Characterisation of the M-locus and functional analysis of the male-determining gene in the housefly

Wu, Yanli

Chapter 5
General discussion

Sex determination systems differ between species of insects and can even vary within species (Sánchez, 2004; Bachtrog et al., 2014; Beukeboom and Perrin, 2014; Blackmon et al., 2017). How this diversity of insect sex determination systems has evolved still remains unclear. As sex determination in the housefly *Musca domestica* is polymorphic, it provides a perfect model to study the turnover of sex determination systems. Recently, based upon differential expression analysis, a male-determining gene was identified and termed *Mdmd* for *Musca domestica male determiner* (Sharma et al., 2017). To further characterise the male determining loci in terms of genomic organisation and function, I addressed several questions in this thesis: What is the genomic organisation of *M*-loci on different chromosomes? What is the coding sequence of *Mdmd*? To what extent are the different *M*-loci conserved? What is the evolutionary relationship between *Mdmd* and its paralog *CWC22/nucampholin*? What is the expression pattern of *Mdmd* in developing embryos? Can expression of *Mdmd* be manipulated to reverse females into males? By answering these questions I aimed at gaining a better understanding of the evolution of sex determination and sex chromosomes in the housefly and in insects in general.

5.1 The significance of characterising male-determining gene(s) in *Musca domestica*

The primary signals for sex determination are bewilderingly diverse in different insect species. In dipterans, a dominant male-determining factor is the primary signal for male differentiation (Marín and Baker, 1998). Recently, the gene *Nix* was identified as the male-determining factor in the mosquito *Aedes aegypti* (Hall et al., 2015) and *Yob* as male-determining factor in the mosquito *Anopheles gambiae* (Krzywinska et al., 2016). In the common housefly, *Musca domestica*, *Mdmd* was the third male-determining gene characterised from an insect (Sharma et al., 2017). The *Nix* cDNA contains 985 nucleotides and it encodes a protein with 288 amino acids (Hall et al., 2015), whereas the *Yob* gene contains 843 nucleotides and it encodes a protein with 56 amino acids (Krzywinska et al., 2016). The *Mdmd* gene is much larger, as $Mdmd^{V}$ cDNA contains 3525 nucleotides of mRNA coding sequence and the protein sequence of Mdmd consists of 1174 amino acids (chapter 3). These three genes do not show any sequence similarity, indicating that each of these three species uses a different male-determining gene. In addition, *Mdmd* is not found outside of *M. domestica*. The significance of these findings is that each insect species may carry a different male-determining gene, and that many more species need to be investigated before generalities may be recognised. Furthermore, no *tra* gene has been found in mosquitos, indicating that the downstream target gene of the

male-determining factor also differs between mosquito and housefly (both Diptera). Hence, characterising sex determination genes in insects may pose an enormous challenge for the future.

Mdmd appears to not be the only male-determining gene in *Musca domestica*. The same differential expression methods can be used to identify the male-determiner of the M^I strain, as well as for other insect species. Notably, the mutation *Arrhenogenic* (*Ag*) is hypothesised to be a variant of a male-determining gene on autosome I. *Ag* is too weak to repress *Mdtra* activity in the soma, but strong enough to suppress maternal *Mdtra* activity in the germ line (Este and Rovati, 1982; Dübendorfer et al., 2003; Hediger et al., 2010). Hence, identification of the male-determining gene(s) on autosome I will shed light on the evolution of various sex-determining genes within *M. domestica.* It is even conceivable that additional male-determiners be discovered if more housefly populations will be studied in the future.

5.2 Genomic processes responsible for the complex structure of *M*-loci

In the housefly, the *M*-locus contains the male-determining gene(s) and is typically located on the Y-chromosome, but it can also be present on any autosome or even the X-chromosome (Wagoner, 1969; Inoue and Hiroyoshi, 1982; Denholm et al., 1983; Inoue et al., 1986). It has been a longstanding question whether *M*-loci on different chromosomes are similar or contain different male-determining genes. Although *Mdmd* was characterised in Sharma et al. (2017), its complete sequence embedding in the *M*-loci remained unknown, as the chromosomal regions adjacent to *Mdmd* orphan contigs had not been investigated. My study is the first to compare *M*-locus sequence and convincingly shows that *M*-loci of different chromosomes are homologous as they partly share similar sequences.

