university of groningen

University Medical Center Groningen

University of Groningen

Male sex in houseflies is determined by Mdmd, a paralog of the generic splice factor gene CWC22

Sharma, Akash; Heinze, Svenia D; Wu, Yanli; Kohlbrenner, Tea; Morilla, Ian; Brunner, Claudia; Wimmer, Ernst A; van de Zande, Louis; Robinson, Mark D; Beukeboom, Leo W

Published in:
Science

DOI:
10.1126/science.aam5498

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:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Sharma, A., Heinze, S. D., Wu, Y., Kohlbrenner, T., Morilla, I., Brunner, C., Wimmer, E. A., van de Zande, L., Robinson, M. D., Beukeboom, L. W., & Bopp, D. (2017). Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene *CWC22*. *Science*, *356*(6338), 642-645. https://doi.org/10.1126/science.aam5498

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.

SEX DETERMINATION

Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene *CWC22*

Akash Sharma,[1,2]* Svenia D. Heinze,[2]* Yanli Wu,[1,3] Tea Kohlbrenner,[2] Ian Morilla,[2]† Claudia Brunner,[2] Ernst A. Wimmer,[3] Louis van de Zande,[1] Mark D. Robinson,[2,4] Leo W. Beukeboom,[1] Daniel Bopp[2]‡

Across species, animals have diverse sex determination pathways, each consisting of a hierarchical cascade of genes and its associated regulatory mechanism. Houseflies have a distinctive polymorphic sex determination system in which a dominant male determiner, the M-factor, can reside on any of the chromosomes. We identified a gene, *Musca domestica male determiner* (*Mdmd*), as the M-factor. *Mdmd* originated from a duplication of the spliceosomal factor gene *CWC22* (*nucampholin*). Targeted *Mdmd* disruption results in complete sex reversal to fertile females because of a shift from male to female expression of the downstream genes *transformer* and *doublesex*. The presence of *Mdmd* on different chromosomes indicates that *Mdmd* translocated to different genomic sites. Thus, an instructive signal in sex determination can arise by duplication and neofunctionalization of an essential splicing regulator.

Genetic mechanisms for sex determination are not conserved among organismal groups. Illustrating this diversity, in insects, systems vary among species at the chromosomal, gene, and gene-regulation level (*1*–*3*). For example, in insects with male heterogamety (XX-XY system), sex can be determined by a dominant Y-linked gene or by X-chromosome dosage (*4*). The *transformer* (*tra*) and *doublesex* (*dsx*) genes are conserved elements of the insect sex determination pathway, but the upstream instructive signals vary (*5*–*7*). The polymorphic sex determination system of the housefly, *Musca domestica*, reflects this diversity in regulation and genes (*8*–*11*). Males can carry a dominant male determiner (M-factor) on the X or Y chromosome or any of the five autosomes (*10*, *12*–*14*).

The M-factor acts as the instructive signal for male development in the housefly. It regulates *transformer* (*Md-tra*), a binary switch that directs female differentiation when active and male differentiation when inactive. *Md-tra* is regulated at the splicing level. The active state of *Md-tra* is initially established by maternally provided *Md-tra*. Once activated, zygotic *Md-tra* will perpetuate its female-promoting function by a positive splicing feedback loop throughout development. Paternally inherited M-factor prevents this maternal activation of the zygotic *Md-tra* self-regulatory loop. The early embryonic presence of male-specific splice products of *Md-tra* indicates that this regulation has already started at the cellular blastoderm stage (*11*).

