

University of Groningen

Sex chromosome evolution in the house fly

Li, Xuan

DOI:
[10.33612/diss.206102060](https://doi.org/10.33612/diss.206102060)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Li, X. (2022). *Sex chromosome evolution in the house fly: Genomic distribution and structure of polymorphic male determining loci*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.206102060>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Samenvatting

De mechanismen die bepalen of een organisme mannelijk of vrouwelijk wordt, variëren van de ene soort tot de andere, ondanks dat seksuele voortplanting vaak voorkomt. In de meeste soorten is er slechts één geslachtsbepalingsmechanisme, maar in sommige soorten kunnen meerdere geslachtsbepalingsmechanismen tegelijkertijd voorkomen. Dit soort polymorfische systemen zijn uiterst geschikt om de evolutie van geslachtsbepalingssystemen en geslachtschromosomen te bestuderen. Huisvliegen hebben een dominant locus (het *M* locus), dat de mannelijke ontwikkeling induceert door de vrouwelijke factor *transformer* (*tra*) te onderdrukken. Dit *M* locus is niet alleen te vinden op het Y chromosoom (M^Y), maar kan ook op het X chromosoom (M^X) of op elk van de vijf autosomen (M^{I-V}) liggen. Eerder onderzoek heeft uitgewezen dat de structuur van de verschillende *M* loci complex is. Het gen dat mannelijke ontwikkeling bepaalt, *Musca domestica male determiner* (*Mdmd*) is geïdentificeerd in de M^I , M^{III} , M^V en M^Y loci, maar niet in het M^I locus wat suggereert dat er naast *Mdmd* ook nog andere genen in de huisvlieg bestaan die mannelijkheid kunnen bepalen. Verder onderzoek van de genomische sequenties van het *M* locus hebben laten zien dat er meerdere kopieën van het *Mdmd* gen zijn met daartussen niet-*Mdmd* sequenties zoals, onder andere, repetitieve elementen en transposons. De aanwezigheid van *Mdmd* in de verschillende *M* loci, in combinatie met de ontdekking van vergelijkbare sequenties tussen de *M* loci op autosoom III en autosoom V hebben geleid tot de hypothese dat de verschillende *Mdmd*-bevattende *M* loci het resultaat zijn van translocatie van het *M* locus tussen verschillende chromosomen. Bovenop deze complexiteit is er ook nog een dominant allel van *tra*, genaamd *tra^D*, dat van nature voorkomt en dat ongevoelig is voor het *M* locus waardoor een huisvlieg die dit allel draagt zich ontwikkeld tot vrouwtje zelfs in de aanwezigheid van het *M* locus. De geografische verspreiding van de verschillende geslachtsbepalingssystemen volgt een breedtegraadgradiënt, met een hogere complexiteit richting de evenaar. In dit proefschrift heb ik het intrigerende polymorfische geslachtsbepalingssysteem van de huisvlieg bestudeerd op zowel genetisch als genomisch niveau om een beter begrip te krijgen van dit complexe systeem en de genomische processen die ten grondslag liggen aan de evolutie van het *M* locus.

Om een gedetailleerd overzicht te krijgen van de geografische distributie en de stabiliteit van het polymorfische geslachtsbepalingssysteem heb ik eerst de

Samenvatting

genomische locatie van de geslachtsbepalingsfactoren geanalyseerd op populatieniveau door gebruik te maken van 5 nabijgelegen Spaanse populaties (SPA1-5; **Hoofdstuk 2**). Ik heb kruisingen uitgevoerd tussen deze populaties en een fenotypische “marker” populatie om de chromosomale locaties van de M loci te identificeren. Ook heb ik de karyotypes bepaald van vliegen afkomstig uit deze populaties om te zien welke geslachtschromosomen er aanwezig zijn. Verder heb ik de allelfrequentie van het tra^D allel in de populaties bepaald door middel van PCR. De distributie van de mannelijke en vrouwelijke geslachtsbepalingssystemen varieerden sterk tussen populaties, ondanks de kleine geografische afstand tussen de populaties. De resultaten lieten ook zien dat er een correlatie is tussen de aanwezigheid van het tra^D allel en het aantal, en de homozygositeit, van de M loci in de populaties. Dit is in lijn met de verwachtingen en komt overeen met eerdere bevindingen. Ik concludeer dat er twee stabiele staten zijn voor deze populaties: (1) één hemizygoot M locus en geen tra^D allel (mannelijk heterogametisch systeem), en (2) een gefixeerd tra^D allel in vrouwtjes (quasi vrouwelijk heterogametisch systeem) en meerdere M loci, zowel homozygoot als hemizygoot. De grote verschillen in M locus en tra^D allel distributie tussen nabijgelegen Spaanse populaties wijst erop dat er andere factoren zijn dan temperatuur en luchtvochtigheid die het polymorfisme in geslachtsbepalingssysteem in stand houden.