In Chapter 2 and Chapter 3, I described the complex structure of the *M*-loci on autosome III and V that contain at least one full sized copy of *Mdmd* and a large number of truncated copies of *Mdmd*. The existence of multiple tandemly repeated and truncated *Mdmd* copies requires explanation. How did these extra sequences come to exist in the *M*-loci? A likely explanation is that these extra *Mdmd* copies are the result of *Mdmd* amplification. One possibility is that unequal crossover between sister chromatids produced tandem repeats of *Mdmd*. When unequal crossing over happens close to the gene *Mdmd*, one sister chromatid can lose *Mdmd* and, conversely, the other sister chromatid can get an extra copy (Fig.

5.1). *Mdmd* copies may accumulate further in the *M*-locus if unequal crossing over continues between sister chromatids in the same region. The second possibility for the duplication of *Mdmd* in the *M*-locus is replication slippage or slipped-strand mispairing. Replication slippage usually occurs in a region with tandemly repeated sequences and the *M*-locus provides an ideal template for this. During the replication process of the *M*-locus, the DNA polymerase may dissociate from the DNA and cause a pause in DNA replication. When DNA polymerase reassembles to the template strand, it may accidentally align to the previous position and copy the same sequence more than once, leading to sequence duplication in the *M*-locus. Additionally, sister chromatid gene conversion following DNA double-strand breaks could cause *Mdmd* duplication (Chen et al., 2007). Further molecular study is required to explore how *Mdmd* is being amplified in the *M*-locus.

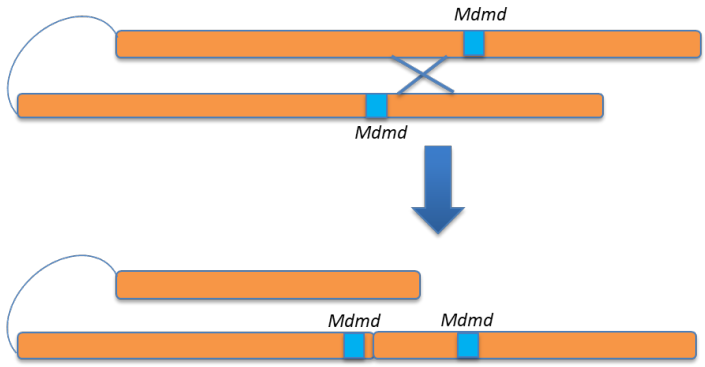

Figure 5.1: Unequal crossing over between sister chromatids causes the duplication of *Mdmd*. Misaligned sister chromatids causes unequal crossing over, yielding a sister chromatid deficient for *Mdmd* and a sister chromatid with two copies of *Mdmd*.

Another feature of the complex *M*-loci is the presence of transposons and repetitive genomic sequences shared with the female genome. The accumulation of repetitive sequences and transposons may in fact promote amplification of *Mdmd* copies through the abovementioned genomic processes. At this stage there is no evidence for an active involvement of transposons in translocating *Mdmd* sequences between chromosomes. In the fly *Megaselia scalaris*, the male-determining factor *M* appears to move between chromosomes through transposition (Traut and Willhoeft, 1990), but we have no evidence that this is also the case in the housefly. Further study is required to determine the causes and effects of the observed association between the *M*-loci containing chromosomes and these repetitive sequences and transposons in the housefly genome.

5.3 Evolution of sex chromosomes in *Musca domestica*

Musca domestica is particularly suited to investigate the evolution of sex determination and sex chromosomes as it harbours several sex determination systems: sex determination based on a male-determining gene on the Y-chromosome, sex determination based on an autosomal male-determining gene(s), sex determination based on a dominant female-determining gene and maternal sex determination. My results provide support for a birth-decay-rebirth model of sex chromosome evolution in *M. domestica* (Fig. 5.2; Rice, 1996; Beukeboom and Perrin, 2014). The male-determining gene *Mdmd* might have arisen by duplication from an autosomal gene *Md-ncm.* After *Md-ncm2* duplicated from *Md-ncm* and translocated among chromosomes, it acquired the novel male sex determination function and became *Mdmd*. The Y-chromosome gradually degenerated due to the insertion of a sex-determining gene(s). To counteract the negative effects brought about by Y-chromosome degeneration, *Mdmd* amplified itself to form the *M*-locus to ensure enough male-determining activity. After amplification, *M*-locus translocated to autosomes and re-arranged afterwards.