We hypothesized that the M-factor gene encodes a product present only in early male embryos to prevent establishment of *Md-tra* function. Exploiting *Musca* genetics, we isolated and sequenced RNA from unisexual embryos (fig. S1). Among the top 14 male-specifically expressed sequences that were absent in the female *M. domestica* genome assembly (*15*), we identified five orphan contigs of the same transcription unit (Fig. 1A and table S1), which we termed *Mdmd* (for *Musca domestica male determiner*). Subsequent analysis revealed that these sequences are present in males that carry an M-factor on chromosome Y, II, III, or V (Fig. 1B). Reverse transcription polymerase chain reaction (RT-PCR) amplification confirmed the exclusive presence of *Mdmd* transcripts in male embryos (Fig. 1C). Zygotic *Mdmd* transcripts first appear in 2- to 3-hour-old embryos (cellularized blastoderm stage), coinciding with the first zygotic expression of *Md-tra* (*11*). *Mdmd* expression is then maintained throughout male development until adulthood (Fig. 1D). *Mdmd* encodes a protein with high homology to CWC22 (complexed with Cef-1), also known as NCM (nucampholin), a spliceosome-associated protein that is required for precursor mRNA splicing and exon junction complex (EJC) assembly (*16*) (Fig. 1E). A BLAST survey of *Mdmd* over female genome scaffolds (*15*) identified a paralog (LOC101896466) of *Mdmd* that is structurally closely related to *ncm* genes of other insect species. In contrast to *Mdmd, ncm* is present and expressed in both sexes (Fig. 1, B to D). On the basis of its high sequence identity to the *ncm* gene in *Drosophila* and its conserved synteny evidenced by linkage to *bicoid stability factor*, we refer to this autosomal gene as the ortholog of *ncm* and name it *Md-ncm*. Mdmd shares a high degree of amino acid identity with Md-NCM in the MIF4G (85%) and MA3 (79%) domains and flanking sequences, but it displays a substantial level of divergence in the N-terminal and C-terminal regions (Fig. 1A and fig. S2). Sequence alignments reveal that *Md-ncm* groups with prototype *ncm* genes of other insect species. However, the *Mdmd* sequences from different *M. domestica* strains form a distinct outgroup, suggesting that after the duplication event, *Mdmd* rapidly diverged from *Md-ncm* (Fig. 1E and fig. S3).

Multiple nonfunctional copies of *Mdmd* were found next to the *Mdmd* gene in the genome of the M^{III} (M-factor located on chromosome III) strain (fig. S4). These copies may have arisen from local amplification to preserve *Mdmd* functionality in a nonrecombining region (fig. S4). Because of its long open reading frame (ORF), *Mdmd* is particularly vulnerable to the accumulation of deleterious mutations. We identified a similar arrangement of multiple *Mdmd* copies in M^{II}, M^{V}, and M^{Y} males (fig. S4). This suggests that the various M loci originated from a common ancestral *Mdmd* sequence, which first locally multiplied and then translocated as a cluster to different sites in the genome (fig. S5).

On silencing of *Mdmd* by injecting double-stranded RNA (dsRNA) into syncytial embryos of different M strains, all of the surviving M-factor–carrying individuals developed externally as males, but 56 to 88% contained fully differentiated ovaries instead of testes, with the notable exception of M^{I} males (Fig. 2, A to C, and fig. S6). From this result, we infer that *Mdmd* is essential for specifying the male gonadal and germline fate, which is consistent with genetic findings that M-factor and its target *Md-tra* govern the sexual identity of both soma and germ line (*11*). Incomplete feminization may be explained by the transient nature of embryonic RNA interference. A 70% reduction of *Mdmd* transcript levels was observed in $M^{III}/+$ embryos 10 hours after dsRNA injection, whereas after 20 hours, levels were comparable to those in control individuals, suggesting a recovery of *Mdmd* expression (Fig. 2, D and E). Given that substantial levels of *Mdmd* transcripts were also detected in nongonadal tissues of male adults with ovaries, restored activity of *Mdmd* at late stages apparently prevented systemic female differentiation (Fig. 2E). To conclusively test whether *Mdmd* is required for overall male differentiation, loss-of-function alleles were generated in *Mdmd* coding sequences by nonhomologous end joining–mediated disruption with Cas9. On targeting *Mdmd* in the M^{III} strain, we recovered 59 fertile males, of which at least 10 sired female progeny carrying dominant markers tightly linked to the M^{III} locus, indicating loss of its male-determining function (fig. S7). These M-factor–containing individuals are phenotypically normal fertile females (Fig. 3A). Sequence analysis confirmed that these females carry structural aberrations in the *Mdmd* cluster (Fig. 3B). Lines M32 and M36 are most informative because the lesions

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Post Office Box 11103, 9700 CC, Groningen, Netherlands. [2]Institute of Molecular Life Sciences, University of Zurich, CH-8057 Zurich, Switzerland. [3]Department of Developmental Biology, Johann-Friedrich-Blumenbach Institute of Zoology and Anthropology, Göttingen Center of Molecular Biosciences, University of Goettingen, Göttingen, Germany. [4]SIB Swiss Institute of Bioinformatics, University of Zurich, Zurich, Switzerland.
*These authors contributed equally to this work. †Present address: Laboratorie Inflamex, LAGA-INSERM UMR 1149, Université Paris 13 Sorbonne, Paris, France. ‡Corresponding author. Email: daniel.bopp@imls.uzh.ch

