LETTER

https://doi.org/10.1038/s41586-018-0086-2

Male-killing toxin in a bacterial symbiont of *Drosophila*

Toshiyuki Harumoto[1]* & Bruno Lemaitre[1]*

Several lineages of symbiotic bacteria in insects selfishly manipulate host reproduction to spread in a population[1], often by distorting host sex ratios. *Spiroplasma poulsonii*[2,3] is a helical and motile, Gram-positive symbiotic bacterium that resides in a wide range of *Drosophila* species[4]. A notable feature of *S. poulsonii* is male killing, whereby the sons of infected female hosts are selectively killed during development[1,2]. Although male killing caused by *S. poulsonii* has been studied since the 1950s, its underlying mechanism is unknown. Here we identify an *S. poulsonii* protein, designated Spaid, whose expression induces male killing. Overexpression of Spaid in *D. melanogaster* kills males but not females, and induces massive apoptosis and neural defects, recapitulating the pathology observed in *S. poulsonii*-infected male embryos[5–11]. Our data suggest that Spaid targets the dosage compensation machinery on the male X chromosome to mediate its effects. Spaid contains ankyrin repeats and a deubiquitinase domain, which are required for its subcellular localization and activity. Moreover, we found a laboratory mutant strain of *S. poulsonii* with reduced male-killing ability and a large deletion in the *spaid* locus. Our study has uncovered a bacterial protein that affects host cellular machinery in a sex-specific way, which is likely to be the long-searched-for factor responsible for *S. poulsonii*-induced male killing.

Endosymbiotic bacteria have evolved sophisticated strategies to manipulate their hosts to increase their transmission, and sex ratio distorters of arthropods are examples of this manipulation. These bacteria are transmitted exclusively through female hosts, and as a result, several lineages have evolved the ability to bias infections towards females, either by turning males into females (feminization), causing clonal reproduction (parthenogenesis), or eliminating males (male killing)[1]. Male killing has independently evolved in at least six bacterial taxa, including *Spiroplasma*, *Wolbachia*, and *Rickettsia*[1]. The genetic basis of male killing is a longstanding mystery in the field of insect symbiosis. Male killing by *S. poulsonii* (Fig. 1a) was first described in the 1950s in *Drosophila*[2]. Previous studies of *S. poulsonii* attributed the selective killing of male progeny to an unknown substance called 'androcidin', assumed to be secreted by the bacterium[12]. The identification of this toxin has been hampered by the lack of practical methods for molecular biology, as has been the case for other symbiotic bacteria.

S. poulsonii symbionts of *D. melanogaster* (strain MSRO for '*melanogaster* sex ratio organism') kill all male progeny (for example, MSRO-H99; Extended Data Fig. 1a). We unexpectedly identified a *S. poulsonii* mutant strain that shows reduced male-killing ability (MSRO-SE; the partial male-killing strain), where almost half of the male progeny survived (Extended Data Fig. 1a–c). The reduced male killing was not linked to host genetic background or low bacterial titre (Extended Data Fig. 1b, d). To identify the genetic basis of reduced male killing, we sequenced the genome of MSRO-SE and compared it with that of MSRO-H99 (Extended Data Fig. 2). We found a candidate gene that was altered in the partial male-killing strain, encoding a 1,065-amino-acid protein with ankyrin repeats and the OTU (ovarian tumour) deubiquitinase domain. We named this *S. poulsonii* protein Spaid (*S. poulsonii* androcidin; Fig. 1b) based on our subsequent functional studies. We focused our analysis on this gene, because (i) both ankyrin repeats and the OTU domain are conserved in proteins across eukaryotes and are also present in some

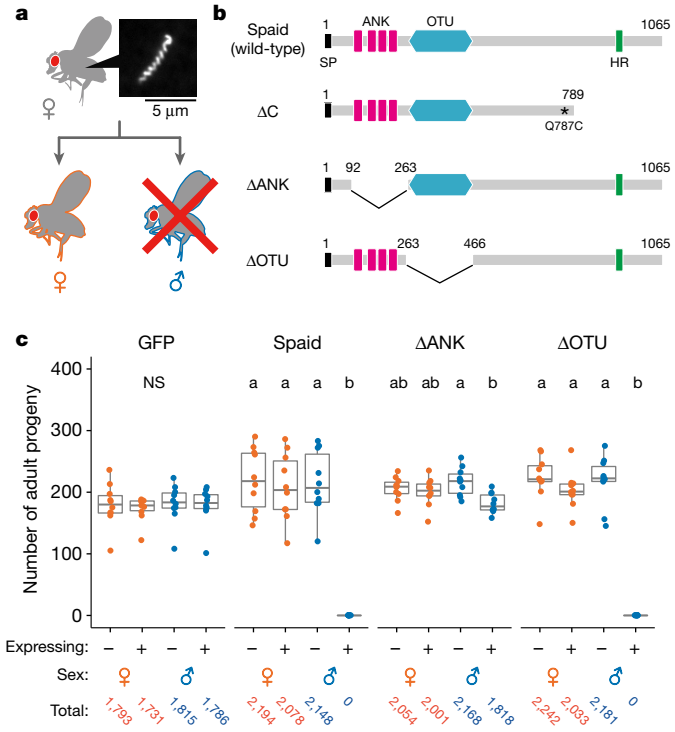


Fig. 1 | Expression of Spaid selectively eliminates male offspring. **a**, Male killing in *Drosophila* induced by *S. poulsonii*. Infected females (top) produce only female offspring (bottom). Inset shows a male-killing *S. poulsonii* of *D. melanogaster* detected by DNA staining. **b**, Protein structure of Spaid, which contains ankyrin repeats (ANK, red), the OTU deubiquitinase domain (blue), an N-terminal signal peptide (SP, black), and a C-terminal hydrophobic region (HR, green). Spaid ΔC of the partial male-killing strain encodes a protein with an amino acid substitution (Q787C) and a C-terminal truncation. The structures of the two deletion constructs of Spaid (ΔANK and ΔOTU) are also indicated. The numbers represent amino acid residues. **c**, The number of adult progeny obtained from crosses between the *actin–GAL4* line and four *UAS* transgenic lines (GFP, Spaid, ΔANK, and ΔOTU; $n = 10$ independent crosses for each transgene). The *UAS–GFP* line was used as a negative control. We counted the number of resultant offspring (females, red; males, blue) having both *actin–GAL4* and *UAS* (+) and siblings having only *UAS* as internal controls (−). Different lowercase letters indicate statistically significant differences (Steel–Dwass test (see Supplementary Table 2)). $P < 0.0001$ in all instances except ΔANK ($P < 0.05$). NS, not significant ($P > 0.05$). Box plots indicate the median (bold line), the 25th and 75th percentiles (box edges), and the range (whiskers). Dot plots show all data points individually. The total numbers of adult counts for each genotype and sex are shown at the bottom.