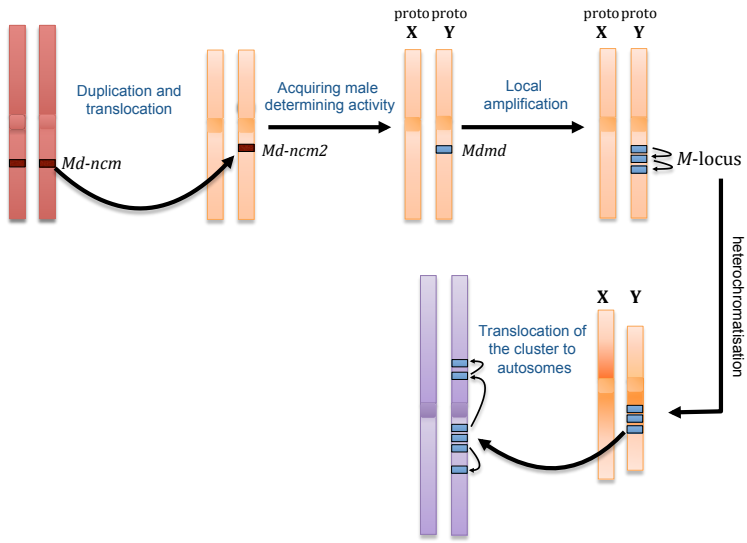

Figure 5.2: A model for the evolution of *Mdmd*: *Mdmd* arose by duplication of *Md-ncm* and acquired the novel male sex determination function. It translocated to an autosome that became a proto-Y chromosome. The Y-chromosome gradually degenerated due to recombination arrest. To counteract the negative effects of Y-chromosome degeneration, *Mdmd* amplified itself to form the complex *M*-locus. After amplification, the *M*-locus translocated to autosomes and underwent further re-arrangements (From Sharma et al., 2017).

Specifically, the high sequence similarity of $Mdmd^{II}$, $Mdmd^{III}$, $Mdmd^{V}$ and $Mdmd^{Y}$ suggests that all *Mdmd* genes originated from a common ancestral sequence. A comparison of Mdmd protein sequences and its paralog CWC22/NCM in Chapter 3 suggests a scenario of *M*-locus evolution, whereby the male-determining gene *Mdmd* evolved after a single duplication event of *Md-ncm* generating a proto-Y chromosome. Gene duplication plays an important role in the origin of new genes (Lynch and Katju, 2004). There is some prior evidence that sex determination genes can arise from genes with other functions, such as a hormone-producing gene or immunity-related gene (Rey et al., 2003; Hattori et al., 2012; Yano et al., 2012). The housefly provides evidence that a sex determination gene can also arise from a splicing regulatory gene.

According to the model of Rice (1996), the genomic region that carries a sex-determining gene is a hot spot for sexually antagonistic genes. The accumulation of sex antagonistic genes in such a region would reduce chromosome recombination surrounding the sex-determining gene. This would be followed by accumulation of deleterious mutations, including insertion of repetitive DNA sequences and transposons due to a lack of recombination on proto-sex chromosomes (Bachtrog, 2005, 2006, 2013). Hence, this model predicts that over long evolutionary time, *Mdmd* will become associated with transposable elements and frameshift mutations. In Chapter 2, I indeed found many repetitive sequences and transposable elements in the flanking regions of *Mdmd* homologous sequences consistent with other studies of early stages of the Y-chromosome evolution (Charlesworth, 1991). For example, in the fish *Oryzias latipes*, it was found that young Y-chromosomes accumulated inactive repetitive elements and transposable element-like sequences in the male-specific region (Nanda et al., 2002; Kondo et al., 2004). The nascent Y-chromosome in the threespine stickleback, *Gasterosteus aculeatus*, has also accumulated duplications and transposable elements (Peichel et al., 2004).

