*. To this end, we selected two target sites

in the second exon of *MdY* for non-homologous end joining (NHEJ)-mediated disruption by the CRISPR-Cas9 system. Preassembled ribonucleoprotein complexes (RNPs) composed of purified Cas9 protein loaded with two different sgRNAs (*sgY2* and *sgY3*) were injected into early syncytial embryos. Both *sgY2* and *sgY3* target two different sites in coding exon 2 *separated by* 340 bp, (**Fig 2a**). As host strain, we used the M^{III} strain that carries the male determining factor on the chromosome III linked to wildtype alleles of *bwb* and *pw*. Females of this strain are homozygous mutant for both markers and are brown coloured with pointed-wings, while males are heterozygous and phenotypically wildtype for both markers ([15,16]) (**Fig. 2b**). This genetic background facilitates detection of possible somatic effects of *MdY* disruption in males that carry only one wildtype allele of *bwb* (**Fig. 1a**). The two sgRNAs were preloaded separately on purified recombinant Cas9 protein (A.M. and G.S. unpublished results) and a 1:1 mix was micro-injected into 1 h old embryos of the M^{III} strain[17]. Of 2565 embryos injected with a 1:1 mix of solubilized Cas9 RNPs containing *sgY2*, and *sgY3*, we recovered 188 surviving adult houseflies. While 106 of these adults were males, none of the males displayed any patches of brown coloration, indicating that our targeting procedure does not introduce significant somatic mutation mosaicism. We proceeded with crossing these injected G0 males to *bwb*-mutant females of the same M^{III} strain (2 males and 6 females per cross) to screen for possible germline effects. Screening the F1 generation, we observed 17% of crosses that produced brown-colored M^{III} pw^{+} males that can only arise from paternal transmission of a mutant *bwb* allele (**Tab. 1, Fig. 2C**). This observation reveals that our Cas9 RNP-mediated mutagenesis protocol introduced *bwb* mutations by targeting the *MdY* gene in germ cells of the syncytial embryo. Of note, the injected males vary greatly in the proportion of potentially *MdY*-disrupted F1 individuals they sire (**Tab. 1),** revealing variable germline mosaicism resulting from Cas9 mutagenesis akin to observations in other model organisms [17].

To characterize the putative lesions induced in the *MdY* gene by sgRNAs *sgY2* and *sgY3*, we PCR-isolated genomic sequences of exon 2 from brown-bodied M^{III} pw^{+} F1 males by PCR and examined the targeted sites in sub-cloned and Sanger-sequenced fragments using CrispRVariants [18]. Overall, the position and extent of the induced lesions vary between individual lines but show a clear preference for lesions at the *sgY3* target site (**Table 2, Supplementary Fig. 4**). In seven lines (*MdY#2, MdY#9, MdY#10, MdY#13, MdY#16, MdY#33, MdY#36,* and *MdY#38*) we found small indels exclusively in the target site of *sgY3* (Table 2). In contrast, only line *MdY#14* carries an allele of *MdY* with a lesion (2 bp deletion) in the *sgY2* site (**Table 2, Supplementary Fig. 4**). In line *MdY#19*, we found evidence that both sites were targeted (Supplementary Fig. 4). The absence of a deletion spanning the two sites suggests that

two dsDNA break events in our recovered alleles must have occurred at different times or with different kinetics, allowing the cellular repair system to independently join the breaks. In contrast, line *MdY#40* carries a large deletion of 1038 bp which removes both target sites and sequences downstream of *sgY2*; *MdY#40* likely had both sites simultaneously targeted by Cas9 and repaired to result in a larger deletion. Taken together, these data establish functional CRISPR-Cas9-mediated mutagenesis targeting the *MdY* locus in *Musca domestica* using *in vitro*-assembled Cas9-sgRNA RNPs.

To test for additional complementation, we crossed brown males of lines *MdY#10* which carries a 15 bp deletion in the *sgY3* site of *MdY* and *MdY#16* containing a frame-shifting 10 bp deletion at the same site with brown females of the M^{III} strain carrying the MdY^{a2} and MdY^{b} alleles and the *aabys* strain, homozygous for the MdY^{a1} allele. In all crosses, all progeny displayed *bwb* phenotype. Lack of melanization in these animals is consistent with our hypothesis that *MdY* is required for proper pigmentation of the cuticle. Furthermore, non-complementation of the Cas9-induced *MdY* alleles with all mutant *bwb* alleles of the *aabys* and M^{III} strains confirms our initial hypothesis that *bwb* corresponds to the *yellow* ortholog of *Musca domestica.*

Two lines, *MdY#29* and *MdY#40*, were of particular interest, as we did not detect any sequence modifications at the two target sites in exon 2 (**Table 2**). Nonetheless, both these lines produced both brown M^{III} pw^{+} males and also *pw* females with normal melanization. The reciprocity of these sex-specific phenotypes suggests Cas9-mediated double-strand breaks in *MdY* induced an intragenic recombination event between the bwb^{+} allele on the M^{III} chromosome and the mutant *bwb* allele on the corresponding homolog. This event may have created a recombinant *MdY* allele with abolished activity on the chromosome that contains the *M* factor. In line with this hypothesis, we found brown males in line *MdY#29* that are unaffected at the two Cas9 target sites, but homozygous for the MdY^{b} signature (translational stop at position 65) (**Fig. 3a**). These sequences are likely the products of a reciprocal recombination event that may also have resulted in a reconstitution of a wildtype *MdY* allele on the non-*M* chromosome in females. Indeed, we found that melanized females are heterozygous for the MdY^{b} signature in exon 1. Moreover, in the mutant *bwb* males, we detected heterozygosity for allele-specific polymorphisms just downstream of the target site *sgY3* that correspond to the paternal genotype (**Fig. 3a**). We interpret these observations as evidence for Cas9-mediated DNA double-strand breaks at, or upstream of, the *sgY3* site that induced an intragenic recombination event between the wildtype *MdY* allele and the MdY^{b} allele *in trans* in germ cells of Cas9 RNP-injected heterozygous males (**Fig. 3b**). This data indicates that Cas9-

mediated mutagenesis in *Musca domestica* can lead to break-point-guided recombination in injected germ cells.