[1]Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. *e-mail: toshiyuki.harumoto@epfl.ch; bruno.lemaitre@epfl.ch

bacterial effectors that manipulate host cellular processes[13,14], and (ii) *spaid* was located on a putative plasmid[15] (Extended Data Fig. 2b), like other bacterial virulence factors. Further analysis predicted an N-terminal signal peptide for secretion and a C-terminal hydrophobic region (Fig. 1b). The *spaid* locus in the partial male-killing strain contained an 828-bp deletion (Extended Data Fig. 3), resulting in a truncated protein lacking the hydrophobic region, as well as a single amino acid substitution (Q787C) (ΔC; Fig. 1b). Of note, this gene was not present in an earlier published version of the *S. poulsonii* MSRO genome[16] (Supplementary Data), and we found no obvious homologous proteins in our BLAST searches.

To test whether Spaid is responsible for male killing, we used the GAL4/UAS system to overexpress the gene as a C-terminal green fluorescent protein (GFP) fusion protein in *D. melanogaster*. Notably, Spaid expression with the *actin–GAL4* ubiquitous driver eliminated all male offspring, but had no impact on female emergence (Fig. 1c). Thus, Spaid kills flies in a sex-dependent manner.

Male killing in *Drosophila* induced by *S. poulsonii* is associated with abnormal apoptosis[7,9] and neural disorganization[5,8,9] during embryogenesis. The mechanism of the neural defects is not known, but has been suggested to be independent of apoptosis[8,9]. If Spaid is the bona fide male-killing factor, its expression in embryos should phenocopy the pathology described above. We employed the *nanos-GAL4* maternal driver to express Spaid during early embryogenesis and monitored apoptosis (Fig. 2; with terminal deoxynucleotidyl transferase dUTP nick-end labelling (TUNEL) staining) and neural organization (Extended Data Fig. 4; using the neuron-specific marker Elav[17]). The sex of embryos was determined by antibody staining for the Sex lethal (Sxl) protein, which is expressed only in females[18]. We found that Spaid expression induced strong apoptosis and neural disorganization in male but not female embryos (Fig. 2 and Extended Data Fig. 4). The level of apoptosis in male embryos that expressed Spaid increased as their development proceeded (Fig. 2c), similar to embryos infected with *S. poulsonii*[9]. Although numerous cells underwent apoptosis in male embryos that expressed Spaid, the neural cells did not appear to suffer ectopic cell death (Extended Data Fig. 4c). Thus, the expression

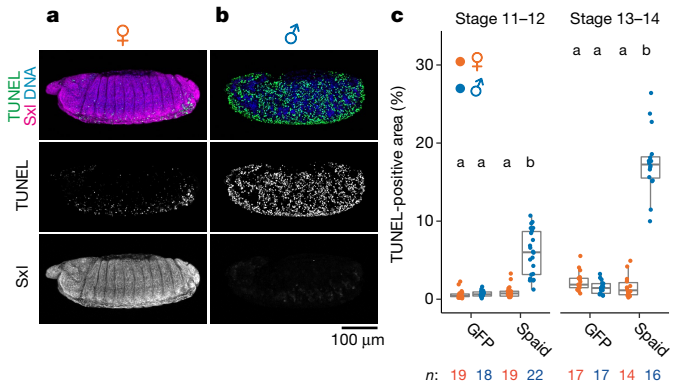


Fig. 2 | Expression of Spaid reproduces male-killing phenotypes during embryogenesis. a, **b**, Representative images of stage 13–14 female (**a**, $n = 14$) and male (**b**, $n = 16$) embryos maternally expressing Spaid, stained for apoptosis (TUNEL; green), Sxl (magenta), and DNA (blue). Single-channel images of TUNEL and Sxl are also shown. **c**, Quantification of TUNEL signals in stage 11–12 and 13–14 embryos (females, red; males, blue). Different characters indicate statistically significant differences ($P < 0.0001$; Steel–Dwass test; see Supplementary Table 2). Box and dot plots as in Fig. 1c. Sample sizes (*n*, number of embryos) are shown at the bottom. Embryos were co-stained for Elav, TUNEL, Sxl, and DNA, and selected channels are shown in **a**, **b** and Extended Data Fig. 4.