Insertions of transposons may play a dynamic and early role in proto-Y chromosome degeneration and may cause functional genes to gradually lose their function (Bachtrog, 2005). Accumulation of transposable elements and associated tandem repeats can also induce heterochromatin (Lippman et al., 2004), thereby speeding up the degeneration of the Y-chromosome (Bachtrog et al., 2008). For example, in *Drosophila miranda*, the neo-Y chromosome originated only about 1.2 million years ago (Bachtrog and Charlesworth, 2002) and genome analysis has shown that more than 20% of its total DNA consists of repetitive sequences (Bachtrog et al., 2008). Thus the *M*-carrying Y-chromosome may have gradually lost its functional genes except for the *M*-loci. The negative effects brought about by the Y-chromosome degeneration may have created a selection

pressure for *Mdmd* to amplify and translocate in the genome (see below), thereby establishing the complex *M*-locus that is shared by different chromosomes. The finding of the complex *M*-loci in M^{II}, M^{III}, M^{V} and M^{Y} males presented in Chapter 2 reflects this process. The *M*-locus may have translocated multiple times from the Y-chromosome to an autosome and/or subsequently between autosomes. In addition, my cladogram analysis of sequences obtained from genome walking in Chapter 2 revealed that to some extent different sequences exist in different autosomes, indicating that after translocation, the *M*-locus underwent further independent amplification and differentiation. The existence of multiple different autosomal *M* variants in the housefly provides a unique opportunity to further test predictions of early stage sex chromosome evolution. In addition, introducing an active *Mdmd* copy into the genome through germline transformation may be used to study the genomic processes that act on novel sex chromosomes.

5.4 The importance of studying the function of *Mdmd* in the *Musca domestica* sex determination pathway

In Chapter 2, I described the complex *M*-locus on autosome III that contains at least one full sized copy of *Mdmd* plus multiple tandemly repeated and truncated *Mdmd* copies. A similar arrangement of the *M*-loci was also found on autosomes II and V, and the Y-chromosome. Are these tandemly repeated and truncated *Mdmd* copies somehow involved in the male-determining function, or are they merely non-functional remnants of the genomic arrangements discussed above? One possible scenario is that sex in *M. domestica* is determined by more than one male-determining gene, and that the truncated copies somehow contribute to this functionality. Hediger et al. (1998) reported that there are at least two male-determining factors (*M*) located on each arm of the Y-chromosome. They showed that *M* on the short arm of the Y-chromosome provides almost full masculinising activity, whereas the *M* located on the long arm has a weak masculinising activity. My results do not provide unequivocal evidence for multiple activity *Mdmd* genes within an *M*-locus. In Chapter 3, I show high sequence similarity for the *Mdmd* ORF isolated from M^{II}, M^{III}, M^{V} and M^{Y} strains. Hence, until evidence is provided to the contrary, I assume that male sex in the housefly may be determined by a single active *Mdmd* gene.

Targeted disruption of *Mdmd* resulted in complete sex transformation of genotypic males to females, convincingly demonstrating that *Mdmd* is required for male differentation (Sharma et al., 2017). When I tried to express functional

Mdmd^{V} in early blastoderm embryos by injecting mRNA after *in vitro* synthesis (chapter 4), no masculinised flies were observed. These results could suggest that *Mdmd*^{V} is not sufficient to turn genotypic females into males. However, a more likely explanation for this failure to transform genotypic females into males is technical problems. Another approach towards answering the question whether expression of *Mdmd* is sufficient to convert genotypic females into males would be to use *piggyBac* transformation to repeatedly express *Mdmd*^{V} during the whole life-cycle of the housefly. In Box 4.1, I described the cloning of a pBac[3×P3-EGFP, hsp70-*Mdmd*^{V}] transgene. This transgene will be used in future experiments to assess the masculinising activity of *Mdmd*^{V}.

There are several arguments to propose that *Mdmd* encodes a splicing factor. First, *Mdmd* has two conserved domains: MIF4G and MA3. MIF4G is part of eukaryotic initiation factor 4G (eIF4G) that is conserved from human to yeast (Ponting, 2000). The MA3 domain forms a second eIF4A-binding site (Imataka and Sonenberg, 1997). Both domains have important roles in mRNA processing (Ponting, 2000). Moreover, *Mdmd* is a paralog of the splicing regulatory gene *CWC22/nucampholin* (Sharma et al., 2017). Protein CWC22 is required for pre-mRNA splicing and exon junction complex (EJC) assembly (Steckelberg et al., 2012). EJC is involved in mRNA export, mRNA localisation, pre-mRNA splicing and nonsense-mediated mRNA decay (Tange et al., 2004). As alternative splicing is a hallmark of insect sex determination, for example the alternative splicing of the *tra* and *dsx* genes is a conserved feature, *Mdmd* may somehow interfere with the splicing regulation of *Mdtra*.