Discussion

Our work reveals that the classic *bwb* phenotype in houseflies is caused by mutations in the *Musca* homolog of the *DCE* gene, which encodes the enzyme that in *Drosophila* has been implicated in the process that converts DOPA and/or dopamine into black melanin. Functional studies in *Drosophila* have provided evidence that the DCE-encoding gene *yellow* is involved in global body pigmentation [19], whereas in the coleopteran *Tribolium castaneum* the loss of *yellow* only affects pigmentation of the hindwing [5]. Also, in the hemimetabolous *Oncopeltus fasciatus*, silencing of *yellow* affects only specific body parts such as abdomen and hindwings [20]. These observations led to the proposition that *yellow* belongs to a network of melanin synthesis genes which by differential deployment can generate a wide range of colors and spatial patterns [20]. Here, we present genetic evidence that, as in *Drosophila*, the *Musca* DCE homolog *MdY* is required for melanin production in the whole body. We base our conclusion of two major lines of evidence: First, complete absence of melanization found in two different *bwb* strains (M^{III} and *aabys*) correlates with homozygosity for predicted nonsense alleles of *MdY*. In these strains, we identified three mutant alleles MdY^{a1}, MdY^{a2} and MdY^{b}, all of which result in premature stop codons. We note that MdY^{a1} and MdY^{a2} only differ with regard to a 1.5 kb insertion which is present specifically in the *5′ UTR* of MdY^{a1}. It is thus possible that these two alleles, both of which carry the same 4 bp insertion downstream of codon 67, have a common origin and that the 1.5 kb insertion has been acquired or lost later during a secondary mutational event. Second, we generated a set of new *MdY* alleles by adapting the CRISPR-Cas9 mutagenesis for disrupting the coding region in *MdY* exon 2. All of the new alleles have confined lesions at least in one of the two targeted sites. The majority of these mutations are deletions that remove parts of the coding sequence and generate frame-shifts. We hence consider these allelic variants to be enzymatically non-functional, if not null alleles of *MdY*. These Cas9-based *MdY* mutants fail to complement, and thus behave allelic to, the previously identified *MdY* variants found in *bwb*-mutant backgrounds.

The CRISPR-/Cas9 protocol using solubilized RNPs used in our study appears to be highly effective in the *Musca* germ line given that at least 1 of 6 RNP-injected males

transmitted an *MdY* allele with a defined lesion at one of the targeted sites. Nonetheless, none of these F0 males displayed patches of non-melanized cuticle indicative of somatic mutation mosaicism. In *Drosophila*, *y* mutations behave cell-autonomously and even small mutant clones can be readily detected in the cuticle: when Cas9 mRNA was injected with sgRNA targeting *y* in *Drosophila* has strong somatic effects (86%) not only in hemizygous males with one wildtype target, but even in females with two wildtype copies of *y* [21]. This study pointed out that the efficiency of inducing somatic clones not only depends on the concentration of sgRNA injected, but more importantly on the selection of the site in the *y* gene that was targeted. It is thus conceivable that the sgRNAs used in our work were able to efficiently disrupt the gene in the germline cells but not in somatic cells. Germline transmission is a prerequisite for the investigation of any new mutation generated and, in this regard, our main objective was to recover fertile adults transmitting mutant *MdY* alleles. The lack of somatic effects in G_0 adults can be a beneficial feature to avoid sterility that may be inflicted by the presence of mutant somatic tissue.

Recent work proposed that the CRISPR-Cas9 system can be used for genetic mapping by inducing targeted recombination events in meiotic and mitotic cells [22]. Our finding of an intragenic recombination event in line *MdY#29* suggests that Cas9 induced breaks allows recombination between homologs in the germ line of housefly males. This observation suggests a possible Cas9-mediated strategy for male meiotic mapping in future studies. In addition, homologous recombination (HR)-mediated repair of Cas9 induced double-strand breaks in *Musca* offers the opportunity to attempt template-based editing of genomic sequences for targeted knock-ins in the future.

To our knowledge, our study is the first report showing that Cas9 can be effectively deployed for NHEJ mediated-disruption of genes in *Musca domestica*, an important addition to the toolkit of molecular methods that have already been established in *Musca* for gene function analysis. Together with transient RNAi-based gene silencing [23,24] and stable germline transformation [16,25], this new genome editing system provides a means to investigate evolutionary diversification of developmental pathways such as the polymorphic sex determination system of the housefly [15,26]. Our successful attempt promises that this genome editing technology can be used in the housefly to study the function of any candidate gene of interest.

Methods

Rearing of houseflies

Rearing of larvae and adult flies has been described in [27,28]. Since low density of larvae on standard medium can cause substantial decrease in survival rates, we reared injected embryos and the surviving larvae on porcine manure. To dispose of mites and other parasites and to avoid contamination with eggs or larvae from wild-type populations, manure was stored at -70° C for at least two weeks prior to use.

Strains of *Musca domestica*

(1) wildtype strain was collected in Siat, Switzerland: females X/X*; bwb^+/bwb^+* and males X/Y*; bwb^+/bwb^+*

(2) multi-marked strain *aabys*: females X/X*; ac/ac; ar/ar; bwb/bwb; ye/ye; snp/snp* and males X/Y *ac/ac; ar/ar; bwb/bwb; ye/ye; snp/snp*; [12]

(3) autosomal M^{III} strain: females X/X*; pw, bwb, w/ pw, bwb, w* and males X/X*; pw^+, M^{III}, bwb^+, w / pw, bwb, w* [16]

Genomic DNA extraction

For genomic DNA extraction a single fly was collected in a 1.5 ml tube, frozen in liquid nitrogen and ground in 1 ml of extraction buffer (0.1 M Tris-HCl, pH 9; 0.1 M EDTA; 1% SDS and 1% of DMDC added freshly). After 30 min incubation at 70 °C, 140 µl of 8M potassium acetate was added and sample was gently inverted and incubated for 30 min on ice. After 15 min of centrifugation at 4°C at 13'000 rpm, supernatant was transferred to a new tube, and 550 µl of isopropanol was added. The mixture was centrifuged at RT for 5 min at 14'000 rpm and the supernatant was removed. The pellet was washed with 500 µl of 70% EtOH (-20 °C) and centrifuged at RT for 2 min at 14'000 rpm. The DNA pellet was finally dissolved in 30 or 50 µl of 10 mM Tris and 1 µl of RNaseA (10mg/ml) to remove RNA. Amplifications for sequence analysis were performed with following primers. For exon 1 we used forward Y-ORF-F3 (5'-TGCTGTGGACATTGGCAAGA -3') and reverse RE1: (5'-TCTCATTCACATCCACACCGT-3').