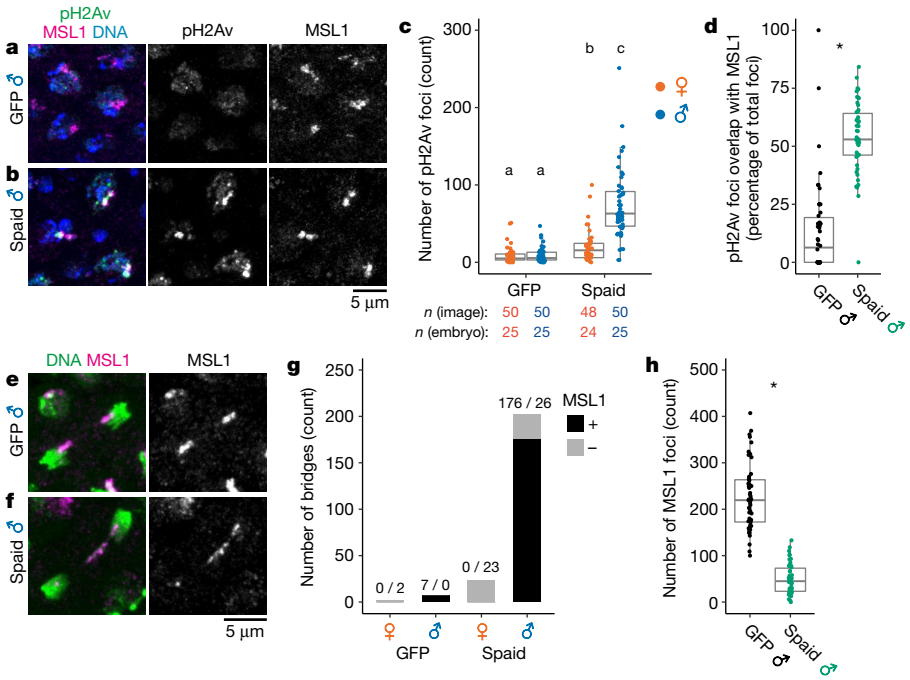


Fig. 3 | Spaid acts through the MSL complex. a, **b**, Epithelial cells in stage 9 male embryos expressing GFP (**a**) and Spaid (**b**), stained for pH2Av (green), MSL1 (magenta), and DNA (blue). **c**, Quantification of pH2Av foci in embryos expressing GFP and Spaid. **d**, Percentage of pH2Av foci overlapping with MSL1 signals in male embryos expressing; GFP, 6.3% (0–19.3%) (median (interquartile range)) and Spaid, 52.9% (46.2–64.2%). **e**, **f**, Dividing cells in stage 9 male embryos expressing GFP (**e**, proper segregation) and Spaid (**f**, a broken bridge) stained for DNA (green) and MSL1 (magenta). **g**, The number of chromatin bridges containing (black; numbers on the left) or not containing (grey; numbers on the right) MSL1 signals in embryos expressing GFP and Spaid. **h**, The number of MSL1 focal signals in male embryos expressing GFP (black) and Spaid (green). The same datasets of stage 8–10 embryos were analysed in **a**–**h** ($n = 50$ or 48 images per condition). Different lowercase letters or asterisks indicate significant differences: **c**, $P < 0.001$, Steel–Dwass test; **d**, $P < 0.0001$, χ^2 test; and **h**, $P < 0.0001$, two-tailed Mann–Whitney *U* test (see Supplementary Table 2). Box and dot plots in **c**, **d**, **h** are as in Fig. 1c and sample sizes are shown at the bottom in **c**. All *UAS* transgenes were expressed maternally.

of Spaid is sufficient to induce the two different pathologies caused by male-killing *S. poulsonii*.

In *Drosophila* (XX female, XY male), the single male X chromosome is hyper-transcribed by twofold to equalize gene expression levels between sexes. This dosage compensation system is mediated by a protein–RNA complex called the male-specific lethal (MSL) complex, which is selectively recruited to the male X chromosome[19]. Previous studies have uncovered a link between the male-killing action of *S. poulsonii* and the host's dosage compensation machinery[6,10,11]. Genetic experiments revealed that *S. poulsonii* fails to kill males lacking the MSL components, while it can induce death in females ectopically expressing the MSL complex[6,10,11]. This suggests that *S. poulsonii* either targets the MSL complex directly or targets its downstream chromatin modifications (for example, acetylation of histone H4 at K16)[19]. Remarkably, we found that the expression of Spaid triggered massive apoptosis in transgenic female embryos engineered to express the MSL complex (Extended Data Fig. 5a–d), indicating that this toxin mediates its effects through the dosage compensation machinery. We previously showed that *S. poulsonii* infection triggers DNA damage and segregation defects on the male X chromosome, thereby activating male-specific apoptosis[11]. To test whether Spaid expression can reproduce these phenotypes, we monitored DNA damage on the male X chromosome by antibody staining for the phosphorylated histone H2Av (pH2Av)[20] and MSL1, a component of the MSL complex labelling the male X chromosome. In control embryos expressing GFP and female embryos expressing Spaid, a few H2Av foci were detected, whereas male embryos expressing Spaid exhibited numerous bright foci (Fig. 3a–c) that frequently overlapped with MSL1 signals (Fig. 3d). We also found that Spaid expression in male embryos caused the formation of chromatin bridges, of which 87.1% contained MSL1 signals (Fig. 3e–g). The MSL1-labelled chromosomes were often fragmented and unevenly distributed after mitosis (Fig. 3f; 47.7% of the MSL1-containing bridges). In addition, MSL1 signals were reduced in male embryos that expressed Spaid, and we observed this phenotype at an even earlier time point (Fig. 3h and Extended Data Fig. 5e, f; stages 8–10) than has been described for embryos infected by *S. poulsonii* (from stage 13 onward)[10]. This is probably because the high expression of Spaid driven by the GAL4/UAS system results in much more severe effects than *S. poulsonii* infection. Finally, we examined the subcellular distribution of Spaid in larval salivary glands, whose large cellular sizes facilitate detailed cytological analyses. In both sexes, Spaid–GFP was found at plasma membranes, in the cytoplasm, and throughout nuclei, but was absent from the nucleolus (Fig. 4a, b). Notably, Spaid–GFP was enriched on MSL1-labelled chromosomes in male nuclei (Fig. 4b). This result is consistent with its reliance on the dosage compensation machinery, although equivalent localization analyses in embryos are needed.

To better characterize the action of Spaid, we made deletion constructs lacking either ankyrin repeats (ΔANK) or the OTU domain (ΔOTU) (Fig. 1b). Expression of ΔANK with *actin–GAL4* did not kill males, pointing to a pivotal role of ankyrin repeats in Spaid activity (Fig. 1c). The expression of ΔOTU with *actin–GAL4* still eliminated adult males, but they were killed later in development (pupal stage) compared with wild-type Spaid (second instar larval stage) (Fig. 1c). Furthermore, and in contrast to wild-type Spaid, weaker expression of ΔOTU with *armadillo–GAL4* failed to kill males (Extended Data Fig. 6), indicating that the OTU domain is required for full activity. Notably, although its distribution pattern was otherwise indistinguishable from that of Spaid–GFP in salivary gland cells, ΔANK–GFP was not enriched on MSL-labelled chromosomes (Fig. 4c, d). By contrast, ΔOTU–GFP was only weakly detected within nuclei regardless of sex, and colocalization with MSL was not apparent (Fig. 4e, f). The male-killing activity of ΔOTU when overexpressed may be caused by excess amounts of protein overriding its localization defect. All of these observations can be reconciled in a model of the chromosomal targeting of Spaid in which the OTU domain promotes nuclear localization and ankyrin repeats interact with the MSL complex or downstream histone modifications (Extended Data Fig. 7).

