Sex chromosome evolution in the house fly

Li, Xuan

DOI:
10.33612/diss.206102060

Thesis summary

The mechanisms that determine maleness and femaleness vary from one species to another, despite that sexual reproduction is ubiquitous. In most species only a single sex determination mechanism is present, but in some species, multiple systems co-occur. Such polymorphic systems are of interest for studying the evolution of sex determining systems and sex chromosomes. In house flies, a dominant male determining locus (*M*), that induces male development by interfering with the female determining factor *transformer* (*tra*), is not confined to the Y chromosome (M^Y) but can also be located on the X chromosome (M^X) as well as on any of the five autosomes ($M^{I\text{-}V}$). Previous studies revealed the complex sequence structures of the various *M* loci. The male determining gene *Musca domestica male determiner* (*Mdmd*) has been identified in M^{II}, M^{III}, M^V and M^Y loci, but not in the M^I locus, suggesting intraspecies variation exists for the male determining gene. Further characterization of the genomic sequences of the *M* locus demonstrated that multiple copies of the *Mdmd* gene exist within the *M* locus intervening by non-*Mdmd* sequences including repeats and transposable elements. The presence of *Mdmd* in different *M* loci together with the discovery of sequence similarity between the *M* loci on autosome III and autosome V led to the hypothesis that different *Mdmd*-containing *M* loci were the result of inter-chromosome translocation. To add to this complexity, a dominant allelic mutation of *tra*, termed tra^D, exists in nature that is insensitive to the interference of the *M* locus and directs female development even when the *M* locus is present in the genome. The distribution of the various sex determining factors follows a latitudinal gradient, with increasing complexity towards the equator. In this thesis, I have investigated this intriguing polymorphic sex determining system of the house fly at the genetic and genomic level to get a better understanding of its polymorphic sex determination system and the genomic processes underlying *M* locus evolution.

To obtain a more detailed understanding of the geographical distribution and the stability of polymorphic sex determining systems, I first analyzed the genomic localization of sex determining factors at the population level using five neighboring Spanish populations (SPA1-5; **Chapter 2**). I applied mapping crosses with a phenotypic marker-strain to determine chromosomal locations of *M* loci. I further examined the sex chromosome composition by karyotyping. Additionally, I determined tra^D allele frequencies of populations in sampled females by diagnostic

PCR. I found that the distribution of male and female determiners varied strongly between the populations, despite their small geographic distance. The results also revealed a correlation between the presence of the *tra*^{D} allele and the number and homozygosity of *M* loci in males, which is consistent with predictions and previous findings. I conclude that there are two stable states for these populations: (1) only one hemizygous *M* locus and no *tra*^{D} (male heterogametic system), and (2) with fixed *tra*^{D} in females (quasi female heterogametic system) and multiple *M* loci both homozygous or hemizygous. The strong differences in *M* locus and *tra*^{D} distribution between neighboring Spanish populations suggest that factors other than temperature and humidity may play a role in maintaining such polymorphisms.

To gain more insight into *M* locus translocation history at the chromosome level, I determined the physical location of *M* loci utilizing fluorescence *in situ* hybridization (FISH) (**Chapter 3**). I focused on the localization of *Mdmd*-containing *M* loci and designed an *Mdmd* specific probe for this purpose. I located *M* loci on the Y chromosome, autosome II and autosome III near the centromeres, suggesting a tendency for *M* loci to be located to pericentromeric regions. The *M* locus on the X chromosome was located in the middle of one chromosome arm. In addition, I compared the localization of *M* loci in populations of different geographical origins. No variation was found between populations from Spain and Italy for the chromosomal position of the *M* locus on the X chromosome, autosome II and autosome III. This supports the hypothesis that *M* locus translocation to each autosome and the X chromosome was a single event, and that these *M* locus-carrying chromosomes subsequently spread in Italian and Spanish regions. In addition to the *Mdmd* specific probe, I designed a "mixed" probe for FISH, containing not only *Mdmd* sequences but also non-*Mdmd* sequences of the *M* locus. Using this “mixed” probe, I localized the *M* locus but also discovered *M* locus And Sex chromosome (MAS) located repeats, which differ in copy number between the X and Y chromosomes. MAS repeats are present on the *M*-carrying Y and X chromosomes, but also on X chromosomes without *M* locus. Moreover, MAS repeat regions likely overlap with the chromosomal locations of *M*^{Y} and *M*^{X}. The sequence homology of MAS repeat regions may promote homologous pairing and recombination between the XY chromosomes. Considering that the *M* locus seems to be adjacent to, and presumably exists within, the MAS^{Y} region, I hypothesize that the translocation of the *M* locus from the Y to the X is a non-allelic homologous recombination event mediated by MAS sequences (**Chapter 6**).