For exon2 we used forward FE4 (5'-CAGGTATACCAGCCACATTGA-3') and reverse Y-ORF-R5 (5'-CTAATGATGGGCGGATGTGGA-3').

For insertion we used flanking primer forward Y-GAP1-F1 (5'-GGCCGAAGTGAGACAGAGAA-3') and Y-EXON1-R (5'-CTAGTGGCGAAAAACCATTAA-3').

sgRNA synthesis and RNP complex assembly

sgRNA were designed using MultiTargeter Website (http://www.multicrispr.net). Possible OFF-target sites were excluded with the program Cas-OFFinder (http://www.rgenome.net/cas-offinder/) and by directly BLASTing selected sequences against the published housefly genome sequence [12]. Following templates for sgRNA production were generated:

sgY2: 5'-GAAATTAATACGACTCACTATA GGCTTTGTCGCCCATTCGTT GTTTTAGAGCTAGAAATAGC-3'

sgY3: 5' -GAAATTAATACGACTCACTATA GGCATAGGGACAGGGGTTGG GTTTTAGAGCTAGAAATAGC-3'

Common sgReverse (PAGE purified): 5'AAAAGCACCGACTCGGTGCCACTTTTTCAA GTTGATGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC AAC

For synthesis of sgRNA we followed the instructions of Megatranscript T7 kit (Ambion) using 400 ng of target template with a 5' flanking T7 promoter as starting material. After RNA synthesis template was removed by incubating with TurboDNase (Mmessage Mmachine T7 Ultra Kit, Ambion) for 15 min at 37 °C.

Cas9 was expressed as an His-tagged protein and purified from bacteria (A.M. and G.S. unpublished results). The injection cocktail was prepared by mixing 1.5 µl purified Cas9 protein (9 mg/ml) with 2µl of sgRNA in 1.36 µl of 2M KCl in a total of 10µl. Prior injection the mix was incubated for 10 min at 37°C [17]. For injection we prepared a 1:1 mix of *sgY3*-preloaded Cas9 RNPs and *sgY2*-preloaded Cas9 RNPs.

Microinjection of sgRNA-Cas9 complexes

Embryos of the M^{III} host strain were collected 1 hour after egg lay and chorion membrane was removed by incubating embryos in 3% sodium hypochlorite solution (NaOCl) for 1.5 min. Dechorionated embryos were then rinsed thoroughly with water and Ringer's solution. Embryos were aligned on a cover slip with posterior ends pointing to injection site, dehydrated for 4 min in a silicagel chamber and then covered with 3S /10S (1:4) Voltalef oil (Prolabo). A glass needle was filled with the preloaded sgRNA-Cas9 mix which was injected into the posterior end of 0 to 1 hour old embryos. After injection excess Voltalef oil was carefully removed and cover slip were put on an agar plate overnight at 25°C. Surviving larvae were

transferred after 24 hours to small beaker filled with porcine manure. G_0 male individuals were collected shortly after eclosing and crossed with untreated virgin females of the M^{III} strain.

Genomic analysis of CRISPR-Cas9 mediated lesions in *MdY*

To examine for the presence of lesions in *MdY* caused by NHEJ, genomic DNA of *bwb* mutant F1 males was extracted following the protocol described above. The region encompassing the two target sites was amplified with the following primer:
forward primers FE3: TCTGGCAAACCACAACAAGT or F4
CAGGTATACCAGCCACATTGA; reverse primers Y-ORF-R1:
GACGAATGCCAACAACCCAC or Y-ORF-R5 CTAATGATGGGCGGATGTGGA
PCR products were purified with the Wizard® Genomic DNA Purification Kit (Promega), subcloned in the pGEM®-T Easy Vector (Promega) and sent for Sanger sequencing (GATA BIOTECH). CrispRVariants and the generation of panel plots was performed from primary sequencing data as previously described. [17,18]

References

1. Malacrida A, Grigolo A, Gasperi G, Sacchi L. [Activity of the enzyme dopa-decarboxylase in brown body (bwb) mutants of Musca domestica L]. Boll Soc Ital Biol Sper. 1974;50: 1566–1569.

2. Wright TR. The genetic and molecular organization of the dense cluster of functionally related, vital genes in the DOPA decarboxylase region of the Drosophila melanogaster genome. Results Probl Cell Differ. 1987;14: 95–120.

3. Wittkopp PJ, Williams BL, Selegue JE, Carroll SB. Drosophila pigmentation evolution: divergent genotypes underlying convergent phenotypes. Proc Natl Acad Sci USA. National Acad Sciences; 2003;100: 1808–1813. doi:10.1073/pnas.0336368100

4. Wittkopp PJ, Stewart EE, Arnold LL, Neidert AH, Haerum BK, Thompson EM, et al. Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in Drosophila. Science. American Association for the Advancement of Science; 2009;326: 540–544. doi:10.1126/science.1176980

5. Arakane Y, Dittmer NT, Tomoyasu Y, Kramer KJ, Muthukrishnan S, Beeman RW, et al. Identification, mRNA expression and functional analysis of several yellow family genes in Tribolium castaneum. Insect Biochem Mol Biol. 2010;40: 259–266. doi:10.1016/j.ibmb.2010.01.012

6. Futahashi R, Fujiwara H. Regulation of 20-hydroxyecdysone on the larval pigmentation and the expression of melanin synthesis enzymes and yellow gene of the swallowtail butterfly, Papilio xuthus. Insect Biochem Mol Biol. 2007;37: 855–864. doi:10.1016/j.ibmb.2007.02.014