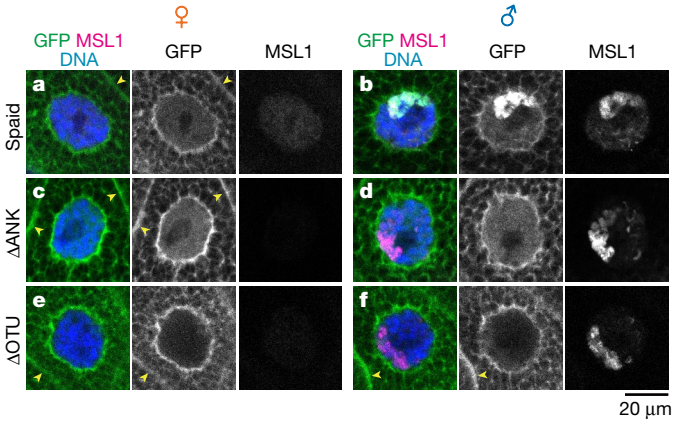


Fig. 4 | Subcellular localization of Spaid. a–f, Larval salivary glands expressing Spaid–GFP (**a**, female, $n = 13$; **b**, male, $n = 17$), ΔANK–GFP (**c**, female, $n = 9$; **d**, male, $n = 16$), and ΔOTU–GFP (**e**, female, $n = 15$; **f**, male, $n = 12$) stained for MSL1 (magenta) and DNA (blue). For GFP (green), raw fluorescence signals were detected. Magnified views of nuclei are shown. Dark spots inside nuclei in GFP images represent the nucleolus. Arrowheads indicate GFP signals associated with plasma membranes.

In summary, we have discovered an *S. poulsonii* gene that is likely to be responsible for male killing. Notably, expression of this single gene was sufficient to recapitulate the phenotypes associated with male killing. Our analysis revealed that Spaid is the sole ankyrin-repeat protein in the *S. poulsonii* genome (Supplementary Data). This contrasts with *Wolbachia*, whose genomes encode more than 20 ankyrin-repeat proteins[21,22]. Notably, the *Wolbachia* strain *w*Mel has a prophage-associated gene *WD0633*, that contains ankyrin repeats and the OTU domain[23]. Nevertheless, its transgenic expression in *D. melanogaster* produced no obvious phenotypes in a previous report[24]. Recent studies discovered *Wolbachia* genes that cause cytoplasmic incompatibility, a reproductive manipulation whereby symbiont-free females are unable to reproduce when mated with infected males[25–27]. Some of the genes that cause cytoplasmic incompatibility along with Spaid contain deubiquitinase domains, raising the possibility that host ubiquitination pathways are a common target in distinct strategies. Future research will focus on the identification of host cellular targets of Spaid, thereby deciphering the principles of its sex-specific activity. A thorough understanding of the reproductive manipulations induced by symbionts would not only provide novel insights into fundamental aspects of development, sex determination, and their evolution in insects, but could also provide clues to control insect populations.

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0086-2.

Received: 9 November 2017; Accepted: 21 March 2018;
Published online 2 May 2018.

1. Hurst, G. D. D. & Frost, C. L. Reproductive parasitism: maternally inherited symbionts in a biparental world. *Cold Spring Harb. Perspect. Biol.* **7**, a017699 (2015).
2. Williamson, D. L. & Poulson, D. F. in *The Mycoplasmas, Volume III: Plant and Insect Mycoplasmas* (eds Whitcomb R.F. & Tully J.G.) Ch. 6, 175–208 (Academic, New York, 1979).
3. Williamson, D. L. et al. *Spiroplasma poulsonii* sp. nov., a new species associated with male-lethality in *Drosophila willistoni*, a neotropical species of fruit fly. *Int. J. Syst. Bacteriol.* **49**, 611–618 (1999).
4. Haselkorn, T. S. The *Spiroplasma* heritable bacterial endosymbiont of *Drosophila*. *Fly (Austin)* **4**, 80–87 (2010).
5. Tsuchiyama-Omura, S., Sakaguchi, B., Koga, K. & Poulson, D. F. Morphological features of embryogenesis in *Drosophila melanogaster* infected with a male-killing. *Spiroplasma. Zool. Sci.* **5**, 375–383 (1988).