Downloaded from http://science.sciencemag.org/ on May 14, 2017

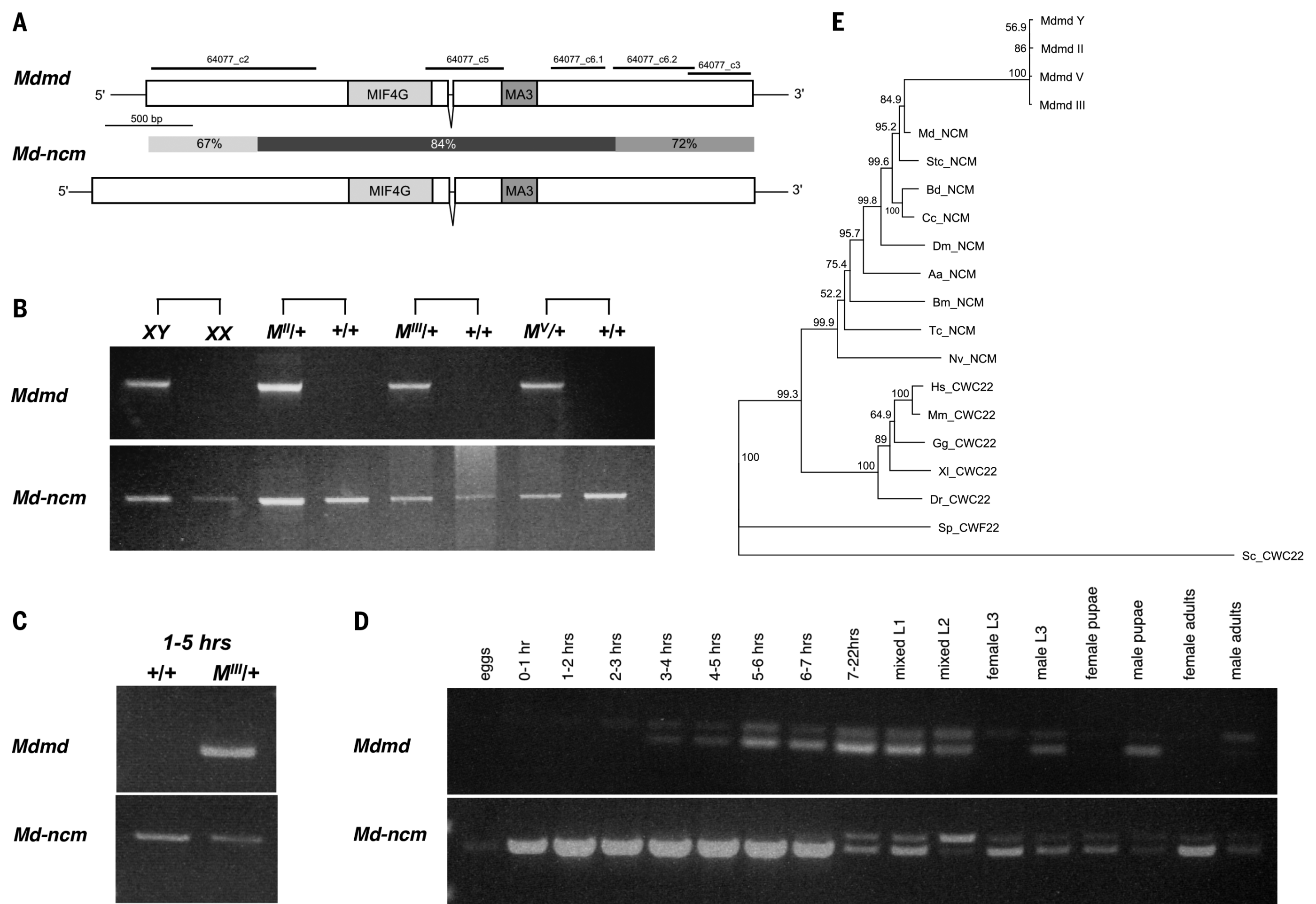


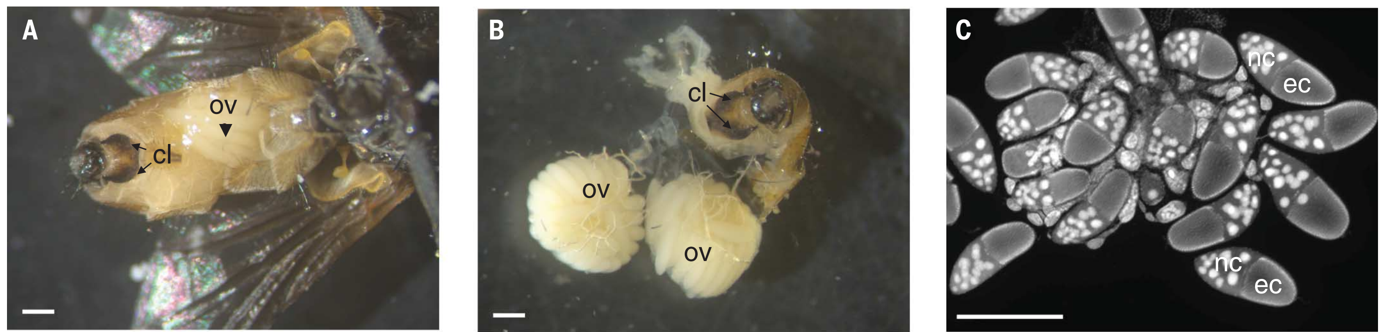
Fig. 1. *Mdmd* is a male-specific paralog of the *M. domestica CWC22* ortholog, *Md-ncm*. (**A**) Comparison of the two paralogs, *Mdmd* and *Md-ncm*. *Mdmd* was initially identified by extending five male-specific RNA contigs (lines above *Mdmd*). Each exon contains highly conserved MIF4G and MA3 domains. Nucleotide identity is indicated in percentages. (**B**) Genomic amplifications with paralog-specific end primers show that *Mdmd* sequences are present only in males of XY, M^{II}, M^{III}, and M^{V} strains, whereas *Md-ncm* is present in both males and females (+/+) of each strain. (**C**) RT-PCR confirming the presence of *Mdmd* transcripts in 1- to 5-hour-old male embryos (M^{III}/+) but not in female embryos (+/+). (**D**) Developmental expression profiles of *Mdmd* and *Md-ncm* based on RT-PCR with intron-spanning primers. The upper bands in both profiles correspond to unspliced RNA and/or genomic DNA contamination, whereas the lower bands represent spliced transcripts. (**E**) Neighbor-joining phylogenetic tree (branch label, percent consensus support) for Mdmd and NCM/CWC22 proteins (see also the MrBayes tree in fig. S3B). Full species names and sequences are listed in the supplementary materials. L, larval stage; bp, base pairs.

specifically disrupt the ORF of *Mdmd* (Fig. 3C) and only abolish the protein-coding function of this *Mdmd* copy. We conclude that the *Mdmd* gene is indispensable for normal male development and may be the only gene in the cluster providing male function. Consistent with the role of M-factor as an upstream repressor of *Md-tra*, individuals that have *Mdmd* abolished by CRISPR-Cas9 exclusively express the female splice variants of *Md-tra* and *Md-dsx* (Fig. 3D).