Om meer inzicht te krijgen in de translocatiegeschiedenis van het M locus heb ik de fysieke locatie van de M loci bepaald met behulp van fluorescentie-in-situhybridisatie (FISH) (**Hoofdstuk 3**). Hiervoor heb ik een *Mdmd*-specifieke probe ontworpen om *Mdmd*-bevattende M loci te kunnen detecteren. Op het Y chromosoom, autosoom II en autosoom III zijn de M loci dichtbij het centromeer te vinden wat suggereert dat M loci vaker te vinden zijn in de pericentromerische regio's van de chromosomen. Het M locus op het X chromosoom is echter te vinden in het midden van één van de chromosoomarmen. Verder heb ik de locaties van de M loci die op hetzelfde chromosoom liggen, maar die afkomstig zijn van geografisch diverse populaties, vergeleken. Daarbij vond ik geen verschillen tussen de locatie van M loci op het X chromosoom, autosoom II en autosoom III van Spaanse en Italiaanse populaties. Dit bevestigt de hypothese dat de translocatie van het M locus naar elk chromosoom één keer heeft plaatsgevonden en dat deze M locus-dragende chromosomen vervolgens zijn verspreid in Italiaanse en Spaanse regio's. Ik heb ook een gemengde probe ontworpen die zowel *Mdmd* en niet-*Mdmd* sequenties van het M locus bevat. Met behulp van deze gemengde probe heb ik M loci gedetecteerd, maar heb ik ook ontdekt dat er regio's met repetitieve elementen, genaamd “ M locus And Sex

chromosome (MAS) repeat”, bestaan die zowel op het X chromosoom als het Y chromosoom liggen, maar in verschillende aantallen. MAS regio’s zijn zowel aanwezig op *M*-dragende Y en X chromosomen als op X chromosomen zonder *M* locus. Verder overlapt de locatie van MAS repetitieve elementen waarschijnlijk met de locatie van het *M* locus op deze chromosomen. De homologie van de MAS regio’s tussen de geslachtschromosomen stimuleert mogelijk paarfomatie en recombinatie tussen de X en Y chromosomen. Aangezien het *M* locus grenst aan en mogelijk zelfs bestaat binnen het MAS^Y locus, hypotheetiseer ik dat de translocatie van het *M* locus van het Y chromosoom naar het X chromosoom heeft plaatsgevonden door middel van niet-allelische homologe recombinatie via de MAS sequentie (**Hoofdstuk 6**).

Om meer te leren over de genomische structuur van het *M* locus en om meer inzicht te krijgen in de mechanismen die hebben geleid tot de structuur van het *M* locus, heb ik het gehele *M*^{III} locus geassembleerd en geannoteerd (**Hoofdstuk 4**). De uiteindelijke genoomassemblage bevat twee *M* locus contigs die samen het ~591 kb *M*^{III} locus vormen. Het *M* locus bevat 88 *Mdmd* kopieën waarvan er slechts één de complete eiwit-coderende sequentie bevat en daarom waarschijnlijk het functionele *Mdmd* gen is. De andere kopieën zijn in verschillende mate gedegeneerd. Buiten *Mdmd* sequenties om, bevat het *M*^{III} locus ook enkele andere sequenties van bekende genen, al is het niet duidelijk of deze andere genen functioneel zijn. Het gehele *M* locus bevat een groot aantal gedupliceerde sequenties. Het grootste deel hiervan zijn de *Mdmd* kopieën en de direct omliggende sequenties, maar er zijn ook repetitieve sequenties te vinden zonder *Mdmd*. Ik hypotheetiseer dat de duplicaties van *Mdmd* en de omliggende sequenties mogelijk een positief effect hebben op het behouden van de mannelijke ontwikkelende regio omdat ze specifiek zijn voor het *M* locus en daarmee divergentie tussen het *M* locus en het homologe chromosoom stimuleren. Daarbij kan deze divergentie potentieel een rol spelen bij het reduceren van recombinatie en daarmee bijdragen aan het in stand houden van de hemizygote staat van de *M* locus regio. Verder heb ik sequentie restanten gevonden die erop wijzen dat “double strand breakage” en homologe reparatie een rol hebben gespeeld in de duplicaties van sequenties binnen het *M* locus. Ik hypotheetiseer dat ditzelfde mechanisme mogelijk betrokken is bij de translocatie van de *M* loci tussen chromosomen (**Hoofdstuk 6**).