7. Futahashi R, Sato J, Meng Y, Okamoto S, Daimon T, Yamamoto K, et al. yellow and ebony are the responsible genes for the larval color mutants of the silkworm Bombyx mori. Genetics. Genetics; 2008;180: 1995–2005. doi:10.1534/genetics.108.096388

8. Ito K, Katsuma S, Yamamoto K, Kadono-Okuda K, Mita K, Shimada T. Yellow-e determines the color pattern of larval head and tail spots of the silkworm Bombyx mori. J Biol Chem. American Society for Biochemistry and Molecular Biology; 2010;285: 5624–5629. doi:10.1074/jbc.M109.035741

9. Wittkopp PJ, Vaccaro K, Carroll SB. Evolution of yellow gene regulation and pigmentation in Drosophila. Curr Biol. 2002;12: 1547–1556.

10. Weller GL, Foster GG. Genetic maps of the sheep blowfly Lucilia cuprina: linkage-group correlations with other dipteran genera. Genome. 1993;36: 495–506.

11. Meisel RP, Scott JG, Clark AG. Transcriptome Differences between Alternative Sex Determining Genotypes in the House Fly, Musca domestica. Genome Biol Evol. Oxford University Press; 2015;: evv128. doi:10.1093/gbe/evv128

12. Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, et al. Genome of the house fly, Musca domestica L., a global vector of diseases with adaptations to a septic environment. 2014;: 1–16. doi:10.1186/s13059-014-0466-3

13. Begun DJ, Aquadro CF. Molecular population genetics of the distal portion of the X chromosome in Drosophila: evidence for genetic hitchhiking of the yellow-achaete region. Genetics. 1991;129: 1147–1158.

14. Gao J-R, Deacutis JM, Scott JG. Characterization of the nicotinic acetylcholine receptor subunit gene Mdalpha2 from the house fly, Musca domestica. Arch Insect Biochem Physiol. Wiley Subscription Services, Inc., A Wiley Company; 2007;64: 30–42. doi:10.1002/arch.20158

15. Dübendorfer A, Hediger M, Burghardt G, Bopp D. Musca domestica, a window on the evolution of sex-determining mechanisms in insects. Int J Dev Biol. 2002;46: 75–79.

16. Hediger M, Henggeler C, Meier N, Perez R, Saccone G, Bopp D. Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. Genetics. 2010;184: 155–170. doi:10.1534/genetics.109.109249

17. Burger A, Lindsay H, Felker A, Hess C, Anders C, Chiavacci E, et al. Maximizing mutagenesis with solubilized CRISPR-Cas9 ribonucleoprotein complexes. Development. Oxford University Press for The Company of Biologists Limited; 2016;143: 2025–2037. doi:10.1242/dev.134809

18. Lindsay H, Burger A, Biyong B, Felker A, Hess C, Zaugg J, et al. CrispRVariants charts the mutation spectrum of genome engineering experiments. Nature Biotechnology. Nature Research; 2016;34: 701–702. doi:10.1038/nbt.3628

19. Wittkopp PJ, True JR, Carroll SB. Reciprocal functions of the Drosophila yellow and ebony proteins in the development and evolution of pigment patterns. Development. 2002;129: 1849–1858.

20. Liu J, Lemonds TR, Marden JH, Popadić A. A Pathway Analysis of Melanin Patterning in a Hemimetabolous Insect. Genetics. Genetics; 2016;203: 403–413. doi:10.1534/genetics.115.186684

21. Bassett AR, Tibbit C, Ponting CP, Liu J-L. Highly efficient targeted mutagenesis of Drosophila with the CRISPR/Cas9 system. Cell Rep. 2013;4: 220–228. doi:10.1016/j.celrep.2013.06.020

22. Sadhu MJ, Bloom JS, Day L, Kruglyak L. CRISPR-directed mitotic recombination enables genetic mapping without crosses. Science. American Association for the Advancement of Science; 2016;352: 1113–1116. doi:10.1126/science.aaf5124

23. Hediger M, Burghardt G, Siegenthaler C, Buser N, Hilfiker-Kleiner D, Dübendorfer A, et al. Sex determination in Drosophila melanogaster and Musca domestica converges at the level of the terminal regulator doublesex. Dev Genes Evol. 2004;214: 29–42. doi:10.1007/s00427-003-0372-2

24. McGregor AP, Shaw PJ, Hancock JM, Bopp D, Hediger M, Wratten NS, et al. Rapid restructuring of bicoid-dependent hunchback promoters within and between Dipteran species: implications for molecular coevolution. Evol Dev. 2001;3: 397–407.

25. Hediger M, Niessen M, Wimmer EA, Dübendorfer A, Bopp D. Genetic transformation of the housefly Musca domestica with the lepidopteran derived transposon piggyBac. Insect Mol Biol. 2001;10: 113–119.

26. Bopp D. About females and males: continuity and discontinuity in flies. J Genet. 2010;89: 315–323.

27. Hilfiker-Kleiner D, Dübendorfer A, Hilfiker A, Nöthiger R. Genetic control of sex determination in the germ line and soma of the housefly, Musca domestica. Development. 1994;120: 2531–2538.

28. Schmidt R, Hediger Niessen M, Nöthiger R, Dübendorfer A. The Mutation *masculinizer (man)* Defines a Sex-Determining Gene With Maternal and Zygotic Functions in *Musca domestica* L. Genetics. 1997;: 1–11.

Acknowledgements

We thank Claudia Brunner for technical assistance, Raymond Grunder and Daniel Christen for help with the rearing of housefly cultures. We are grateful to Elena Chiavacci for her help with the target design and processing of CRISPR sequence data. This work was supported by the Canton of Zürich, a Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (SNSF) professorship [PP00P3_139093] and a Marie Curie Career Integration Grant from the European Commission to C.M.; a University of Zurich URPP Translational Cancer Research Seed Grant to A.B. G.S. and A.M. were visiting researchers at University of Zurich

(November 2015), supported by University Federico II of Naples (International Exchange Program to GS and by the Biology PhD program of University Federico II of Naples to AM).