On the basis of sequence similarity, we infer that *Mdmd* is a paralog of *Md-ncm* (*CWC22*), which, as noted, encodes a spliceosome-associate protein that is indispensable for the assembly of the EJC (*16*, *17*). The essential functions of *CWC22* are likely to be provided by *Md-ncm*, given that embryonic silencing of this gene leads to early lethality in both males and females (fig. S8). However, the effect of EJC on splicing is limited to certain genes (*18*). Changes in expression levels of EJC components also affect the splice site selection of alternatively spliced genes (*19*). Considering that *tra* is one of the targets on which EJCs preferentially assemble in *Drosophila* (*18*), it is conceivable that *Md-ncm* plays a crucial role in the splicing regulation of *Md-tra*. Because the target of M-factor, *Md-tra*, is alternatively spliced, this posttranscriptional regulatory function makes *Mdmd* an excellent candidate M-factor. *Mdmd* may act as a direct regulator of *Md-tra* by selectively promoting the male or preventing the female splicing mode. Alternatively, the high level of sequence similarity to its paralog opens the possibility that *Mdmd* behaves as a dominant-negative, interfering with the functions of *Md-ncm* in promoting female splicing of *Md-tra*. Further study needs to elucidate the precise role of this gene in *Md-tra* splicing and can contribute to a better understanding of alternative splicing regulation.

