Sex chromosome evolution in the house fly

Li, Xuan

DOI:
10.33612/diss.206102060

Chapter III

Fluorescence *in situ* hybridization reveals localization of *Musca domestica male determiner* (*Mdmd*) loci and associated repeat regions

Authors: Xuan Li, Sander Visser, Louis van de Zande, Leo W. Beukeboom

Affiliations: Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands.

Abstract

In many dipteran insects, male sex is determined by a dominant locus (*M*) on the Y-chromosome. In the house fly, *Musca domestica*, *M* loci have been found not only on the Y chromosome, but also on the X chromosome or on any of the five autosomes. To get more insight in the fine scale chromosomal location and genomic properties of *M* in house fly populations, we applied fluorescence in situ hybridization (FISH) to different chromosomes with mapped *M* loci. Based on the sequence of the recently identified male determining gene, *Musca domestica male determiner* (*Mdmd*), two types of *M*-derived probes were generated: (1) a "mixed" probe, containing *Mdmd* sequences and additional *M* associated sequences and (2) an *Mdmd* specific probe. Using probe (1), we identified *M* and Sex chromosome located (MAS) repeats, which differ in number between the X and Y. Apart from being located within the *M* locus, on both the Y and the X chromosome, MAS repeats were also found on X chromosomes without *M*. In all instances, probe (2) revealed *M* loci on the Y chromosome, autosome II and autosome III to be located near the centromeres, whereas the *M* locus on the X chromosome is located in the middle of one chromosome arm. *M* loci were not detected for autosome I, as this locus does not contain *Mdmd*, but also not on autosome V, probably for technical reasons. The observed locations support the hypothesis that for all *M*-containing autosomes, the transpositions were single events. These results are discussed in the light of *M* locus translocation, and sex chromosome evolution in the house fly.

Introduction

Sex determination mechanisms are variable among taxonomic groups, but they are generally fixed at species level. Polymorphic systems are however known from a number of species, including several fish, frogs and invertebrates (Beukeboom and Perrin 2014). Such systems are useful for studying the evolution of sex determination systems. The house fly, *Musca domestica*, is a dipteran species that has a unique system of polymorphic sex determination (Dübendorfer et al. 2002; Hamm et al. 2015). The typical house fly karyotype consists of five pairs of autosomes and a pair of sex chromosomes (XY) (Hediger et al. 1998b). A male determining locus (*M*), has been identified not only on the Y chromosome (M^Y) but also on any of the five autosomes ($M^{I\text{-}V}$, collectively known as M^A) or even on the X chromosome (M^X) (Hiroyoshi 1964; Franco et al. 1982; Denholm et al. 1986; Tomita and Wada 1989; Denholm et al. 1990; Hamm et al. 2005; Feldmeyer et al. 2008; Kozielska et al. 2008). All these *M*-carrying chromosomes are considered of recent origin and they are collectively referred to as "neo-Y" chromosomes (Meisel et al. 2017). These polymorphic "neo-Y" chromosomes reflect an ongoing process of sex chromosome turnover in the house fly. Studying the nature of such polymorphisms can provide novel insights into genomic processes that underlie sex chromosome evolution and mechanisms of evolutionary changes in sex determination systems.

Global studies demonstrated that the geographic distribution of different *M* loci follows a latitudinal cline in house fly populations (Tsukamoto et al. 1980; Franco et al. 1982; Hamm et al. 2005; Kozielska et al. 2008). The M^Y locus is mainly found in the high latitude regions, whereas M^A loci are mostly found in the low latitude regions. In addition, because of the presence in low latitudinal populations of a dominant female determining allele, tra^D, which overrules the function of *M,* homozygous M^A loci are more frequently found in such regions.

All published studies of *M* factor localization have used autosomal markers. Although this method is sufficient to determine the specific chromosome location, it also has limitations. For instance, there are no known markers for the X and the Y chromosome that makes it difficult to distinguish the presence of M^X or M^Y loci. In addition, the method cannot accurately determine the within chromosome *M* locus location, although differences in recombination rates between marker and *M* locus may hint at such intra-chromosomal variation. Hence, many questions about the polymorphic sex determination system of the house fly remain. How did multiple *M*

loci originate? Did autosomal *M* originate from a single translocation event of M^Y and subsequently spread all over the world, or did multiple translocations occur and always to a particular region on each of the chromosomes? The Y chromosome possesses two distinctive regions with male determining function (Hediger et al. 1998a), can this also be the case for other *M*-carrying chromosomes? Alternative approaches to localize *M* loci in the house fly genome, such as fluorescence *in situ* hybridization (FISH), can be applied to obtain more detailed knowledge of the genomic dynamics of *M* loci.

Sharma et al. (2017) identified a male determining gene, *Musca domestica male determiner* (*Mdmd*), as a part of the *M* locus on the Y chromosome as well as on autosome II, III and V, indicating that the autosomal *M* loci are the results of multiple translocation events. Partial sequence analysis revealed that *M* loci are complex and contain multiple *Mdmd* copies, most of them incomplete, while the intervening sequences include repetitive sequences (Sharma et al. 2017; Wu 2018).

In this study, we apply FISH to localize *M* loci in more detail on various house fly chromosomes. We ask where on the chromosomes *M* loci are located and whether any of the *M*-carrying chromosomes possess more than one *M* locus. In addition, we compare the chromosomal locations of *M* loci in house fly populations of various geographical origins to detect potential inter-population variation. To do so, we first designed a probe that is composed of a mixture of *Mdmd* and additional *M* locus sequences (the Mix probe). This maximizes the detection of possible *M* loci, as this will be not dependent on detecting *Mdmd* homology only. In addition, we prepared a second probe (the *Mdmd* probe) that specifically targeted *Mdmd* sequences. We discuss how our findings increase our understanding of the evolution of the house fly polymorphic sex determination system.

Materials and Methods

House fly strains

Eleven house fly strains were used for the localization of *M* loci. Three are long-term laboratory strains. The universally used *aabys* strain that has an XY sex chromosome system (Tomita and Wada 1989, described in Chapter 2). The M3 strain of which males carry an *M* locus (hemizygous) on autosome III that is linked to the wild type allele of *brown body* (*bwb*) and *pointed wings* (*pw*) mutations (see Chapter 2 for

detailed information of phenotypical markers). Females are homozygous for the recessive mutant *bwb* and *pw* alleles and do not carry the *M* locus. Both males and females carry two X chromosomes and no Y chromosome. The M5 strain in which males carry a hemizygous *M* on autosome V that is linked to the wild type allele of *ocra eyes* (*ocra*). Females are homozygous for the recessive *ocra* mutant allele and do not carry the *M* locus. Both males and females have two X chromosomes.