6. Veneti, Z., Bentley, J. K., Koana, T., Braig, H. R. & Hurst, G. D. D. A functional dosage compensation complex required for male killing in *Drosophila*. *Science* **307**, 1461–1463 (2005).
7. Bentley, J. K., Veneti, Z., Heraty, J. & Hurst, G. D. D. The pathology of embryo death caused by the male-killing *Spiroplasma* bacterium in *Drosophila nebulosa*. *BMC Biol.* **5**, 9 (2007).
8. Martin, J., Chong, T. & Ferree, P. M. Male killing *Spiroplasma* preferentially disrupts neural development in the *Drosophila melanogaster* embryo. *PLoS One* **8**, e79368 (2013).
9. Harumoto, T., Anbutsu, H. & Fukatsu, T. Male-killing *Spiroplasma* induces sex-specific cell death via host apoptotic pathway. *PLoS Pathog.* **10**, e1003956 (2014).
10. Cheng, B., Kuppanda, N., Aldrich, J. C., Akbari, O. S. & Ferree, P. M. Male-killing *Spiroplasma* alters behavior of the dosage compensation complex during *Drosophila melanogaster* embryogenesis. *Curr. Biol.* **26**, 1339–1345 (2016).
11. Harumoto, T., Anbutsu, H., Lemaitre, B. & Fukatsu, T. Male-killing symbiont damages host's dosage-compensated sex chromosome to induce embryonic apoptosis. *Nat. Commun.* **7**, 12781 (2016).
12. Oishi, K. Spirochaete-mediated abnormal sex-ratio (SR) condition in *Drosophila*: a second virus associated with spirochaetes and its use in the study of the SR condition. *Genet. Res.* **18**, 45–56 (1971).
13. Makarova, K. S., Aravind, L. & Koonin, E. V. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. *Trends Biochem. Sci.* **25**, 50–52 (2000).
14. Al-Khodor, S., Price, C. T., Kalia, A. & Abu Kwaik, Y. Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol.* **18**, 132–139 (2010).
15. Masson, F., Calderon Copete, S., Schüpfer, F., Garcia-Arraez, G. & Lemaitre, B. In vitro culture of the insect endosymbiont *Spiroplasma poulsonii* highlights bacterial genes involved in host-symbiont interaction. *MBio* **9**, e00024-18 (2018).
16. Paredes, J. C. et al. Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. *MBio* **6**, e02437–14 (2015).
17. Campos, A. R., Rosen, D. R., Robinow, S. N. & White, K. Molecular analysis of the locus *elav* in *Drosophila melanogaster*: a gene whose embryonic expression is neural specific. *EMBO J.* **6**, 425–431 (1987).
18. Salz, H. K. & Erickson, J. W. Sex determination in *Drosophila*: The view from the top. *Fly (Austin)* **4**, 60–70 (2010).
19. Lucchesi, J. C. & Kuroda, M. I. Dosage compensation in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **7**, a019398 (2015).
20. Madigan, J. P., Chotkowski, H. L. & Glaser, R. L. DNA double-strand break-induced phosphorylation of *Drosophila* histone variant H2Av helps prevent radiation-induced apoptosis. *Nucleic Acids Res.* **30**, 3698–3705 (2002).
21. Siozios, S. et al. The diversity and evolution of *Wolbachia* ankyrin repeat domain genes. *PLoS ONE* **8**, e55390 (2013).
22. Jernigan, K. K. & Bordenstein, S. R. Ankyrin domains across the Tree of Life. *PeerJ* **2**, e264 (2014).
23. Wu, M. et al. Phylogenomics of the reproductive parasite *Wolbachia pipientis* *w*Mel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**, e69 (2004).
24. Yamada, R., Iturbe-Ormaetxe, I., Brownlie, J. C. & O'Neill, S. L. Functional test of the influence of *Wolbachia* genes on cytoplasmic incompatibility expression in *Drosophila melanogaster*. *Insect Mol. Biol.* **20**, 75–85 (2011).
25. Beckmann, J. F. & Fallon, A. M. Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. *Insect Biochem. Mol. Biol.* **43**, 867–878 (2013).
26. Beckmann, J. F., Ronau, J. A. & Hochstrasser, M. A. *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat. Microbiol.* **2**, 17007 (2017).
27. LePage, D. P. et al. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature* **543**, 243–247 (2017).

Acknowledgements We thank the Bloomington Stock Center in the USA and the Department of *Drosophila* Genomics and Genetic Resources at Kyoto Institute of Technology in Japan (DGGR) for fly stocks; the Developmental Studies Hybridoma Bank at the University of Iowa for monoclonal antibodies; the *Drosophila* Genomics Resource Center in the USA for DNA vectors; J. Jaenike, T. Murata, and M. Kuroda for providing fly strains; T. Fukatsu for the *S. poulsonii*-infected fly stocks that were originally established in his laboratory by T.H; J. Lucchesi for providing antibodies; the BioImaging & Optics Platform (BIOP) in EPFL for confocal microscopy; K. Harshman and the Lausanne Genomic Technologies Facility (GTF) at the University of Lausanne (UNIL) for whole-genome sequencing and draft genome assemblies; and C. Vorburger and S. Perlman for comments and suggestions on the manuscript; and the members of the laboratory for discussion and support. This work was supported by the European Research Council (ERC) Advanced Grant 339970 and the Swiss National Science Foundation (SNSF) Sinergia Grant CRSII3_154396. The purchase of the PacBio RS II system in UNIL was supported in part by the Loterie Romande through the Fondation pour la Recherche en Médecine Génétique.

Reviewer information *Nature* thanks P. Ferree and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions T.H. conceived of and designed the study, performed the experiments, analysed the data and wrote and edited the manuscript. B.L. conceived and administered the study and wrote and edited the manuscript.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0086-2.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-018-0086-2.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to T.H. or B.L.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Fly stocks and genetics. Laboratory stocks of *D. melanogaster* were maintained at 25 °C with standard cornmeal medium. The fly stocks used in this study were cultured in tetracycline-containing medium (0.7–0.8 mg ml^{-1}) for one generation to eliminate possible contamination with *Wolbachia* and *Spiroplasma*. After treatment, about ten females and/or males were checked by diagnostic PCR with specific primers for *Wolbachia* (wsp_81F/691R)[28] and *Spiroplasma* (16SA1F/TKSSsp)[29,30] (Supplementary Table 1). Treated flies were maintained in the normal medium for at least two generations before use in experiments. The following lines were obtained from the Bloomington *Drosophila* Stock Center at Indiana University (BDSC) and the Department of *Drosophila* Genomics and Genetic Resources at Kyoto Institute of Technology (DGGR): *actin5C–GAL4* (*actin–GAL4*; BDSC 4414), *nanos–GAL4::VP16* (*nanos–GAL4*; BDSC 4937), *tubulin–GAL80^{ts}* (BDSC 7108), *armadillo–GAL4* (BDSC 1560), *UASp–eGFP* (DGGR 116071), and *CyO, ActGFP* (the green balancer; DGGR 107783). The *nanos–GAL4* flies, which were found to be *Wolbachia* and *Spiroplasma*-free by PCR, were not treated with antibiotics, because they became sick after treatment. Oregon-R (used as a wild-type line) and *msl1^{L60}/CyO*; *H83M2*[31] flies were generously provided by T. Murata and M. Kuroda, respectively.

To express the Spaid-encoding gene, we used the GAL4/UAS system[32] (see below). For the zygotic expression of Spaid, *actin–GAL4/CyO* (Fig. 1c) or *armadillo–GAL4* (homozygous; Extended Data Fig. 6) flies were crossed to homozygous *UAS* transgenic flies. To determine the lethal phase during larval stages, we analysed the number of teeth present in mouth hooks of killed larvae[33]. For maternal expression[34], *UAS* females were crossed to *nanos–GAL4* males, and the resultant female progeny were mated with Oregon-R males (Figs. 2, 3 and Extended Data Fig. 4 and 5e, f). For the ectopic formation of the MSL complex in females, *msl1^{L60}/CyO, ActGFP*; *H83M2* flies were used instead of Oregon-R. The *H83M2* transgene expresses the *msl-2* coding sequence (lacking Sxl-binding sites and resistant to translational repression) under the control of a heat shock-inducible promoter[31], but we used its leaky basal expression. The resultant embryos were distinguished by GFP staining and only GFP$^+$ embryos (wild-type for *msl1*) were analysed (Extended Data Fig. 5a–d). For the expression of Spaid in larval salivary glands (Fig. 4), we used a recombined *actin–GAL4, tubulin–GAL80^{ts}/CyO* line to avoid male lethality and let them grow until the third instar larval stage. Crosses maintained at 18 °C for 7–8 days were shifted to 29 °C and kept for 1 day before dissection. Only GFP$^+$ wandering third instar larvae were dissected.