Authors contributions statement

DB. and GS conceived the experiments, SDH, TK, DI, AM. and AB. conducted the experiment(s). All of the authors discussed the data and helped manuscript preparation. DB, GS, and CM wrote the manuscript with intellectual input from all authors.

Competing financial interests statement

The authors declare no competing financial interests.

FIGURE LEGENDS

Figure 1. ***bwb* phenotype is associated with nonsense mutations in *MdY***

Phenotypes are displayed from left to right: multimarked *aabys* fly, *bwb* mutant female from M^{III} strain and *bwb* wildtype male from M^{III} strain. Genotypes are indicated; *bwb*: brown body, *pw*: pointed wings (notches along the edge of the wing, see arrow), *w*: white eyes. (B) schematic drawing of the *MdY* locus in the *bwb* mutant *aabys* strain and the Siat *bwb* wildtype strain. The aabys allele, MdY^{a1}, contains a 1.5 kb insertion in the 5' UTR and an additional 4 bp insertion in the ORF of exon 1. This frame-shift leads to a premature TAA stop in the 5' end of exon 2. The Siat *MdY* allele has an intact ORF (boxed yellow). (C) Genomic amplification with flanking primers Y-GAP1-F1 and Y-EXON1-R show that the 1.5 kb insertion is present in males and females of the *bwb* mutant aabys strain, but not in the *bwb* wildtype Siat flies (D) Affected part of the coding region (position 63 to 80) of the nonsense alleles of *MdY*. Deviations from the wildtype sequence (Siat) are marked in red and translational stops in bold. Location of intron is indicated with a triangle.

Figure 2. Strategy for CRISPR/Cas9 mediated disruption of *MdY*

(A) A schematic of the *MdY* gene showing the positions of the two target sites in exon 2. Sequences used for the design of the two sgRNAs, sgY3 and sgY2, are indicated. Both sequences are flanked by a PAM motif (in red) and separated by 343 bp. (B) Crossing scheme for screening mutational events affecting melanization. Injected G_0 males (with a mix of sgY3-CAS9 and sgY2-CAS9) are crossed with *bwb* females and F1 is examined for occurrence of *bwb* males. (C) Left a *bwb* mutant F1 male from line MdY#16 which is heterozygous for a CRIPSR induced 10 bp deletion in sgY3 over MdY^{b}. Right an unaffected *bwb* wildtype F1 male from the same line with the paternal genotype (MdY^{+} over MdY^{b})

Figure 3. Intragenic recombination in *MdY* mediated by CRISPR/Cas9

(A) Excerpts of chromatograms of exon 1 and 2 showing allele-specific polymorphisms. Mutant F1 *bwb* males are homozygous for the two variants in exon 1 (MdY^{b} genotype) amplified with primers Y-ORF-F3 and RE1, but heterozygous for the three variants in exon 2 like in the paternal wildtype *bwb* G_0 male amplified with primers FE4 Y-ORF-R5. Arrows point to the allele-specific polymorphisms examined. (B) This sequence analysis suggests that an intragenic recombination occurred downstream of the MdY^{b} specific TGA stop codon in

exon 1 and upstream of polymorphisms examined in exon 2. Locations of the SNPs between the two target sites (sgY3 ad sgY2) are indicated with short arrows.

Table 1. Lines with *bwb* males in F1 generation

From a total of 43 crosses with 2 injected G_0 males each, 14 lines with *bwb* mutant F1 males were recovered. MdY#29 and MdY#40 also produced bwb^+ wildtype females. The numbers of F1 flies with different phenotypes are shown for each line. Presence of M^{III} indicates a male phenotype.

Table 2. Overview of *MdY* lesions detected in *bwb* males

14 CRISPR lines are listed with lesions detected in the target sites, Δbwb sg3 and Δbwb sg2. The majority of mutations caused by NHEJ occurred in the Δbwb sg3 region ranging in size from 1 bp insertion to 174 bp deletions. In line MdY#38 sequence changes in and downstream of Δbwb sg3 but we were unable to determine the extent of this putative lesion (nd). In lines MdY#29 and MdY#40 none of the isolated sub-clones harbored a visible lesion in the region of the two target sites. Since both lines also produced recombinant *pw*, bwb^+ females, it is likely that the pw^+, *bwb* males are products from a recombinant event rather than from a NJEH induced lesion.

Highlights

- The *brown body* locus of Musca encodes the dopachrome conversion enzyme *yellow*
- CRISPR-Cas9 technology is highly efficient in generating new loss-of-function alleles in Musca
- CRISPR-Cas9 can be used for interchromosomal recombination in male houseflies