There likely exists a large source of primary signal genes, which contribute to the high diversity of sex-determining mechanisms in insects. Recently, two male determiners were characterized in mosquitoes, *Nix* in *Aedes aegypti* (*20*) and *Yob* in *Anopheles gambiae* (*21*). These genes show sequence homology neither to each other nor to *Mdmd*, further pointing toward the species-specific acquisition of novel male determiners in insects. Moreover, *Mdmd* appears to be absent in the *M. domestica* strain that has an M-factor mapped to chromosome I (fig. S4), suggesting that even intraspecific variation exists at the level of the primary signal. Because insect sex determination is based on alternative splicing, its role in splicing regulation may have pre-equipped *ncm* for attaining a sex determination function. The recruitment of a *CWC22* duplicate for male function may be unique to the housefly, given that *ncm* paralogs have thus

Downloaded from http://science.sciencemag.org/ on May 14, 2017



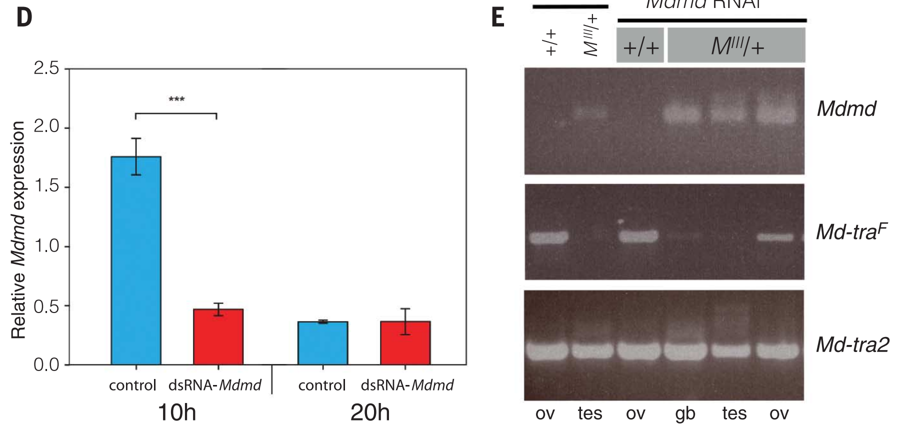


Fig. 2. Embryonic silencing of *Mdmd* is transient and leads to ovarian differentiation in males. In (A) to (C), $M^{III}/+$ individuals were injected with dsRNA against *Mdmd*. Scale bars, 1 mm. (**A**) Adult abdomen with male genital structures [claspers (cl) indicated by arrows] and, inside, fully differentiated ovaries (ov, arrowhead). (**B**) Dissected ovaries from the same male. (**C**) DAPI (4′,6-diamidino-2-phenylindole)–stained ovaries containing normal cysts composed of nurse cells (nc) and egg chambers (ec). (**D**) Relative levels of *Mdmd* mRNA 10 and 20 hours (h) after injections with dsRNA against *Mdmd* and dsRNA against *M112* control in $M^{III}/+$ male embryos. (**E**) RT-PCR analysis of *Mdmd* transcripts and female transcripts of *Md-tra* (*Md-tra*^{F}) in normal (+/+) ovaries and ($M^{III}/+$) testes (tes) and in *Mdmd* dsRNA-injected ($M^{III}/+$) gonadectomized bodies (gb), testes, and ovaries.