The remaining eight strains were established from collections of wild populations. They include the strain NL1 from Gerkesklooster, the Netherlands and five strains (SPA1-5) from Spain that have been described in Chapter 2. Additionally, two Italian strains (ITA1, ITA3) were used. ITA1 was collected in Altavilla Silentina and ITA3 in Castelianeta marina in October, 2013. Linkage mapping identified these populations as M^{III} for ITA1 and M^{II} plus M^{III} for ITA3 (S. Visser & L. W. Beukeboom unpublished). Both hemizygous and homozygous *M* loci were found in both strains.

At the start of this study, the Italian strains had been maintained in the laboratory for approximately 90 generations, the Spanish strains for approximately 50 generations and the Dutch strain for approximately 40 generations.

Chromosome preparations

Chromosome slides were prepared from brain tissues of third instar larvae. Spreads of mitotic chromosomes were made according to the method of Carabajal Paladino et al. (2014) with slight modification. In short, larval brains were dissected in Ringer's solution (182 mM KCl, 46 mM NaCl, 3 mM $CaCl_2$, 10 mM Tris; pH to 7.2), pre-treated in hypotonic solution (75 mM KCl) for 10 minutes and then fixed in Carnoy's fixative (ethanol:acetic acid, 3:1) for 10 minutes. Fixed tissues were then transferred to glass slides (Thermo Fisher Scientific SuperFrost Microscope Slides) with a drop of 60% acetic acid and spread with a tungsten needle on a 45 °C heating plate. Slides were examined under a phase contrast microscope (Carl Zeiss Axio Lab.A1) to check whether the nuclei were appropriately spread before FISH.

After brains were dissected out, the remaining larval tissue was used for DNA extraction using a high salt protocol (Aljanabi and Martinez 1997) followed by PCR with primers designed on *Mdmd* specific sequences (Mdmd_1as and Mdmd_6as according to (Sharma et al. 2017)) to determine whether the larva possessed the *Mdmd* sequences. PCR was conducted with Platinum II *Taq* Hot-Start DNA

Polymerase (Thermo Fisher Scientific) according to the manufacturer's instructions. The total 20 µl PCR mixture contained ~200 ng template DNA. The thermocycling program was as follows: initial denaturation at 94 °C for 2 min; 30 cycles of 15 sec denaturation at 94 °C, 15 sec annealing at 60 °C, 1 min 15 sec extension at 72 °C. PCR products were visualized on a 1% agarose gel in TAE buffer to evaluate the presence of *Mdmd* in the samples. As the used primers are designed to amplify intervening sequences between *Mdmd* copies, successful amplification not only confirms presence of *Mdmd* sequences but also of multiple copies. Only the chromosome slides from confirmed *Mdmd*-carrying larvae were used for FISH and for each strain, 2-3 individuals were investigated.

Probe preparation

Two probes were prepared for FISH experiments. The first probe, the Mix probe, contains multiple *Mdmd* copies with sequence variation, of which some contain insertions of transposable elements or small repeats. It can target both *Mdmd* and non-*Mdmd M*-locating sequences. Owing to complex signals obtained in FISH experiments using the Mix probe, we also prepared a more specific *Mdmd* probe consisting of only *Mdmd*-specific sequences, using a cloned *Mdmd* gene (sequence shown in Figure S3.1).

To prepare the *Mdmd*-probe, we first amplified the *Mdmd* sequences with the *Mdmd* specific primer pair Mdmd_FISHs and Mdmd_FISHa (Table S3.1) and M3 male genomic DNA. PCR was conducted using Platinum II *Taq* Hot-Start DNA Polymerase according to the manufacturer's instructions. The PCR mixture and condition were same as above and resulted in fragments sizes of approximately 2-5 Kb. PCR products were run on a 1% agarose gel in TAE to verify successful amplification and then purified with the GeneJET PCR Purification Kit (Thermo Fisher Scientific).

The Mix probe was prepared directly from the aforementioned purified PCR products (amplified with Mdmd_FISHs and Mdmd_FISHa primers). Labeling was performed with digoxigenin(DIG)-11-deoxyuridine triphosphate (dUTP) using the DIG-Nick Translation Mix (Roche) according to the manufacturer's instructions. The total 20µl labeling mixture consisted of ~1 µg DNA, 4 µl DIG-Nick Translation Mix and was incubated at 15 °C for 90 min. DIG-Nick Translation Mix randomly breaks the DNA molecules while labeling resulting in probe fragments that range mainly from 200-700 bases. The labeled products were purified with the GeneJET PCR Purification Kit

and used as probes in FISH experiments.

To generate the *Mdmd* probe, we ligated the purified PCR products into the pGEM-T vector and transfected the construct into competent *E. coli* JM109. The *Mdmd* insert was sequenced and mapped against the published *Mdmd* sequence (Sharma et al. 2017) to assess the fragment identity after alignment with Clustal Omega (Madeira et al. 2019). Plasmids containing the *Mdmd*-insert were isolated with GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The *Mdmd* sequences were then amplified using the plasmid DNA and pGEM-T primers using Platinum II *Taq* Hot-Start DNA Polymerase according to the manufacturer's instructions with 1 min extension time. Finally, the products were labeled with DIG-11-dUTP and purified.

Fluorescence *in situ* hybridization

The FISH procedure was adapted from Huang et al. (2007) with minor modifications. Chromosome slides were pretreated with 100 µg/ml RNase A in 1×PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4; pH 7.0) for 1 hour at 37 °C, followed by washing three times with 2×SSC (300 mM sodium citrate, 30 mM NaCl; pH 7.0) at room temperature for 5min each. Subsequently, the slides were denatured in 2xSSC containing 70% formamide at 68°C for 3.5 min, dehydrated by passing them through an ice-cold ethanol series (70%, 90%, 100%; 5 min each) and air-dried. The 20 µl probe mixture contained 200-300 ng digoxin-labeled probe DNA, 50% (v/v) deionized formamide, 10% (v/v) dextran sulphate in 2×SSC. The probe was denatured at 90 °C for 5 min and rapidly cooled on ice for 10 min. The denatured probe mixture was then applied to the slides and left to hybridize at 37 °C for at least 14 hours.

After hybridization, slides were washed with 2×SSC, 50% formamide at 42 °C for 10 min, followed by three washes with 2×SSC at 42 °C for 5 min each. Slides were blocked with 3% (w/v) bovine serum albumin (BSA) blocking buffer (dissolved in 4XSSC with 0.1% Tween 20). Probes were detected with Anti-Digoxigenin-Rhodamine (Roche) by incubating at 37 °C for an hour. Slides were then washed three times with washing buffer (4×SSC with 0.1% Tween 20) at 37 °C for 5 min each. After washing, slides were shortly rinsed with 2×SSC and air-dried. Chromosomes were counterstained with ProLong Diamond Antifade Mountant with DAPI (Thermo Fisher Scientific). Signals were detected with a Leica epifluorescence microscope (DMI6000 B) equipped with a Leica CCD camera (DFC365 FX) and analyzed with Leica

Application Suite X (3.4.2.18368.1.2). Chromosomes were numbered according to Hediger et al. (1998b) in which autosomes were ranked based on their relative length. Sex chromosomes (XY) were distinguished by their large shape differences, the Y being much smaller than the X (Boyes et al. 1964; Wagoner 1967; Hediger et al. 1998b). Signals were only considered as a successful hybridization if they were observed with consistent chromosomal location on at least 20 metaphase nuclei.