a

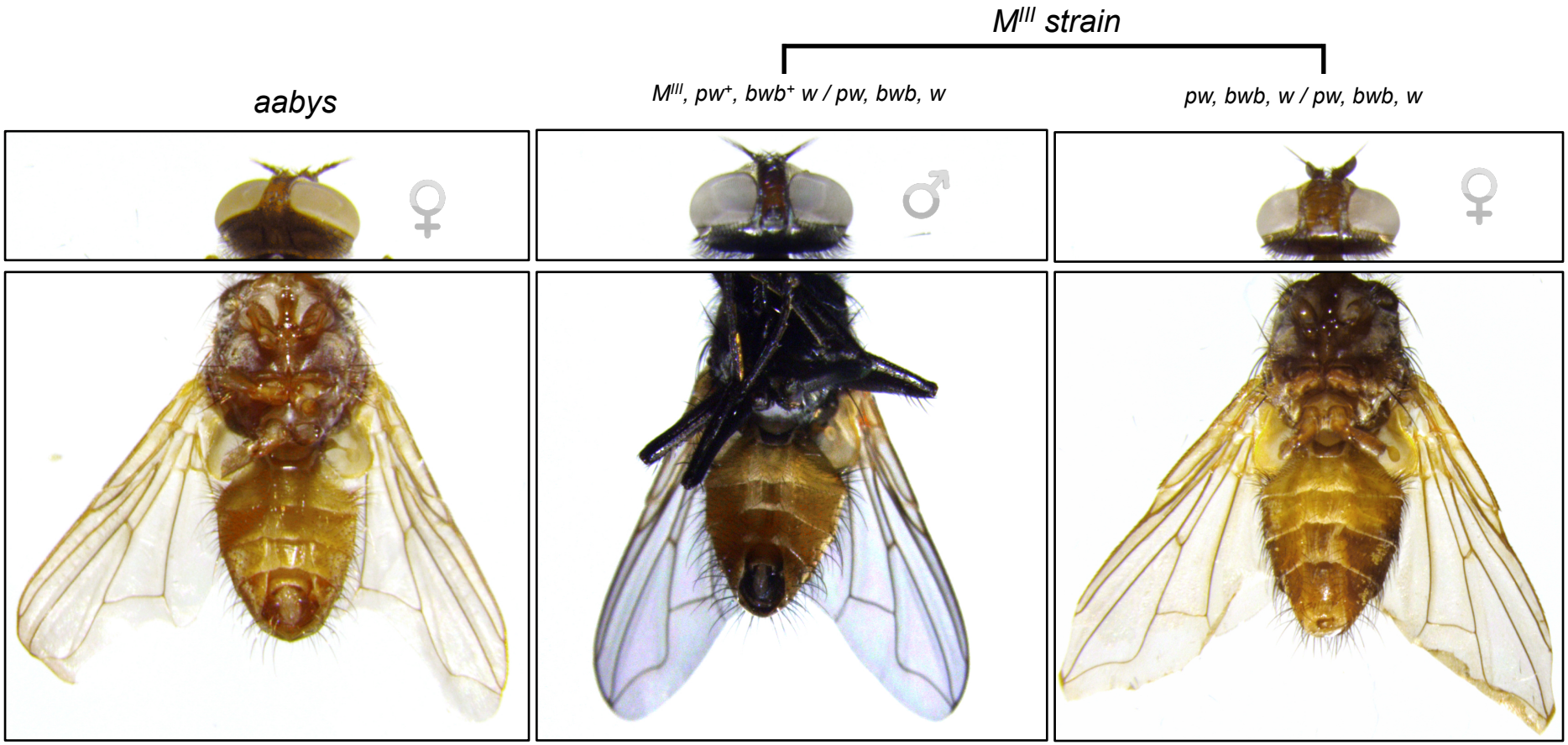
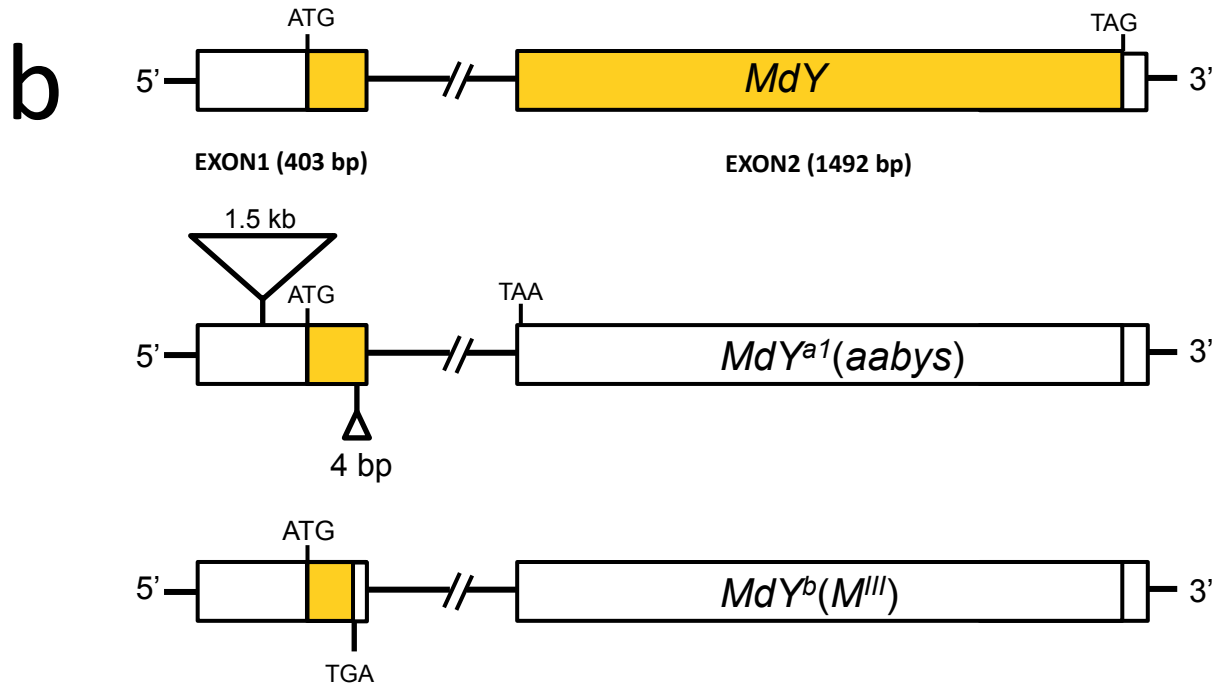
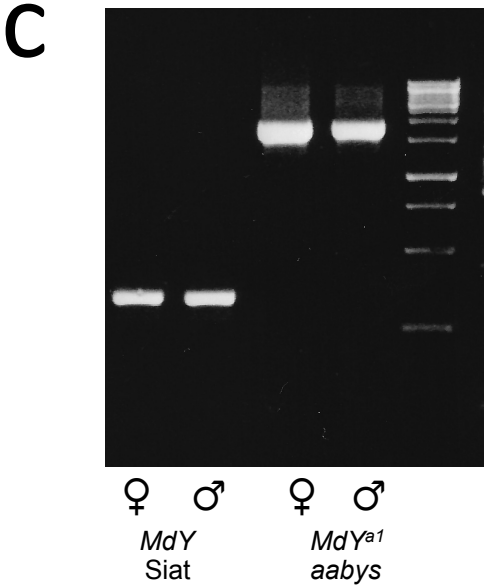


Figure 1 Heinze et al.