Construction of transgenic fly lines. For the gene synthesis of the Spaid coding sequence by the GeneArt service (Thermo Fisher Scientific), TGA stop codons, which encode tryptophan in *Spiroplasma*, were mutated into TGG. Codon usage was optimized for the expression in *D. melanogaster*. Consequently, GC content was modified from 23.3% to 50.4%. We divided the Spaid coding sequence into two parts and synthesized them separately: one was a 2,367-bp fragment that corresponds to Spaid ΔC and the other was an 838-bp fragment encoding the remaining 3′ portion of full-length Spaid. To obtain the full-length Spaid coding sequence (3,195 bp, without a stop codon), the two fragments were fused using PCR with a 30-bp overlap and cloned into the pENTR vector using the pENTR/D-TOPO cloning kit (Thermo Fisher Scientific). To generate Spaid deletion constructs, we amplified two PCR fragments (nucleotide positions 1–276 and 787–3195 for ΔANK; nucleotide positions 1–789 and 1396–3195 for ΔOTU) from the synthetic Spaid coding sequence, fused them using PCR with a 24-bp overlap and cloned the product into the pENTR vector. We used PrimeSTAR HS DNA Polymerase (Takara Bio) for all PCR reactions. The Gateway cassette containing the Spaid coding fragments was transferred into the pPWG destination vector (The *Drosophila* Genomics Resource Center 1078; The *Drosophila* Gateway Vector Collection) by the LR clonase II enzyme mix kit (Thermo Fisher Scientific) to construct pUASp-Spaid-eGFP, pUASp-Spaid.ΔANK-eGFP, and pUASp-Spaid.ΔOTU-eGFP plasmids. Transgenic fly lines were generated by the standard microinjection method for P-element transformation (BestGene).

Identification and characterization of the partial male-killing *S. poulsonii* strain. Male-killing *S. poulsonii* can be easily transferred to other fly stocks by haemolymph injection[2]. *S. poulsonii* containing haemolymph was collected from a naturally infected *D. melanogaster* line Ug-SR provided by J. Jaenike[35], and transferred to wild-type Oregon-R flies that were used as the source of haemolymph for subsequent artificial infection. We established two artificially infected *D. melanogaster* stocks, *Df(3L)H99* (DGGR 106395) and *Sxl-eGFP* (BDSC 24105) in 2012 (Extended Data Fig. 1a). These infected stocks had shown perfect male killing (no emergence of adult males) for a period of at least 20 generations after the establishment of stable infection[9]. Afterwards, however, the infected *Sxl-eGFP* line started to produce male escapers, while the *Df(3L)H99* line showed complete male killing. We collected haemolymph from these fly stocks and injected them into Oregon-R flies to exclude the effect of host genetic background. We confirmed that only the *S. poulsonii* strain collected from the *Sxl-eGFP* line showed partial male killing (Extended Data Fig. 1b, c). The male-killing *S. poulsonii* strain in *D. melanogaster* is conventionally designated as MSRO (*melanogaster* sex ratio organism). Therefore, we call these *S. poulsonii* strains MSRO-H99 (the complete male killer as a control) and MSRO-SE (the partial male killer) after the genotypes of the host fly lines from which they were derived (*Df(3L)H99* and *Sxl-eGFP*, respectively). We also refer to the original male-killing strain from the Ug-SR line as MSRO-Ug according to the previous study[35].

We suspect that the partial male-killing strain might have resulted from genome rearrangements as reported in other *Spiroplasma* species maintained in the laboratory[36]. Repeat-rich sequences like viral fragments are often associated with genomic instability of *Spiroplasma*[37], and accordingly, the *spaid* gene of the partial male-killing strain seemed to be truncated by rearrangements between genes containing repetitive sequences (for example, *integrase* and *p58*; Extended Data Fig. 3b). To keep the partial male-killing strain, we used aged females (7–11 days) harbouring high bacterial titres to prevent the loss of the symbiont during vertical transmission. Because the titre of the partial male-killing strain appears to be higher than the original strain in some aged females, this may have led to the propagation of mutant bacteria and promoted their transmission (Extended Data Fig. 1d).

Analysis of *S. poulsonii* titres by quantitative PCR. Oregon-R virgin females infected with MSRO-Ug, MSRO-H99, and MSRO-SE were collected and aged for 0 (virgin), 7 or 14 days and individually transferred to a 1.5-ml tube and stored at −80 °C. They were then homogenized with plastic pestles by hand, and genomic DNA was purified using the DNeasy Blood & Tissue Kit (Qiagen). A portion of the *S. poulsonii dnaA* gene and the *D. melanogaster RpS17* gene were amplified with specific primer sets (dnaA_109F/246R[38] and rps17_615F/695R[39], respectively) (Supplementary Table 1) under the following PCR conditions: 95 °C for 5 min and then 45 cycles of 95 °C for 10 s, 55 °C for 20 s and 72 °C for 20 s. The 10-μl PCR mixture contained 2 μl of genomic DNA, 1× LightCycler 480 SYBR Green I master mix (Roche) and 0.5 μM primers. Quantitative PCR was performed by LightCycler 480 II (Roche). Cp values were obtained by the second derivative maximum method and the values of technical duplicates were averaged. We excluded samples if their duplicates had a difference in Cp values >0.5 cycles or if their Cp values were more than 30 cycles (2 out of 63 samples in total). To calculate the relative copy number of *S. poulsonii dnaA* gene against the host *RpS17* gene, we followed the Pfaffl method[40]. To estimate the PCR efficiency of each primer set, we performed quantitative PCR by using six tenfold serial dilutions of genomic DNA purified from ten female adult flies infected with MSRO-Ug, aged for 4–6 days after eclosion. PCR efficiency values were 82% (dnaA_109F/246R) and 98% (rps17_615F/695R), respectively.