Om meer inzicht te krijgen in hoe het *M* locus zich heeft verplaatst naar de verschillende chromosomen en in de evolutionaire relatie tussen de verschillende *M* loci, heb ik de sequentie verschillen tussen *M* loci op autosoom II, autosoom III,

Samenvatting

autosoom V, en het Y chromosoom vergeleken (**Hoofdstuk 5**). Door Illumina datasets van verschillende *M* loci te vergelijken met de sequentie van het *Mdmd* gen, heb ik gevonden dat er een verschil in *Mdmd* kopieën is tussen de verschillende *M* loci. Het M^{III} locus blijkt het grootste aantal *Mdmd* kopieën te hebben en M^V de minste. Ook observeerde ik dat er verschillen in het aantal *Mdmd* kopieën zijn tussen M^V datasets van verschillende populaties wat suggereert dat M^V in verschillende populaties onafhankelijk evolueert. Het mappen van dezelfde datasets tegen het M^{III} locus liet zien dat het M^V locus van één populatie het meest lijkt op het M^{III} locus, terwijl het M^{II} locus meer lijkt op het M^V locus van twee andere populaties. De genomische regio waarin het complete *Mdmd* gen, met intact eiwit-coderende sequentie, ligt, is aanwezig in alle onderzochte *M* loci (M^{II} , M^{III} , M^V en M^Y). Dit laat zien dat deze regio van het *M* locus belangrijk is voor het induceren van mannelijke ontwikkeling. Gebaseerd op de bevindingen, presenteer ik drie potentiële routes die de evolutionaire geschiedenis van het *Mdmd*-dragende *M* locus kunnen beschrijven. De meest waarschijnlijke route stelt dat het M^V locus het originele *M* locus is waarna het *M* locus via translocatie op de andere chromosomen terecht is gekomen (**Hoofdstuk 6**).

Ter conclusie, ik heb laten zien dat geslachtsbepalingssystemen sterk kunnen variëren op regionaal niveau in huisvliegpopulaties. Deze resultaten suggereren dat er aanvullende factoren zijn, naast de eerder voorgestelde omgevingsfactoren, die een rol spelen in het behouden van het polymorfische geslachtsbepalingssysteem van de huisvlieg. Ik heb de genomische organisatie beschreven van een aantal “jonge” *M* loci en, gebaseerd op een vergelijking van de verschillende *M* loci, presenteer mogelijke mechanismen hoe het *M* locus zich tussen chromosomen verplaatst heeft. Ik heb indicaties gevonden voor degeneratie, door middel van accumulatie van mutaties in het M^{III} locus. Dit kan wijzen op gereduceerde recombinatie in het *M* locus, wat overeen komt met de verwachting naar aanleiding van het algemeen geaccepteerde model van geslachtschromosoomevolutie. Gebaseerd op de resultaten presenteer ik verschillende routes voor het ontstaan van meerdere *M* loci in het huisvliegengenoom. Mijn onderzoek heeft bijgedragen aan nieuwe inzichten met betrekking tot de populatie dynamiek en genoomstructuur van polymorfische geslachtsbepalingssystemen in de huisvlieg en draagt bij aan een beter begrip van de evolutie van geslachtsbepalingsloci en geslachtschromosoomevolutie in het algemeen.