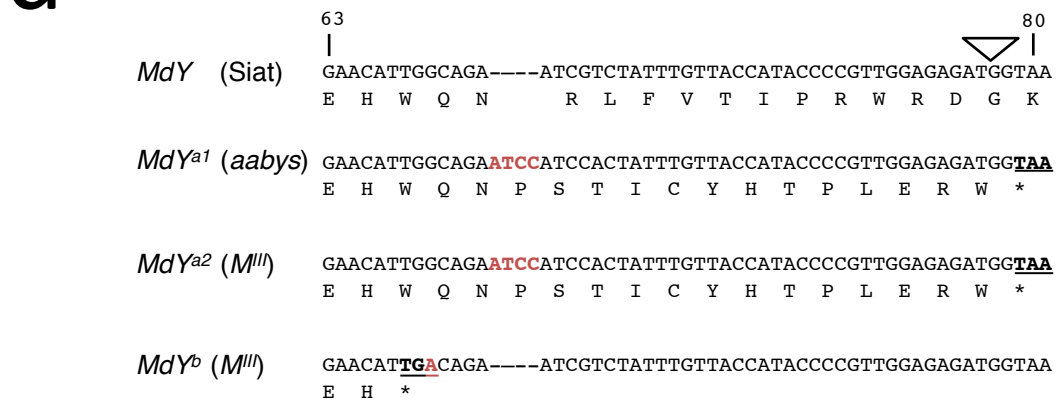
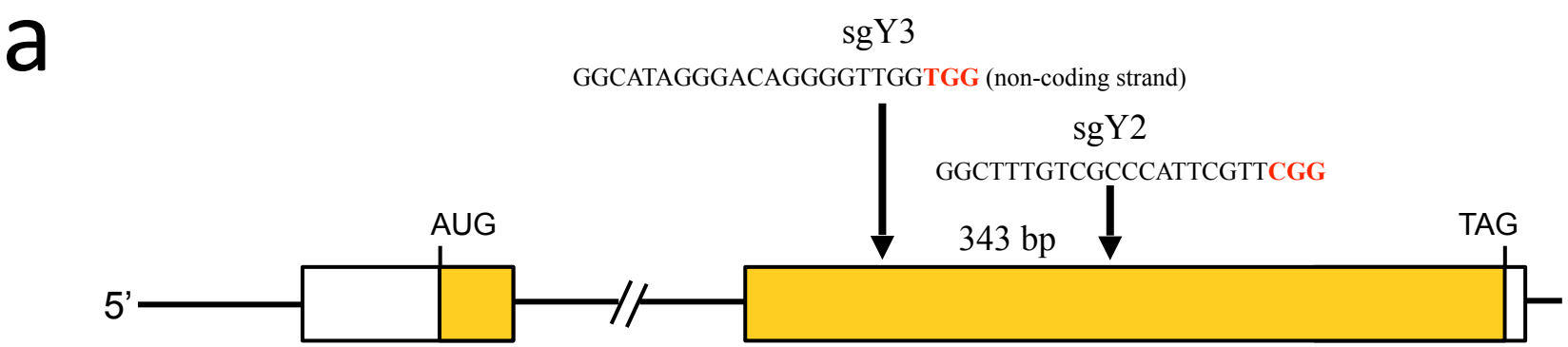
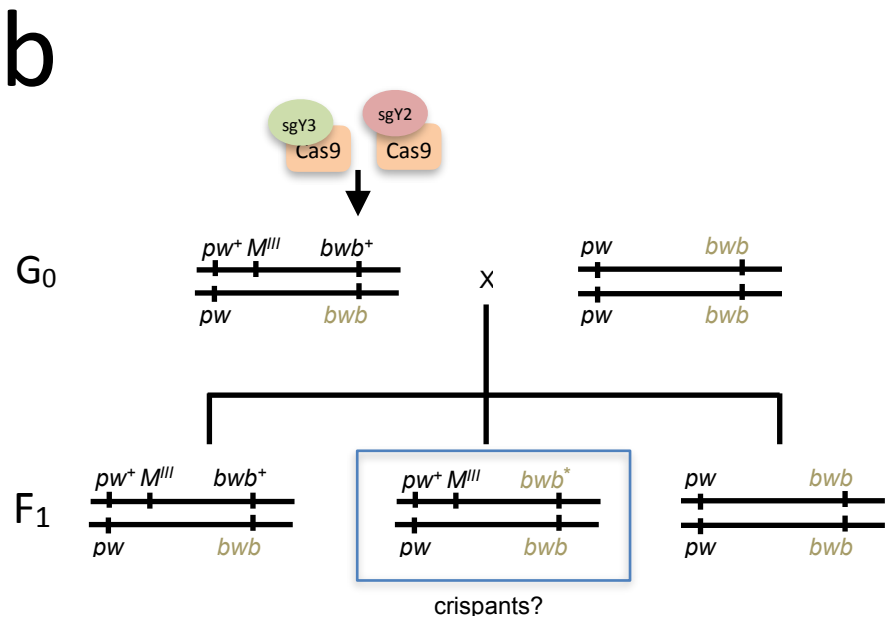


Figure 1 Heinze et al.