As we obtained additional signals on the X and Y chromosomes using the Mix probe, the size of those signals was measured with Adobe Photoshop CC 2018 software. For each signal, the average length in pixel units of three measurements was used to minimize measurement errors. The signal length difference was expressed as the ratio of Y:X signal. The Y:X ratios were calculated only within a nucleus to avoid variation in hybridization strength and chromosome density between nuclei.

The FISH method applied in this study has limited sensitivity, meaning that short sequences, such as a single *Mdmd* copy that is ~3.7 Kb in length, will not be identified due to insufficient fluorescence intensity.

Results

Probe quality

Amplification with Mdmd-FISH primers and M3 male genomic DNA as template, resulted in fragments with varying sequence length, indicating amplification of multiple *Mdmd* copies (Figure 3.1 A1). PCR with M3 female genomic DNA as template did not generate any product indicating that the Mdmd-FISH primers are *M* locus specific (Figure 3.1 A2). The cloned *Mdmd* sequence was ~3 Kb (indicated by the red arrow in Figure 3.1 A). Sequencing of the cloned fragment confirmed that it contained only the *Mdmd* sequence (Figure S3.1). After the labeling procedure, the probe fragments ranged from 200-700 bases (Figure 3.1B), which is in the optimal range for FISH experiments. Both probes resulted in the same size range of probe fragments (Figure 3.1B) ensuring that differences in hybridization patterns are only the result of sequence variation between the probes.

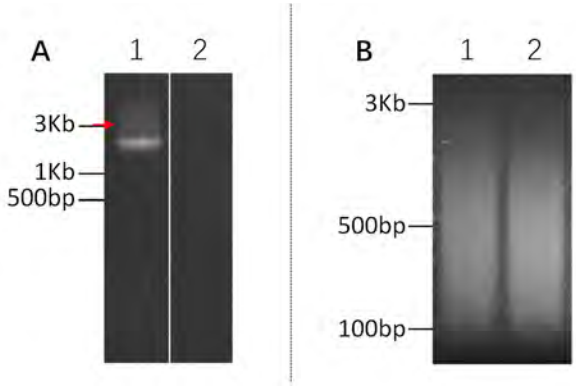

Figure 3.1 The amplification of *Mdmd* sequences and probe labeling. **A:** PCR products using Mdmd-FISH primers and M3 male genomic DNA (1) or M3 female genomic DNA (2). The multiple bands in lane 1 show amplification of different *Mdmd* variants. As is shown in lane 2, Mdmd-FISH primers do not amplify from female genomic DNA and thus are *M* locus specific. The red arrow indicates the fragment that was cloned for probe development. **B:** Labeled probe products using mixed PCR amplified from M3 male genomic DNA (1) and PCR products amplified from the *Mdmd* clone (2). DNA fragments mainly ranged from 200 bp-700 bp for both probes.

Mix probe FISH results

We first used the Mix probe in FISH with male samples of SPA1 and SPA4 as well as both males and females of the *aabys* strain. SPA1 and SPA4 samples possess homozygous M^X, whereas only *aabys* males carry M^Y. Hybridization signals were observed at the end of the short arm of the Y chromosome but also on both ends of the X chromosome in males of *aabys* (Figure 3.2 A). Similar signals on the X chromosomes were detected in SPA1 and SPA4 samples and *aabys* females (Figure 3.3 B, C, D). In SPA1, SPA4 and *aabys* male samples, signals are expected to be, in part, due to hybridization of *Mdmd* sequence regions be the result of hybridization to *M*-loci as they do possess *M* locus. A*abys* females, however, do not possess the *M* locus. Apparently, the observed signals are the result of hybridization of the Mix probe to non-*M* sequences on the X chromosome.

The Mix probe signals cover a large terminal region of the sex chromosomes

suggesting that the hybridized region is of considerable sequence length (likely covering multiple Mbs), indicating that these hybridized regions contain clusters of repetitive sequences. For easier description, from here on, we refer to these repeats as *M* And Sex chromosome located (MAS) repeat and make a distinction between the region found on the Y, MAS^{Y}, and the X chromosomes, MAS^{X}. The lengths of the MAS^{Y} and MAS^{X} were measured and compared in XY male samples (Table S3.2). These data show that there are two regions of MAS on the different arms of the X chromosome, and that one MAS^{X} region (referred as AS^{X1}) is larger than the other (referred as MAS^{X2}). MAS^{Y} contains only a single region. On average, the ratio of total length of MAS^{Y} to MAS^{X} is 0.42 (±0.06) showing that MAS^{X} is much larger than MAS^{Y} (Table S3.2).