How could *Mdmd* prevent activation of zygotic *Mdtra*? As reported before, active *Mdtra* directs female-specific splicing of *Mdtra* with the assistance of other essential co-factors such as *Mdtra2* (Burghardt et al., 2005; Hediger et al., 2010). The male-determining gene(s) is thought to antagonise *Mdtra* function and thereby prevents the activation of the autoregulatory loop. Instead, *Mdtra* is spliced in the non-functional male mode. The female splice variant of *Mdtra* has an intact open reading frame encoding a functional protein (MdTRA^{F}), while the male splice variant of *Mdtra* contains additional sequences that encode truncated nonfunctional proteins (MdTRA^{M}) (Fig. 5.3; Hediger et al., 2010). Mdmd protein might sequester the MdTRA/MdTRA2 complex from binding to *Mdtra* pre-mRNA, resulting in a male-specific variant of *Mdtra* (Fig. 5.3). Additionally, the expression pattern of *Mdmd* in embryos (Chapter 4) unveiled ubiquitous expression starting at an early blastoderm stage throughout embryonic development. This indicates that *Mdmd* is acting at a very early embryonic stage to prevent the starting up of the *Mdtra* loop in every cell. Embryos may need continuous expression of *Mdmd* to secure that the loop cannot be reactivated as

long as maternal *Mdtra* is present. Further studies are required to investigate how Mdmd protein interacts with the MdTRA/MdTRA2 complex. An alternative hypothesis is that *Mdmd* exerts a dominant negative effect on *Md-ncm,* a splicing regulatory gene that may be required for female splicing of *Mdtra*. This dominant negative hypothesis can be tested by over-expression of *Md-ncm*. If over-expression of *Md-ncm* turns genotypic males into females, this indicates that *Md-ncm* promotes female development in the *M. domestica* sex determination pathway.

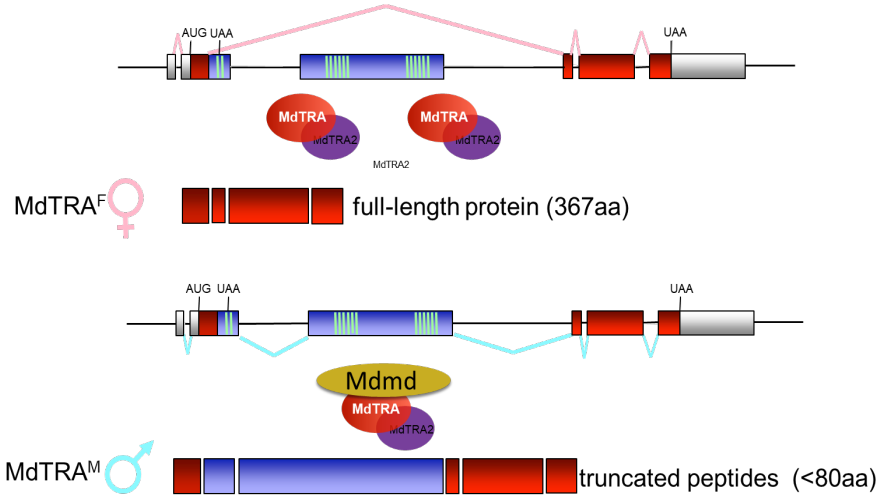

Figure 5.3: A model for splicing regulation of *Mdtra*. The MdTRA/MdTRA2 complex binds to *Mdtra* pre-mRNA, splicing it into female splicing mode. The female splice variant of *Mdtra* with an intact open reading frame expresses a functional protein (MdTRA^F). Mdmd protein might sequester the MdTRA/MdTRA2 complex from binding to *Mdtra* pre-mRNA, resulting in a male-specific variant of *Mdtra*, which contains additional sequences that encode truncated nonfunctional proteins (MdTRA^M). The green lines depict TRA/TRA2 binding sites (Adapted from Hediger et al., 2010).