To learn more about the genomic structure of the *M* locus and to gain insight into the dynamics that may have led to its structure, I assembled and annotated the entire M^{III} locus (**Chapter 4**). I assembled two *M* locus contigs that together make up the M^{III} locus with a size of ~591 kb. In the *M* locus reside 88 *Mdmd* replicates, of which only one contains the complete protein-coding sequence and thus is likely the functional *Mdmd* gene, whereas others have degenerated to various degrees. Other than *Mdmd*, the M^{III} locus contains few sequences of known genes, although it is debatable whether these non-*Mdmd* genes are functional. I found a high degree of sequence duplication throughout the *M* locus. The majority of the duplicated sequences are *Mdmd* replicates and their flanking sequences, however, repetitive sequences that did not contain *Mdmd* were also found. I propose that duplication of *Mdmd* and its flanking sequences may have a positive effect on sustaining this male determining region because they are *M* locus-specific and increase sequence divergence between the *M* locus region and the homologous chromosome. Thus, they can potentially promote recombination arrest and contribute to maintaining the hemizygous status of the *M* locus region. I further found putative evidence that double-strand breakage and homologous repair have played a role in sequence duplications within the *M* locus. I hypothesize that the same mechanisms were involved in the translocation of *M* loci from one chromosome to another (**Chapter 6**).

To gain more insight into *M* locus translocation mechanisms and the evolutionary history at the genomic level, I compared the genomic sequence differences between *M* loci on autosome II, autosome III, autosome V and the Y chromosome (**Chapter 5**). By mapping Illumina sequence datasets of different *M* loci to the *Mdmd* sequence and the M^{III} locus contigs, I first found that the copy number of *Mdmd* replicates varies between *M* loci. M^{III} appears to have the highest number of *Mdmd* replicates, whereas the M^{V} locus contains the fewest. I revealed that sequence variation also occurs between *M* loci on various Y chromosomes, suggesting that M^{Y} in different populations evolved independently and at a different speed. Comparison of interspersed sequences among different *M* loci reveals high similarity between the M^{III} locus and the M^{Y} locus of one strain, whereas the M^{II} locus shows higher sequence similarity to the M^{Y} loci in two other strains. Additionally, the genomic region that contains the *Mdmd* with the complete ORF sequences is universally present in all investigated *M* loci (M^{II}, M^{III}, M^{V} and M^{Y}), indicating the importance of this region for the male determining function. Based on these findings, I proposed three potential routes to describe the evolutionary history of *Mdmd*-containing *M* loci, of which the most plausible one is M^{Y} being the initially established *M* locus with

subsequent translocation to other chromosomes (**Chapter 6**).

In conclusion, I demonstrated that sex determining systems vary strongly at a regional level in house fly populations. These results suggest that additional factors other than previously proposed environmental factors play a role in maintaining this polymorphic system. I described how a series of "young" male determining loci are arranged at the genomic level. Based on comparison of chromosomal locations and sequence structures of various *M* loci I proposed possible genomic mechanisms of *M* locus translocation. I find evidence for degeneration by mutation accumulation in the *M* locus region, possibly as a result of recombination arrest, consistent with the canonical model of sex chromosome evolution. Based on my results, I propose several routes for the origin of the multiple *M* loci in the house fly genome. My study provides novel insights into population dynamics and genomic structure of polymorphic sex determiners in the house fly and contributes to a better understanding of the evolution of sex determining loci and sex chromosome evolution in general.