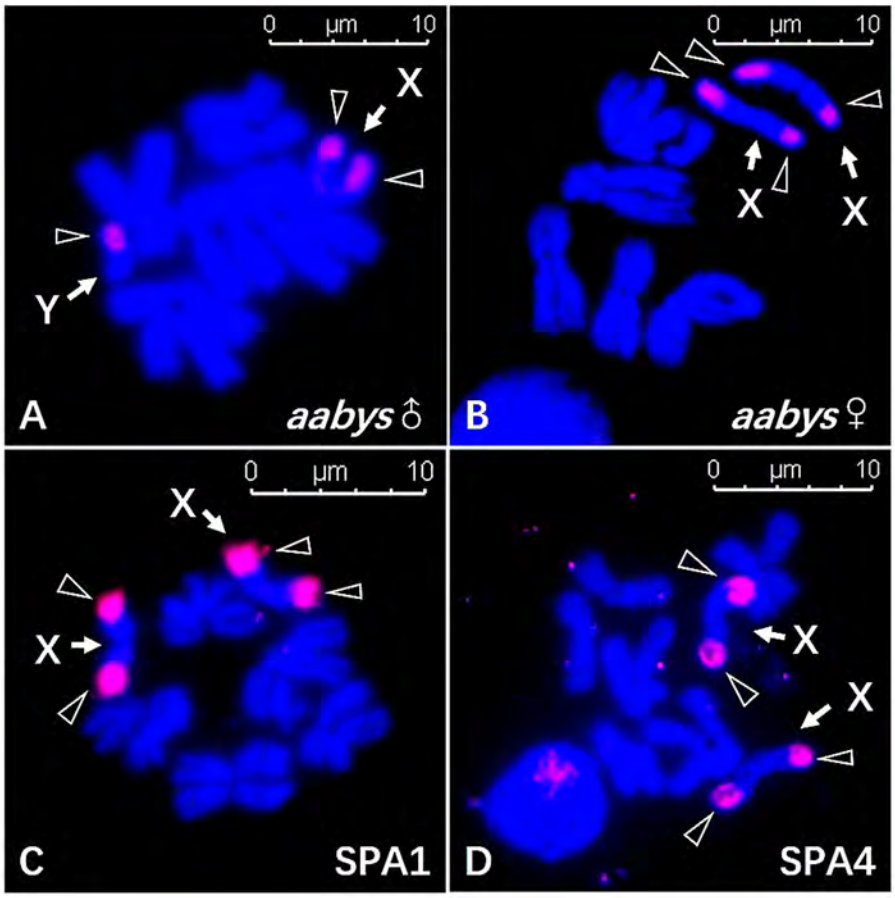

Figure 3.2 FISH results with the Mixed probe. **A**: *aabys* male with one X and M^{Y} on

the one Y chromosome; **B**: *aabys* female with two X chromosomes; **C**: SPA1 carrying *M* on the X chromosomes; **D**: SPA4 carrying *M* on the X chromosomes. Signals are shown in red and indicated by open triangles and sex chromosomes are indicated by arrows. Mitotic chromosomes are shown in blue.

Mdmd probe FISH results

To avoid hybridization of non-*Mdmd* containing regions as seen using the Mix probe, the chromosomal locations of *Mdmd*-containing *M* loci were detected by hybridization with the *Mdmd*-specific probe. Hybridization of the *Mdmd* probe resulted in successful localization of the *M* loci on autosome II, autosome III and both X and Y chromosomes. M^{Y} was localized at the pericentromeric region on the short chromosome arm in males of *aabys*, ITA1 and GK strains (Figure 3.3 A-C).

Homozygous *M* loci were localized on the X chromosomes in samples from four strains, SPA1, SPA2, SPA4 and ITA2 (Figure 3.3 D-G). Unlike M^{Y} that is present at the pericentromeric region, M^{X} is located at the middle region on one arm of the X chromosome in all samples. In addition to M^{X}, in SPA4 and ITA2, a hemizygous M^{II} locus was observed (Figure 3.3 F, G). M^{II} resides on the short arm of autosome II at the pericentromeric region. No variation in location of the M^{II} locus was observed among SPA4 and ITA2 samples.

Hemizygous M^{III} was identified in M3, SPA3 and SPA5 samples (Figure 3.3 H-J). M^{III} is located at the pericentromeric region on the short arm of autosome III. No variation in location of the M^{III} loci was observed among different samples. Note that the sex chromosome pair is XX, without an *M* locus in these M3, SPA3 and SPA5 samples.

None of the three tested samples from the M5 strain showed hybridization signals (Figure 3.3 K), despite testing positive for the presence of *Mdmd* by PCR. This is most likely due to insufficient gene copy numbers.

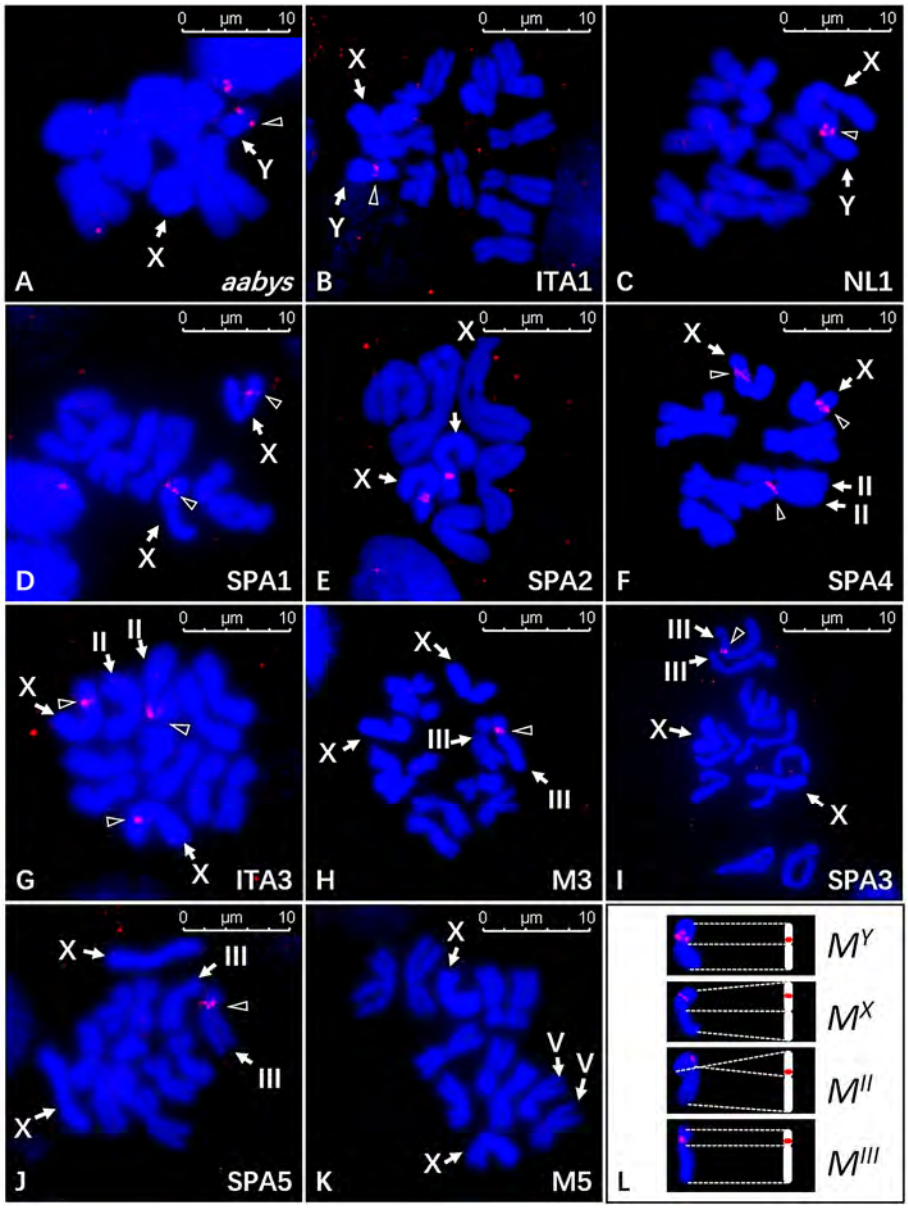

Figure 3.3 FISH localization of *M* loci on autosomes and sex chromosomes in various house fly strains. **A-C:** *M* locus on the Y chromosome in *aabys*, ITA1 and NL1. **D, E:** Homozygous *M* locus on the X chromosomes in SPA1 and SPA2. **F, G:** Homozygous *M* locus on the X chromosomes with an additional hemizygous *M* locus on autosome II in SPA4 and ITA3. **H-J:** Hemizygous *M* locus on autosome III in M3, SPA3 and SPA5. **K:**

No hybridization signal was observed on autosome V in M5. **L:** Schematic drawings of the mitotic chromosomes with chromosomal locations of *M* loci on Y and X chromosomes, autosome II and autosome III. Positive signals are shown in red and indicated by open triangles, chromosomes are indicated by arrows. Mitotic chromosomes are shown in blue.