5.5 Possible applications of *M. domestica* sex determination knowledge

Knowledge of the housefly male-determining mechanism may be useful in pest control. The sterile insect technique (SIT) is a promising method to control pests, which is species-specific and environment friendly. SIT relies on a mass release of sterile males. Currently, irradiation is a common sterilisation method in SIT. The disadvantages of irradiation are that it reduces the sterile insects' fitness and that it is laborious to separate the males from the females. Efficient male

isolation methods are required for improved SIT, such as the use of genetic sexing strains (GSS) that rely on sex-specific mutations for sex separation. Traditionally, development of GSS includes complex physical and genetic manipulations that also reduces the sterile insects' fitness (Munhenga et al., 2016). Based on the complexity to generate GSS strains through traditional methods, it has been suggested to manipulate the sex-determining system to convert females into males, instead of killing them (Saccone et al., 2007). If *Mdmd* is sufficient to perform the male-determining function, a transgenic line may be constructed in which *Mdmd* is able to transcribe under the control of a heat-shock promotion. Masculinisation of genotypic females can be expected by repeated heat-shock induction. As mentioned above, *Mdmd* is not found outside of *M. domestica*, however, as it is quite common in insects that male sex is determined by a dominant male-determining factor (Marín and Baker, 1998; Beukeboom and Perrin, 2014), male-determining factors in other insect species may act the same as *Mdmd*, such as having a function of splicing regulation in male differentiation. Thus, clarifying the role of sex-determining genes in the *M. domestica* sex determination pathway will not only provide fundamental insights into insect sex determination systems but may also have applied relevance.

5.6 A hypothesis for temperature driven transition of *Musca domestica* sex determination

The evolutionary dynamics of sex determination transitions in *M. domestica* are not yet well understood. Previous studies reported that the distribution of the different sex determination variants follows latitudinal clines (Feldmeyer et al., 2008; Kozielska et al., 2008). In the northern hemisphere, a male-determining locus (*M*-locus) on the Y-chromosome is more common at higher latitude and its frequency gradually decreases towards lower latitude where the *M*-locus occurs more often on one of the autosomes (Franco et al., 1982; Denholm et al., 1986; Tomita and Wada, 1989; Cakir and Kence, 1996; Hamm et al., 2005; Kozielska et al., 2008). In populations with high frequencies of autosomal *M*-locus males, females carry a dominant female determining factor $Mdtra^D$, on autosome IV, which is insensitive to male-determining gene(s) repression (Mcdonald et al., 1978). A similar cline is observed in the frequency of $Mdtra^D$. Feldmeyer et al (2008) found that the yearly mean temperature, interacting with humidity, is the main responsible factor for the clinal distribution of $Mdtra^D$, while the temperature gradient appears to correlate with the geographical distribution of autosomal *M*-loci.

These geographical patterns suggest that the environment, in particular temperature, affects the transition of *M. domestica* sex determination systems. It has thus been hypothesised that temperature is the driving force for the turnover in sex determination in the housefly (Fig. 5.4). Specifically, temperature induced competition between *Mdmd* and *Mdtra* expression may drive changes in housefly sex determination. One could imagine that the two genes compete to attain efficient *Mdtra* splicing as the alternative splicing process is known to be temperature sensitive (Jakšić and Schlötterer, 2016). An experimental procedure can be designed to measure the expression level of *Mdmd* and female and male *Mdtra* transcripts under different temperature regimes. The output level of female *Mdtra* transcripts may be higher in the south with high temperature (Fig. 5.4A). This in turn may require increased levels of *Mdmd* to repress the *Mdtra* autoregulatory splicing loop and insure male determination. One could further hypothesise that *Mdmd* on the Y-chromosome is too weak in the south due to Y-chromosome degeneration. Hence, *Mdmd* is selected for a stronger male-determining effect and this may be accomplished by amplification (Fig. 5.4B) and translocation to autosomes, and as a result sex is determined by an autosome carrying an *M*-locus instead of the Y-chromosome (see the appearance of *XX* male in Fig. 5.4C). Next, the appearance of the mutation $Mdtra^{D}$ can be interpreted as escaping from this arms race, as $Mdtra^{D}$ is insensitive to *Mdmd* (Fig. 5.4D). Finally, presence of $Mdtra^{D}$ may then cause accumulation of *M*-loci in single individuals because they also segregate to females (Fig. 5.4E). This hypothesis would explain why $Mdtra^{D}$, which induces female development even in the presence of several *M*-loci, is prevalent at high temperature (Feldmeyer et al., 2008).