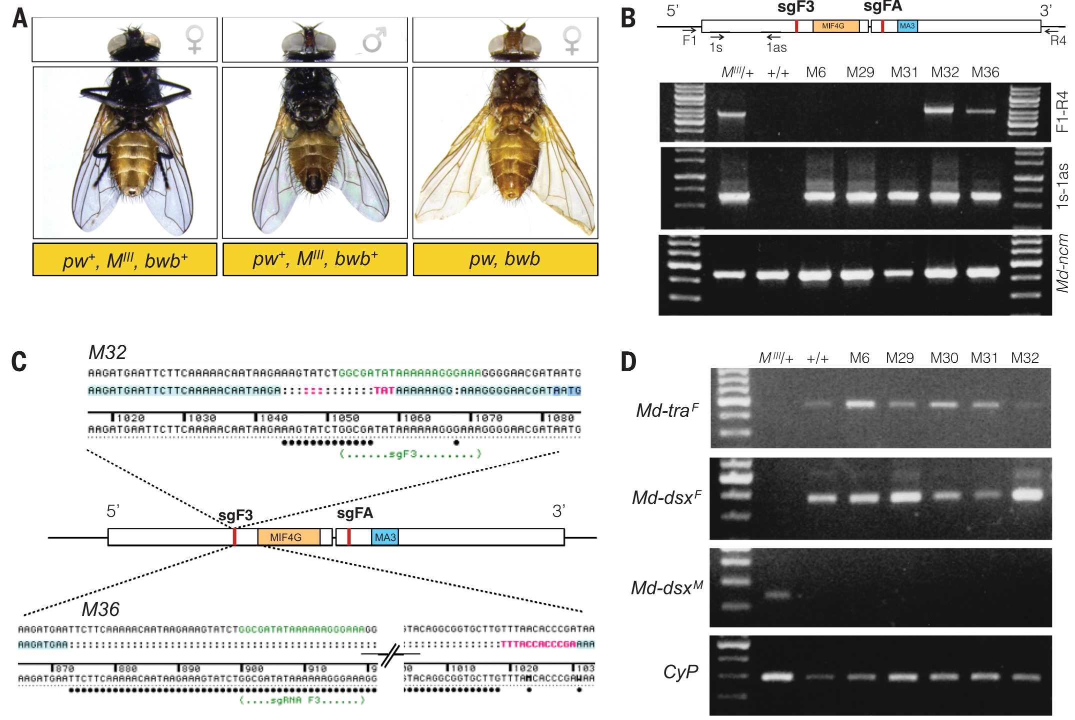


Fig. 3. CRISPR-Cas9–induced disruption of *Mdmd* causes complete male-to-female transformation. (**A**) F1 female of line M32 with *pw*$^+$, *bwb*$^+$ phenotype (left); male sibling with *pw*$^+$, *bwb*$^+$ phenotype (middle), and female sibling with *pw*, *bwb* phenotype (right). The phenotypic markers are described in the materials and methods and in fig. S7A. (**B**) CRISPR-Cas9–targeted sites sgF3 and sgFA in *Mdmd* (red stripes, top) and genomic amplifications of *Mdmd* and *Md-ncm* in F1 females of lines M6, M29, M31, M32, and M36 (bottom). In the upper blot, F1-R4 primers were used to amplify the ORF of *Mdmd* on chromosome III; in the middle blot, 1s-1as primers were used to amplify the 5′ region of different *Mdmd* copies; and in the lower blot, a primer pair was used that specifically amplifies *Md-ncm*. Absence of F1-R4 amplicons in M6, M29, and M31 indicates large deletions. (**C**) In M32 females, a deletion of 14 bp uncovers the sgF3 target site upstream of the MIF4G domain, causing a frame shift. In M36, a deletion of 146 bp removes the same target site and extends into the MIF4G domain. The target sequence is labeled in green in the wild-type reference sequence (pink letters, manually aligned for best fit; blue shading, sequence of mutant allele). (**D**) Expression of *Md-tra* and *Md-dsx* in sex-reverted females of lines M6, M29, M30, M31, and M32. Female splice variants are absent in control males ($M^{III}/+$) but present in control (+/+) and sex-reverted females. The male splice variant of *Md-dsx*, *Md-dsx*^{M}, is only detected in control males ($M^{III}/+$). Expression of cytochrome P450 (CyP) was used as an internal standard.

Downloaded from http://science.sciencemag.org/ on May 14, 2017

far not been found in other higher dipterans. Our study thus demonstrates that novel genes originating from duplication and neofunctionalization can adopt critical roles in essential developmental processes.