Our results are in agreement with a prior linkage-mapping study of *M* loci (this thesis, Chapter 2). With linkage-mapping, SPA3 and SPA5 flies were identified carrying hemizygous M^{III} locus. Homozygous M^{X} was found in SPA1, SPA2 and SPA4 strains. Additionally, SPA4 flies were found occasionally possessing M^{I} and M^{II} loci. SPA2 flies additionally possessed hemizygous *M* on the sex chromosome and M^{II} with low frequency. M^{Y} was found in the NL1 strain. In Italian strains, M^{III} was found for ITA1 males and M^{II} plus M^{III} for ITA3 with linkage-mapping. The M^{II} was indeed found in ITA3 samples also by FISH. Yet, we detected M^{Y} in the ITA1 sample and homozygous M^{X} in ITA3 that was not reported in previous study. This is likely due to small sample sizes in the prior study (only 4-9 tested males per strain), which may thus not be conclusive for the *M* locus composition at the strain level.

Discussion

We cytogenetically investigated the chromosomal locations of male determining (*M*) loci in laboratory and field strains of the house fly. By targeting the male determining gene, *Mdmd* (Sharma et al. 2017), *M* loci were localized on autosome II, autosome III and both X and Y chromosomes in Spanish and Dutch strains. Our results are generally consistent with prior linkage-mapping studies of these strains (this thesis, Chapter 2). An exception is Italian strains ITA1 and ITA3 in which *M* loci were additionally found on the X and the Y chromosome. These differences are likely due to the small sample sizes in the linkage mapping study, which may have led to undetected *M* loci that occur with low frequencies in these strains.

Hediger et al. (1998a) found male determining regions on both arms of the Y chromosome, M^{YS} (short arm) and M^{YL} (long arm), that separately could determine maleness. In our FISH study, only one *M* locus was located at the pericentromeric region of the short arm of the Y chromosome in the *aabys*, NL1 and ITA1 strains. The successful localization of M^{YS} with the *Mdmd* probe confirms the male determining

function of the *Mdmd* gene on the Y. M^{YL} was, however, not visualized with the *Mdmd* probe. One possible explanation is that M^{YL} contains low copy number of *Mdmd*, and thus could not be detected due to sensitivity limitations of the FISH method (like the undetected M^{V} locus, discussed below). An alternative explanation is that M^{YL} does not contain *Mdmd* sequences and its male determining function results from an alternative factor, perhaps similar to M^{I}. Intraspecific variation in primary male-determining genes in the house fly has been previously reported (Sharma et al. 2017).

M loci on both autosome II (in ITA3 and SPA4 strains) and autosome III (in M3, SPA3 and SAP5 strains) are located at the pericentromeric regions on the short arm, similar to the chromosomal location of M^{YS}. The centromere is functionally important for proper chromosome segregation during meiosis. Crossovers in the vicinity of centromeres negatively affect meiotic chromosome segregation (Koehler et al. 1996; Hassold and Hunt 2001; Rockmill et al. 2006). In organisms such as fission yeast and *Drosophila*, double-stranded breaks and centromere-proximal recombination are repressed to prevent these errors (Westphal and Reuter 2002; Ellermeier et al. 2010). The *Mdmd* gene has a long open reading frame that makes it more vulnerable to sequence disruption (Sharma et al. 2017). Therefore, translocations of *M* loci to pericentromeric regions on autosomes might be more stable as the chance of gene disruption as a consequence of crossovers is reduced in these regions of the chromosomes.

No intra-chromosomal differences in location of *M* loci were observed within and between strains. This means that an *M* locus on a chromosome is always found at a similar chromosomal location in samples from the same strain and from different strains. The M^{II} and the M^{III} loci are present at pericentromeric locations. The M^{X} locus is, however, located in the middle of one of the arms on the X chromosome. The lack of variation within chromosomal locations suggests that the *M* loci on the autosomes and the X chromosome each originated from a single translocation event and subsequently spread over Spanish and Italian regions. An alternative hypothesis is that translocations happened more frequently and that the observed chromosomal locations for non-M^{Y} loci are merely the result of selection, i.e. for residing in pericentromeric location where the *Mdmd* gene is sheltered from recombination. Note that in the current study, we only tested a limited number of individuals (2-3 per strain) from Spanish and Italian populations. More samples need to be investigated to confirm the absence of intra-chromosomal variation in *M* location. This is the more important as autosomal *M* loci have been found in house

fly populations from different continents (Tsukamoto et al. 1980; Franco et al. 1982; Tomita and Wada 1989; Hamm et al. 2005; Feldmeyer et al. 2008; Kozielska et al. 2008).

The M^V locus was not localized by FISH in the M5 strain, even though samples used in the experiments were confirmed by PCR to possess *Mdmd* sequences. This is most likely caused by the low number of *Mdmd* copies in M^V (Wu 2018) resulting in low light intensity of fluorescent signals leading to unsuccessful detection by microscopy. In the future, methods with high detection sensitivity such as Tyramide signal amplification-FISH (TSA-FISH) (Schriml et al. 1999; Carabajal Paladino *et al.* 2014) can be applied to increase the detection sensitivity. In addition, TSA-FISH could be used to test whether M^{YL} consists of few copies of *Mdmd*.

Interestingly, using the Mix probe, we localized *M* and sex chromosome located (MAS) repeats. These repeats cover a larger region on the X chromosome than on the Y chromosome indicating differences in copy number between the sex chromosomes. Previous studies have found that the X and the Y chromosomes in the house fly contain many repetitive sequences, of which some are unique to the sex chromosomes (Hediger et al. 1998b; Meisel et al. 2017). Using comparative genomics, Meisel *et al.* (2017) found differential abundance of sequences on the X and the Y chromosomes. They proposed that the morphological differences, with the X chromosome being larger than the Y chromosome, may be the result of these repetitive sequences. Our MAS repeat results are consistent with this prediction.