Whole-genome sequencing of male-killing *S. poulsonii*. To collect a sufficient amount of DNA for whole-genome sequencing of *S. poulsonii* strains MSRO-H99 and MSRO-SE, we recovered fly haemolymph by a previously developed centrifugal separation method[41] with minor modifications. We first prepared a 0.5-ml polypropylene tube (Sarstedt, 72.704) whose bottom has a slit made by a sharp blade (hereafter called cartridge). Under CO_2 anaesthesia, we pricked the preepisternal area of the thorax of infected female adults and put them into the cartridge. The cartridge containing 30–40 flies was inserted into a 1.5-ml tube and centrifuged at 5,000 r.p.m. (2,300g) for 5 min at 4 °C. We usually made between three and ten cartridges at one time and they were kept on ice until centrifugation. After removing cartridges, we checked precipitates that contained host haemolymph and *S. poulsonii* under a stereomicroscope, and confirmed that there were no large debris (embryos and carcasses etc.). The tubes were washed with 400 μl PBS to resuspend and merge all precipitates. The merged suspension was filtered with a 0.65-μm pore size membrane filter (Ultrafree-MC; Merck, UFC30DV00) by centrifugation at 12,000g for 3 min at 4 °C to eliminate large microorganisms including yeasts. The flow-through was centrifuged at maximum speed for another 10 min and the supernatant was removed. Remaining bacterial pellets were stored at −80 °C. After suspending the bacterial pellet in 180 μl buffer ATL (Qiagen), 20 μl proteinase K solution (Qiagen) was added and pulse-vortexed for 10 s, and the mixture was then incubated for 1 h at 56 °C. Then, 4 μl RNase A (100 mg ml^{-1}; Qiagen) was added and treated for 5 min at room temperature. Genomic DNA was purified by the conventional phenol–chloroform extraction method (see PacBio Shared Protocol online, https://www.pacb.com/wp-content/uploads/2015/09/SharedProtocol-Extracting-DNA-usinig-Phenol-Chloroform.pdf). During the extraction, samples were mixed by pulse-vortexing within 20 s to prevent shearing of high molecular weight DNA. We collected haemolymph from 1,843 (MSRO-H99) and 1,378 (MSRO-SE) adult females to recover 18.9 μg and 7.56 μg genomic DNA, respectively (quantified by NanoDrop 1000; Thermo Fisher Scientific).

The genomic DNA was purified with Agencourt AMPure XP beads (Beckman Coulter) and was sheared in a Covaris g-TUBE (Covaris) to obtain 20-kb fragments. After shearing, the size distribution of DNA was checked by the Fragment Analyzer (Advanced Analytical Technologies). Sheared DNA (4 μg, MSRO-H99; 5 μg MSRO-SE) was used to prepare a SMRTbell library with the PacBio SMRTbell

Template Prep Kit 1.0 (Pacific Biosciences) according to the manufacturer's recommendations. The resulting library was size selected on the BluePippin system (Sage Science) for molecules larger than 16 kb (MSRO-H99) or 18 kb (MSRO-SE). The recovered library bound to MagBeads was sequenced on a single SMRT Cell with P6/C4 chemistry per genome by the PacBio RS II system (Pacific Biosciences) at 240-min movie length. Genome assembly was performed with the HGAP (Hierarchical Genome Assembly Process) software (Pacific Biosciences) version 3 and version 2 for the MSRO-H99 and MSRO-SE genomes, respectively. Library preparation, whole-genome sequencing and genome assembly were performed in the Lausanne Genomic Technologies Facility (GTF) at the University of Lausanne (UNIL).

Genomic data analyses and protein domain searches. Genomic data analysis was performed on the Bio-Linux 8 platform (NERC Environmental Omics)[42]. The genome sequences of MSRO-H99 and MSRO-SE assembled into five and two contigs, respectively (Extended Data Fig. 2a), and were aligned with the previously published MSRO genome[16] and ordered using Mauve v.2.4.0[43,44] (Extended Data Fig. 2b). For the MSRO-H99 genome, three major contigs (1–3) were assigned to the main chromosome, while the remaining two minor contigs (4 and 5) were assigned to extra chromosomes. Two contigs of the MSRO-SE genome were assigned to the main chromosome (contig 1; circularized) and the extra chromosome (contig 2). These extra chromosomes could be plasmids because they contain several proteins, such as Soj (the chromosome partitioning protein) and ARPs (adhesion-related proteins; P58, P12, P54, P123 and P18), which are located on plasmids in other *Spiroplasma* species[45].

Whole-genome annotation was conducted by using Prokka v.1.11[46]. We created a custom annotation database by combining published genomic sequences of several *Spiroplasma* species (*S. citri* CII3-3X (AM285301)[47], *S. melliferum* KC3 (AGBZ02000000)[48], *S. melliferum* IPMB4A (AMGI01000000)[49], *S. chrysopicola* DF-1 (CP005077)[37], *S. poulsonii* MSRO (NZ_JTLV00000000)[16], *S. kunkelii* CR2-3X (CP010899)[50]; the numbers in parentheses represent GenBank accession numbers).

To identify conserved protein domains including ankyrin repeats, annotated protein sequences were analysed using the NCBI Conserved Domain Database (NCBI CDD)[51,52] with default parameters (Supplementary Data). Besides *spaid* (SMH99_26490; 3,198 bp), other genes containing the OTU domain (SMH99_25890; 597 bp) and ankyrin repeats (SMH99_25900; 204 bp) were predicted on an identical contig (contig 4; 87,892 bp) in MSRO-H99. It is likely that they are derived from the misassembly or partial assembly of Spaid-encoding reads, because they are located close to the end of the contig (from 4,133 to 5,048 bp) with much lower coverage (below 10×) compared to the *spaid* locus (from 51,206 to 54,403 bp) with high coverage (over 330×). These genes were not found in MSRO-SE. Further domains were predicted using the InterPro protein sequence analysis and classification database[53]. The protein domain structure of Spaid was drawn using the Illustrator for Biological Sequences (IBS) software v 1.0.2[54] and modified manually.