REFERENCES AND NOTES

1. J. B. Pease, M. W. Hahn, *Mol. Biol. Evol.* **29**, 1645–1653 (2012).
2. D. Bachtrog *et al*., *PLOS Biol.* **12**, e1001899 (2014).
3. B. Vicoso, D. Bachtrog, *PLOS Biol.* **13**, e1002078 (2015).
4. J. W. Erickson, J. J. Quintero, *PLOS Biol.* **5**, e332 (2007).
5. L. Sánchez, *Int. J. Dev. Biol.* **52**, 837–856 (2008).
6. E. C. Verhulst, L. van de Zande, L. W. Beukeboom, *Curr. Opin. Genet. Dev.* **20**, 376–383 (2010).
7. D. Bopp, G. Saccone, M. Beye, *Sex Dev.* **8**, 20–28 (2014).
8. M. G. Franco, P. G. Rubini, M. Vecchi, *Genet. Res.* **40**, 279–293 (1982).
9. A. Dübendorfer, M. Hediger, G. Burghardt, D. Bopp, *Int. J. Dev. Biol.* **46**, 75–79 (2002).
10. R. L. Hamm, R. P. Meisel, J. G. Scott, *G3* **5**, 371–384 (2015).
11. M. Hediger *et al*., *Genetics* **184**, 155–170 (2010).
12. T. Hiroyoshi, *Genetics* **50**, 373–385 (1964).
13. I. Denholm, M. G. Franco, P. G. Rubini, M. Vecchi, *Genet. Res.* **42**, 311–322 (1983).
14. M. Kozielska *et al*., *Genet. Res.* **90**, 157–165 (2008).
15. J. G. Scott *et al*., *Genome Biol.* **15**, 466 (2014).
16. A. L. Steckelberg, V. Boehm, A. M. Gromadzka, N. H. Gehring, *Cell Rep.* **2**, 454–461 (2012).
17. A. L. Steckelberg, J. Altmueller, C. Dieterich, N. H. Gehring, *Nucleic Acids Res.* **43**, 4687–4700 (2015).
18. J. Saulière *et al*., *Nat. Struct. Mol. Biol.* **17**, 1269–1271 (2010).
19. Z. Wang, V. Murigneux, H. Le Hir, *Genome Biol.* **15**, 551 (2014).
20. A. B. Hall *et al*., *Science* **348**, 1268–1270 (2015).
21. E. Krzywinska, N. J. Dennison, G. J. Lycett, J. Krzywinski, *Science* **353**, 67–69 (2016).

ACKNOWLEDGMENTS

We thank E. Geuverink and A. H. Rensink for technical advice, A. Meccariello and G. Saccone for providing purified Cas9 protein, and M. Hediger for helpful advice on housefly genetics. This work was supported by the Netherlands Organisation for Scientific Research (grant no. ALW 822.02.009) and an Ubbo Emmius grant to L.W.B.; the U4 Network sponsored by the DAAD (German Academic Exchange Service) "Strategic Partnerships" program and the German Federal Ministry of Education and Research to Georg-August-University Göttingen; and the Swiss National Science Foundation (grant no. 31003A_143883) to M.D.R. We thank M. Schenkel, S. Visser, G. Saccone, and the Groningen Evolutionary Genetics group members for fruitful discussion. The nucleotide sequences of *Mdmd* genes from different M strains have been deposited in the DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank database (accession numbers KY020047, KY020048, KY020049, and KY020050). Deep sequencing data are available from ArrayExpress (accession number E-MTAB-5080).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/356/6338/642/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 and S2
Protein Sequences
References (*22*–*39*)

8 December 2016; accepted 12 April 2017
10.1126/science.aam5498

Downloaded from http://science.sciencemag.org/ on May 14, 2017

Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene *CWC22*

Akash Sharma, Svenia D. Heinze, Yanli Wu, Tea Kohlbrenner, Ian Morilla, Claudia Brunner, Ernst A. Wimmer, Louis van de Zande, Mark D. Robinson, Leo W. Beukeboom and Daniel Bopp (May 11, 2017)

Science **356** (6338), 642-645. [doi: 10.1126/science.aam5498]

Editor's Summary

Disrupting housefly gene reverses sex

Sex comes in many forms, even when considered at the molecular level. In different animals, the chromosomes and specific genes that function in sex determination vary widely. As a case in point, the familiar housefly displays a highly variable sex determination system. In this animal, the male determiner (M-factor) instructs male development when it is active, but female development results when it is inactive. Sharma *et al.* now identify the housefly M-factor, which arose via the co-option of existing genes, gene duplication, and neofunctionalization. The findings elucidate the remarkable diversity in sex-determining pathways and the forces that drive this diversity.

Science, this issue p. 642

This copy is for your personal, non-commercial use only.

Article Tools Visit the online version of this article to access the personalization and article tools:
http://science.sciencemag.org/content/356/6338/642

Permissions Obtain information about reproducing this article:
http://www.sciencemag.org/about/permissions.dtl

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.

Downloaded from http://science.sciencemag.org/ on May 14, 2017