The *M* loci locate in close proximity to the MAS regions. As in *aabys* SPA1 and SPA4 male samples, the signals of MAS regions and *Mdmd* sequences are inseparable as only a single large signal is observed. This indicates that the MAS^Y and MAS^X regions are either in close proximity to M^X or M^Y, or the *Mdmd* sequences reside within the MAS regions on the sex chromosomes. In addition, the chromosomal locations of M^X and M^Y loci revealed by the *Mdmd* probe indicate the adjacency to the MAS regions. Furthermore, the Mix probe is generated from amplification of the M^{III} locus. Therefore, the sequence of the Mix probe that hybridized to the MAS regions must be present in the M^{III} locus as well. The current hypothesis is that the *Mdmd*-containing *M* locus originated on the Y chromosome and then, translocated to other chromosomes such as autosome III as large fragments (Sharma et al. 2017). Therefore, the reason that MAS sequence is found in the M^{III} locus is likely because that it got included during the translocation process.

The connection between the MAS regions and M^X and M^Y loci may provide a new perspective on the translocation of the *M* locus from the Y to the X chromosome. As MAS regions exist at the terminal part of the short arm of the Y chromosome and both ends of the X chromosome, the sequence homology of this region may promote homologous pairing and recombination. Considering that the MAS^Y region seems to be adjacent to, and presumably exists within, the M^{YS} locus, it is possible for the M^{YS} locus to have been included in an unequal crossing over event (non-allelic homologous recombination), resulting in the translocation of M^Y to M^X. As in many species, sex chromosomes not only possess sequences that are unique to either X or Y chromosome (or Z and W), they also include recombining regions (Otto et al. 2011; Raudsepp and Chowdhary 2015). According to this hypothesis, the M^X locus should be located within the MAS region on the X chromosome, which is likely to be the case based on our results. This can also explain why *M* on the X chromosome is located in the center region of one chromosome arm instead of the pericentromeric region on autosomes. Our hypothesis presumes high sequence identity between the M^{YS} and the M^X locus. This can be tested in future studies by comparing the genomic sequence of the M^{YS} and M^X loci. This hypothesis does not necessarily apply to the translocation events of autosomal *M* loci as large MAS regions were not observed on autosomes that do not possess the *M* locus. We can therefore not rule out additional mechanisms for translocation of *M* loci to autosomes.

In conclusion, the current study, cytogenetically demonstrates the chromosomal locations of different *M* loci in the house fly genome. M^{II}, M^{III} and M^Y loci are located near the centromeres suggesting a tendency for pericentromeric locations to harbor autosomal *M* loci. No intra-chromosomal variation in location was found for non-M^Y loci suggesting each of these M^A loci may have originated from a single translocation event and then spread to larger geographical regions. We additionally reported MAS on both the X and Y chromosome with different copy numbers. We hypothesize that non-allelic homologous recombination may have played a role in the translocation of *M* such as from the Y chromosome to the X chromosome. In later chapter 5, we present further results on *M* loci sequence composition and propose additional genomic processes responsible for the polymorphic male-determiners in the house fly. In Chapter 6, the possible scenarios and mechanisms of *M* evolution will be discussed.

Acknowledgments

We thank František Marec and Jan Keijser for technical advice and assistance.

References

Aljanabi SM, Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* **25**(22): 4692-4693.

Beukeboom LW, Perrin N. 2014. *The Evolution of Sex Determination*. Oxford University Press, USA.

Boyes J, Corey MJ, Paterson H. 1964. Somatic chromosomes of higher diptera: IX. Karyotypes of some muscid species. *Canadian Journal of Zoology* **42**(6): 1025-1036.

Carabajal Paladino LZ, Nguyen P, Šíchová J, Marec F. 2014. Mapping of single-copy genes by TSA-FISH in the codling moth, *Cydia pomonella*. *BMC Genetics* **15**(S2): S15.

Dübendorfer A, Hediger M, Burghardt G, Bopp D. 2002. *Musca domestica*, a window on the evolution of sex-determining mechanisms in insects. *International Journal of Developmental Biology* **46**(1): 75-79.

Denholm I, Franco M, Rubini P, Vecchi M. 1986. Geographical variation in house-fly (*Musca domestica* L.) sex determinants within the British Isles. *Genetics Research* **47**(1): 19-27.

Denholm I, Rubini P, Rovati C, Vecchi M. 1990. Genetic basis of sex determination in two South African strains of house-fly (*Musca domestica* L.). *South African Journal of Science* **86**(1): 41-43.

Ellermeier C, Higuchi EC, Phadnis N, Holm L, Geelhood JL, Thon G, Smith GR. 2010. RNAi and heterochromatin repress centromeric meiotic recombination. *Proceedings of the National Academy of Sciences* **107**(19): 8701-8705.

Feldmeyer B, Kozielska M, Kuijper B, Weissing FJ, Beukeboom LW, Pen I. 2008. Climatic variation and the geographical distribution of sex-determining mechanisms in the housefly. *Evolutionary Ecology Research* **10**(6): 797-809.

Franco M, Rubini P, Vecchi M. 1982. Sex-determinants and their distribution in various populations of *Musca domestica* L. of Western Europe. *Genetical Research* **40**(03): 279-293.

Hamm RL, Meisel RP, Scott JG. 2015. The evolving puzzle of autosomal versus Y-linked male determination in *Musca domestica*. *G3: Genes| Genomes| Genetics* **5**(3): 371-384.

Hamm RL, Shono T, Scott JG. 2005. A cline in frequency of autosomal males is not associated with insecticide resistance in house fly (Diptera: Muscidae). *Journal of Economic Entomology* **98**(1): 171-176.

Hassold T, Hunt P. 2001. To err (meiotically) is human: the genesis of human aneuploidy. *Nature Reviews Genetics* **2**(4): 280-291.

Hediger M, Minet AD, Niessen M, Schmidt R, Hilfiker-Kleiner D, Çakır Ş, Nöthiger R, Dübendorfer A. 1998a. The male-determining activity on the Y chromosome of the housefly (*Musca domestica* L.) consists of separable elements. *Genetics* **150**(2): 651-661.

Hediger M, Niessen M, Müller-Navia J, Nöthiger R, Dübendorfer A. 1998b. Distribution of heterochromatin on the mitotic chromosomes of *Musca domestica* L. in relation to the activity of male-determining factors. *Chromosoma* **107**(4): 267-271.

Hiroyoshi T. 1964. Sex-limited inheritance and abnormal sex ratio in strains of the housefly. *Genetics* **50**(3): 373.

Huang X, Hu X, Hu J, Zhang L, Wang S, Lu W, Bao Z. 2007. Mapping of ribosomal DNA and $(TTAGGG)_n$ telomeric sequence by FISH in the bivalve *Patinopecten yessoensis* (Jay, 1857). *Journal of Molluscan Studies* **73**(4): 393-398.

Koehler KE, Hawley RS, Sherman S, Hassold T. 1996. Recombination and nondisjunction in humans and flies. *Human Molecular Genetics* **5**(Supplement_1): 1495-1504.