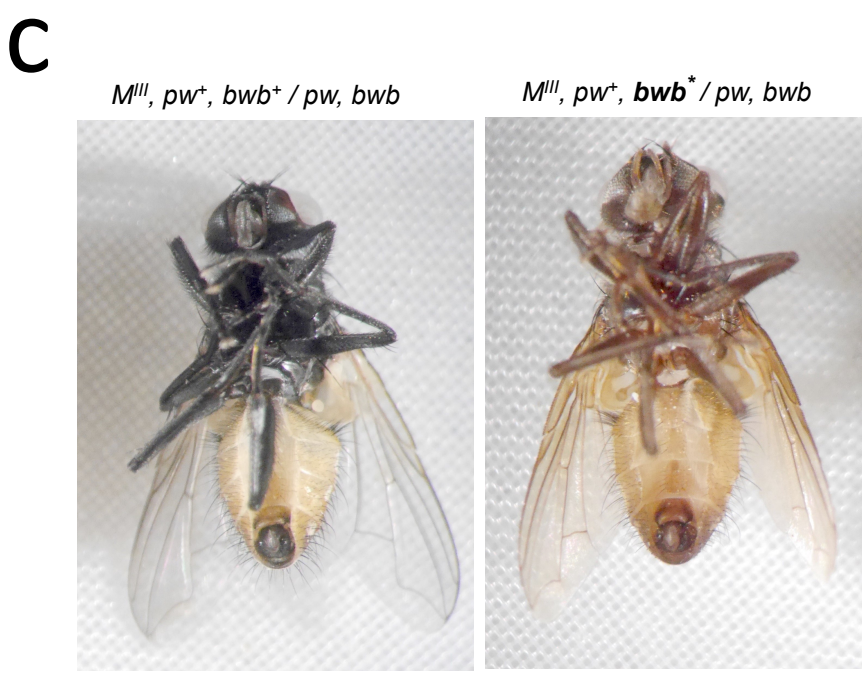


Figure 2 Heinze et al.

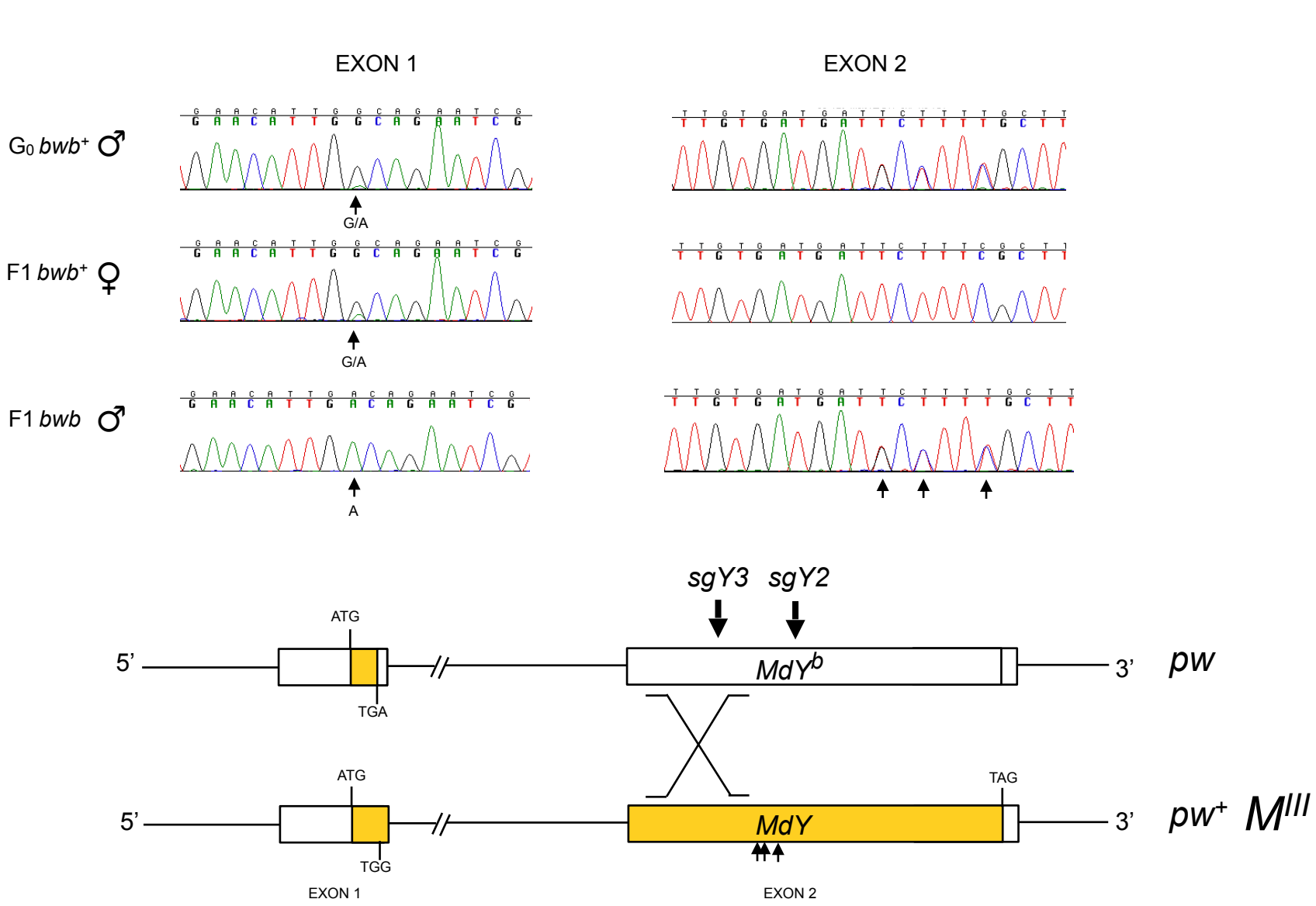


Figure 3 Heinze et al.

	M^{III} *pw*$^{+}$ ***bwb*$^{+}$**	*pw* ***bwb***	M^{III} *pw*$^{+}$ ***bwb***	*pw* ***bwb*$^{+}$**
MdY#2	198	484	18	0
MdY#4	212	183	1	0
MdY#7	136	150	1	0
MdY#9	152	136	3	0
MdY#10	28	86	151	0
MdY#13	110	124	12	0
MdY#14	139	55	1	0
MdY#16	179	174	21	0
MdY#19	147	115	13	0
MdY#29	142	136	68	10
MdY#33	257	211	7	0
MdY#36	163	138	4	0
MdY#38	194	171	5	0
MdY#40	81	62	2	1

Table 1 Heinze et al.

	Δ bwb sg3	*Δ bwb sg2*
MdY#2	1 bp insertion	-
MdY#9	174 bp	-
MdY#10	15 bp	-
MdY#13	11 bp	-
MdY#14	-	2 bp
MdY#16	10 bp	-
MdY#19	14 bp	11 bp
MdY#29	-	-
MdY#33	12 bp	-
MdY#36	43 bp	-
MdY#38	extent not defined	
MdY#40	1038 bp	

Table 2 Heinze et al.