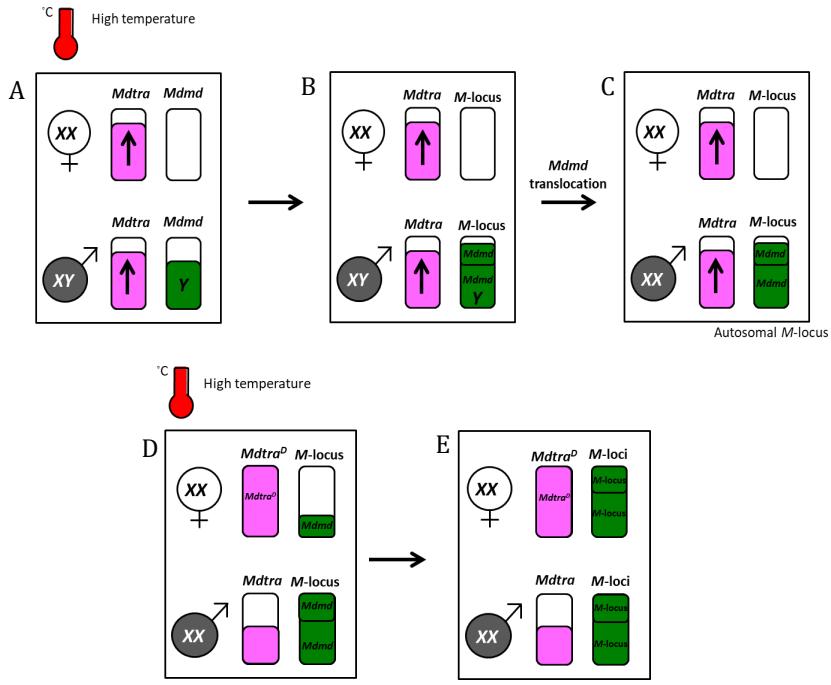

Figure 5.4: A hypothesis for temperature driven transition in housefly sex determination. Temperature affects the relative expression of the female and male-determining genes. A: The output level of female *Mdtra* transcripts may be higher with high temperature. B and C: This may require increased levels of *Mdmd* to repress the *Mdtra* autoregulatory splicing loop, which may be accomplished by *Mdmd* amplification and translocation to autosomes. D and E: Next, *Mdtra* mutated into $Mdtra^D$ that is insensitive to *Mdmd* and presence of $Mdtra^D$ may then cause accumulation of *M*-loci in single individuals. Green color represents the strength of *Mdmd* and pink color of *Mdtra*. More coloring in the *Mdmd* and *Mdtra* bars means greater expression strength (Adapted from L. W. Beukeboom, unpublished).

5.7 Concluding remarks

My study has contributed to a deeper knowledge of how the complex male-determining locus in the housefly is organized and to a better understanding of how sex determination systems and sex chromosomes may evolve. I could show that the male-determining regions in the housefly are homologues and contain repetitive sequences. However, the current assembly of the *M*-loci in the housefly is not yet sufficient and further characterisation using sequencing platforms that produce long reads will be required to determine the precise organisation of the *M*-loci in different *M. domestica* strains.

In my study I characterised the recently discovered *Mdmd* in greater detail and

could show the *Mdmds* in M^{II}, M^{III}, M^{V} and M^{Y} strains share a high level of sequence identity, suggesting that they have been only very recently obtained. Expression analysis of *Mdmd* RNA unveiled ubiquitous expression as early as in the blastoderm stage, which is maintained throughout embryonic development, consistent with *Mdmd*'s predicted role to prevent zygotic activation of the *Mdtra* feedback splicing loop by maternal *Mdtra* (Hediger et al., 2010; Sharma et al., 2017). Embryos may need continuous expression of *Mdmd* to secure that the loop cannot be reactivated as long as maternal *Mdtra* is present. The question remains whether *Mdmd* is solely sufficient for male development in *M. domestica* strains. Further functional analysis of *Mdmd* in embryos through germline transformation is required to determine the precise function of *Mdmd* in male determination.

Investigation of the complex male-determining locus and sex-determining genes provide empirical support for the birth-decay-rebirth model of sex chromosome evolution (Rice, 1996; Beukeboom and Perrin, 2014). It states that sex chromosomes evolved from ordinary autosomes that lost recombination after having acquired a sex-determining role. The origin of *Mdmd* from duplication of *Md-ncm* can be regarded as the birth of a sex chromosome and the presence of repetitive sequences and transposons as the early signs of degeneration. I hope my study will contribute to a further understanding of sex determination systems in *M. domestica* and the evolution of sex chromosomes in the housefly and in insects in general.