Kozielska M, Feldmeyer B, Pen I, Weissing FJ, Beukeboom LW. 2008. Are autosomal sex-determining factors of the housefly (*Musca domestica*) spreading north? *Genetics Research* **90**(02): 157-165.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey

AR, Potter SC, Finn RD. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* **47**(W1): W636-W641.

Meisel RP, Gonzales CA, Luu H. 2017. The house fly Y chromosome is young and minimally differentiated from its ancient X chromosome partner. *Genome Research* **27**(8): 1417-1426.

Otto SP, Pannell JR, Peichel CL, Ashman T-L, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF. 2011. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends in Genetics* **27**(9): 358-367.

Raudsepp T, Chowdhary BP. 2015. The eutherian pseudoautosomal region. *Cytogenetic and Genome Research* **147**(2-3): 81-94.

Rockmill B, Voelkel-Meiman K, Roeder GS. 2006. Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in *Saccharomyces cerevisiae*. *Genetics* **174**(4): 1745-1754.

Schriml L, Padilla-Nash H, Coleman A, Moen P, Nash W, Menninger J, Jones G, Ried T, Dean M. 1999. Tyramide signal amplification (TSA)-FISH applied to mapping PCR-labeled probes less than 1 kb in size. *Biotechniques* **27**(3): 608-613.

Sharma A, Heinze SD, Wu Y, Kohlbrenner T, Morilla I, Brunner C, Wimmer EA, van de Zande L, Robinson MD, Beukeboom LW. 2017. Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene *CWC22*. *Science* **356**(6338): 642-645.

Tomita T, Wada Y. 1989. Multifactorial sex determination in natural populations of the housefly (*Musca domestica*) in Japan. *The Japanese Journal of Genetics* **64**(5): 373-382.

Tsukamoto M, Shono T, Horio M. 1980. Autosomal sex-determining system of the housefly: Discovery of the first-chromosomal male factor in Kitakyushu, Japan. *Journal of UOEH* **2**(2): 235-252.

Wagoner DE. 1967. Linkage group—karyotype correlation in the house fly determined by cytological analysis of X-ray induced translocations. *Genetics* **57**(3): 729.

Westphal T, Reuter G. 2002. Recombinogenic effects of suppressors of position-effect variegation in *Drosophila*. *Genetics* **160**(2): 609-621.

Wu Y. 2018. *Characterisation of the M-locus and Functional Analysis of the Male-determining Gene in the Housefly*. University of Groningen.

Supplementary material – Chapter 3

Table S3.1 Primer sequences used for amplification of *Mdmd* copies, *Mdmd* clone and intergenic sequences among *Mdmd* copies

Primer name	Sequence (5'-3')
Mdmd-FISHs	GGAAGTCGTATTGGAAGTAGT
Mdmd-FISHa	ATTTGGTGCGCCCTTCT
pGEM-Ts	GTAAAACGACGGCCAGT
pGEM-Ta	GGAAACAGCTATGACCATG
Mdmd_1as	GATTGGCTCAGATCGGCGTA
Mdmd_6as	GGTTGACGCGGACAATCAAC

Table S3.2 The length† of MAS regions (in μm) on the Y and X chromosomes of individual cells. MAS^X is the total length of MAS regions on the X chromosome. Ratios are MAS^Y length/MAS^X length.

No.	MAS^Y	MAS^{X1}	MAS^{X2}	MAS^X	Ratio
1	0.85	1.35	1.00	2.35	0.36
2	1.04	1.75	1.37	3.12	0.33
3	0.64	1.12	0.52	1.64	0.39
4	1.09	1.71	1.22	2.94	0.37
5	0.80	1.42	0.99	2.41	0.33
6	0.80	0.97	0.63	1.60	0.50
7	0.61	0.92	0.61	1.53	0.40
8	0.68	0.76	0.57	1.33	0.51
9	0.74	0.86	0.75	1.61	0.46
10	0.61	0.76	0.52	1.28	0.48
11	0.89	1.39	0.83	2.22	0.40
12	0.46	0.77	0.42	1.20	0.38
13	0.55	0.75	0.47	1.22	0.45
14	0.52	0.67	0.47	1.14	0.46
15	0.53	0.94	0.52	1.46	0.36
16	0.89	1.18	0.79	1.97	0.45
17	0.62	1.30	0.58	1.89	0.33
18	0.38	0.65	0.49	1.15	0.33
19	0.84	1.03	0.63	1.66	0.50
20	0.88	1.15	0.62	1.76	0.50
21	0.78	0.98	0.62	1.60	0.49
22	0.86	1.06	0.76	1.81	0.47
23	1.13	1.92	1.38	3.30	0.34
24	1.11	1.63	1.03	2.66	0.42
Average ratio					0.42

† As the signal lengths were initially measured with pixel length as unit. The actual length=signal length/scale bar length*10.

```
Mdmd        AGGGCAAAGGGAAGTCGTATTGGAAGTAGTACGTGATCTATCTTATACTATATGCAAGGA   420
Clone-5end  ----------------------------GTACAGTGATCTATCGTTATACTATATGCAAGGA    34
                                        *           ******************

Mdmd        AGCAGAAGAAAAATTAGTTGAAAGATTTCCTAGAAAGGATGGATCGAATGAAATGCTTCC   480
Clone-5end  AGCAGAAGAAAAATTAGTTGAAAGATTTCCTAGAAAGGATGGATCGAATGAAATGCTTCC    94
            ************************************************************

Mdmd        CAAGGAAGATAGCATAAATACTAACCATAATTACACAACGGATTCAAATGAACACCCAGT   540
Clone-5end  CAAGGAAGATAGCATAAATACTAACCATAATTACACAACGGATTCAAATGAACACCCAGT   154
            ************************************************************

Mdmd        AGAACTAACGACAAAGACAGAAGAATGTAAAAATACAGAGAAGACTAAAAAAAAATCTTT   600
Clone-5end  AGAACTAACGACAAAGACAGAAGAATGTAAAAATACAGAGAAGACTAAAAAAAAATCTTT   214
            ************************************************************

Mdmd        TGTGCGAGCTCTCTCCAAAGATAAACAACTATCCGCCTATAGAAGCCGTTCTAGAAGCAC   660
Clone-5end  TGTGCGAGCTCTCTCCAAAGATAAACAACTATCCGCCTATAGAAGCCGTTCTAGAAGCAC   274
            ************************************************************

Mdmd        ACGTTTAAGTTATTCTGGACACATTAGCCGAACTCATTCAGTAGAAAAAAGTCTATCAAG   720
Clone-5end  ACGTTTAAGTTATTCTGGACACATTAGCCGAACTCATTCAGTAGAAAAAAGTCTATCAAG   334
            ************************************************************

Mdmd        ATATAAAAAAAGTGTATTAAGAAATCGAAGAACTTCCTTTGGACACGGCCGAGATTCTTC   780
Clone-5end  ATATAAAAAAAGTGTATTAAGAAATCGAAGAACTTCCTTTGGACACGGCCGAGATTCTTC   394
            ************************************************************

Mdmd        TACGACCAAAAGGTCTGTCTCCAGGGACAAGGACAATCGACTAAGACGAAGAATAGGATC   840
Clone-5end  TACGACCAAAAGGTCTGTCTCCAGGGACAAGGACAATCGACTAAGACGAAGAATAGGATC   454
            ************************************************************

Mdmd        TAGCAGAAGCCATACTAGATCTCACAGTCGATTTCGAAGGTCGGAAAAGAAGCTTCCATC   900
Clone-5end  TAGCAGAAGCCATACTAGATCTCACAGTCGATTTCGAAGGTCGGAAAAGAAGCTTCCATC   514
            ************************************************************

Mdmd        AAGATCTCCACGACGAATTCGTTCACAAGAAAGACGACATGAACGACGTCGTTCGATGTC   960
Clone-5end  AAGATCTCCACGACGAATTCGTTCACAAGAAAGACGACATGAACGACGTCGTTCGATGTC   574
            ************************************************************

Mdmd        GTCAGATTATGAAAGGATAGCGTTACGCCGATCTGAGCCAATCAAAAGAAGAGATAAAGA  1020
Clone-5end  GTCAGATTATGAAAGGATAGCGTTACGCCGATCTGAGCCAATCAAAAGAAGAGATAAAGA   634
            ************************************************************

Mdmd        TGAATTCTTCAAAAACAATAAGAAAGTATCTGGCGATATAAAAAAGGGAAAGGGGAACGA  1080
Clone-5end  TGAATTCTTCAAAAACAATAAGAAAGTATCTGGCGATATAAAAAAGGGAAAGGGGAACGA   694
            ************************************************************

Mdmd        TAATGGTACTGTAGCAGAACTCGAGGCCAAGATAACAGAACGTCAAAGGAAAAGTCTAGA  1140
Clone-5end  TAATGGTACTGTAGCAAAACTCGAGGCCAAGATAACAGAACGTCGAAGGAAAAGTCTAGA   754
            ********** ********** ********** **********

Mdmd        TATACTAACATCACGTACAGGCGGTGCTTGTTTAACACCCGATAAATTGCGTATGATACA  1200
Clone-5end  TATACTAACATCACATACAGGCGGTGCTTGTTTAACACCCGATAAATTGCGTATGATACA   814
            ********** **********************************************

Mdmd        AGCAGAAATTACAGACAAATCATCGGCTGCATATCAAAGTATAGCTCGGGAAGCTCTTAA  1260
Clone-5end  AGCAGAGATTACGTACAAATCATCGGCTGCATATCAAAGTATA----------          857
            ****** *****  ****************************
```

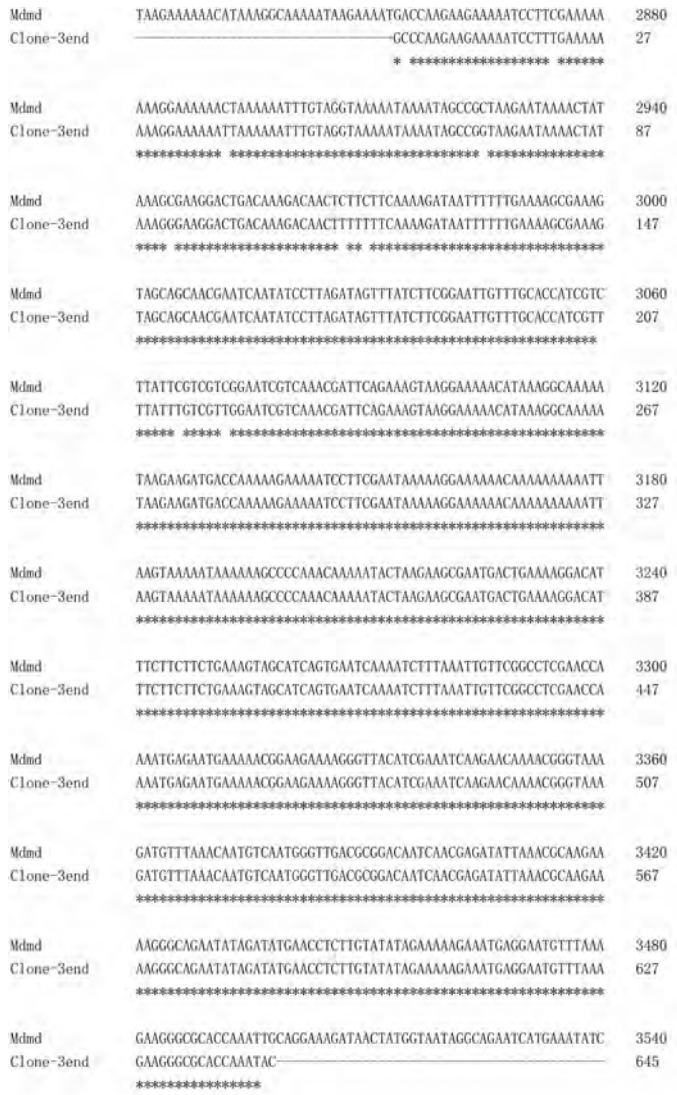

Figure S3.1 Nucleotide alignment of 5’ and 3’-end sequence of inserted fragment against the published *Mdmd* sequence in the clone containing *Mdmd* copy (Sharma *et al*. 2017). Numbers on the right indicate sequence positions.

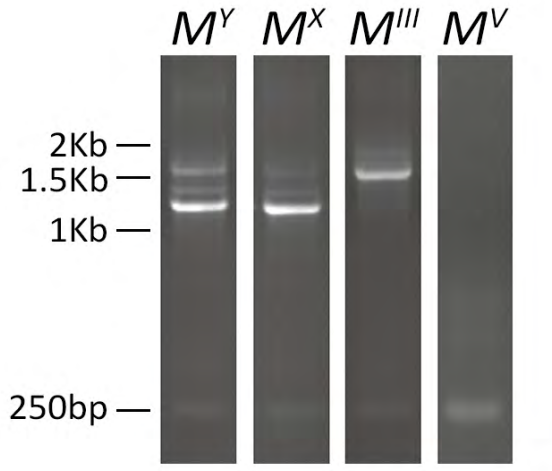

Figure S3.2 Banding pattern of intergenic sequences among *Mdmd* copies for different *M* loci. Multiple bands for M^{III}, M^{X} and M^{Y} indicate presence of multiple *Mdmd* copies. A single band for M^{V} suggests the number of *Mdmd* copies is lower than that of the other chromosomes.