Homology searches. For homology searches, we used the protein BLAST (blastp) program on the NCBI BLAST website with a non-redundant protein sequence database (nr). After searching the entire database without specifying the organism, we also searched bacteria, viruses, and *Drosophila* databases separately; however, we found no protein sequences that aligned to the entire Spaid protein sequence. The top hit having the highest score was a hypothetical protein of *Spiroplasma kunkelii* (WP_053391598), which covers the portion of the N- and C-terminal sequences of Spaid (amino acid positions 1–49 (score, 45.4; *E*-value, 0.71; identities, 48.98%; positives, 69.39%; gaps, 14.29%) and 704–1,065 (score, 385; *E*-value, 2.63×10^{-118}; identities, 58.03%; positives, 69.63%; gaps, 14.07%), respectively).

Sanger sequencing of the *spaid* locus. Ten adult Oregon-R females that were uninfected (as a negative control) or infected with MSRO-Ug, MSRO-H99, and MSRO-SE were collected and genomic DNA was purified using the method described above for quantitative PCR. The 3′ portion of the *spaid* gene was amplified with forward and reverse primer sets (*spaid*_1568F with *spaid*.L_ +136R for MSRO-Ug and MSRO-H99 and *spaid*.S_ +362R for MSRO-SE, respectively) (Supplementary Table 1) under the following PCR conditions: 95 °C for 2 min then 30 cycles of 95 °C for 30 s, 50 °C for 30 s and 60 °C for 3 min. Owing to the low GC content of the *spaid* gene (23.3%), extension at 60 °C instead of 72 °C was essential for successful PCR amplification[55]. The 50-μl PCR mixture contained 2 μl of genomic DNA, 1.25 U of GoTaq G2 DNA polymerase (Promega), 1× green reaction buffer (Promega), 0.5 μM primers, and 0.2 mM dNTP mixture. The PCR fragment was purified from agarose gel using the Wizard SV Gel and PCR Clean-Up System (Promega) and read by direct sequencing. To design primers, we referred to the genome sequences assembled from the PacBio data. No distinct PCR amplification was detected in uninfected Oregon-R female flies, confirming that the *spaid* gene is not encoded by the *D. melanogaster* genome but by the *S. poulsonii* genome.

Staining and imaging. To collect embryos, we put female and male flies in culture vials and waited for two days. These adults were transferred to egg-collection cages with grape-juice agar plates and fed with yeast paste. Developmental staging of embryos was followed refs. 56,57. Embryos were collected from grape-juice agar plates and dechorionated in 2.8% sodium hypochlorite solution, they were subsequently fixed in 1:1 mixture of heptane and 4% paraformaldehyde (EM Grade; Electron Microscopy Sciences, 15710) diluted in PBS for 20 min at room temperature, and devitellinized by vigorously shaking in heptane/MeOH. The embryos were washed in MeOH three times and rehydrated through an EtOH series (95%, 70%, 50% and 35%), and then washed in PBT (PBS containing 0.1% Triton X-100). After treated with a blocking buffer (PBT containing 2% bovine serum albumin (BSA, Fraction V; MP Biomedicals, 02160069)) for 60 min, the embryos were incubated with primary antibodies at 4 °C overnight, washed three times in PBT and incubated with secondary antibodies at room temperature for 90 min. Antibodies were diluted in the blocking buffer. Anti-Sxl and anti-MSL1 antibodies were used for sexing embryos (see below).

The following primary antibodies were used: mouse anti-Sxl (1:20; Developmental Studies Hybridoma Bank (DSHB), M18)[58], mouse and rabbit anti-MSL1 (1:200 and 1:500; kindly provided by J. Lucchesi), rabbit anti-histone H2AvD pS137 (pH2Av; 1:1,000; Rockland Immunochemicals, 600-401-914), chicken anti-GFP (1:500; Aves Labs, GFP-1020), rat anti-Elav (1:20; DSHB, 7E8A10)[59]. Secondary antibodies (1:1,000; Alexa Fluor 488/555/647 conjugate) were purchased from Molecular Probes (Thermo Fisher Scientific). DNA staining was carried out with DAPI (1 μg ml^{-1}; Sigma, D9542) for 10 min at room temperature after secondary antibody staining. TUNEL staining was performed by using the In situ Cell Death Detection Kit, TMR red (Roche), and the embryos stained with primary antibodies were incubated in 50 μl TUNEL reaction mixture with secondary antibodies at 4 °C overnight. Stained embryos were washed three times in PBT, mounted in FluorSave Reagent (Calbiochem) and observed under a confocal microscope (Zeiss LSM 700).

To detect *S. poulsonii* in host haemolymph (Fig. 1a), the abdomen of an adult female was dissected in 5 μl PBS with tweezers on a microscope slide. One microlitre of SYTO 9 nucleic acid stain solution (0.02 mM; Thermo Fisher Scientific) was added and observed under a widefield microscope equipped with a CCD camera (Zeiss Axio Imager Z1/AxioCam MRm).

To monitor the subcellular distribution of the GFP fusion proteins of Spaid, we dissected out salivary glands from wandering third instar larvae and fixed them in 4% paraformaldehyde diluted in PBS for 15 min at room temperature. Blocking and staining were performed as described above. For GFP, we detected raw fluorescent signals without antibody staining.

Image analysis and processing. Confocal *z*-sections were max projected using the Fiji software package[60] with a custom macro. Image analysis was performed using custom R scripts and the EBImage package[61]. For the quantification of TUNEL signals of whole embryos (Fig. 2c and Extended Data Fig. 5d) (acquired using a 20×/0.8 objective with 0.6× zoom scan; frame size: 512 × 512; 1.5 μm, two times optimal intervals), maximum projection images of DAPI and TUNEL staining were binarized and the former was used to make an embryonic mask image. TUNEL signals inside a corresponding mask image were measured by image integration. This value was divided by mask image area to normalize the embryonic size.

For the quantification of focal pH2Av signals (Fig. 3c), embryos from stages 8–10 were triply stained (pH2Av, MSL1, and DAPI) and two images were acquired per embryo (using a 63×/1.4 oil immersion objective; frame size, 1,024 × 1,024; 0.3 μm optimal intervals). We compiled 20 serial *z*-sections from the top (from 25 sections in total) to make projected images of pH2Av and MSL1 signals. These images were smoothed using a Gaussian filter and binarized using the moving average method, respectively. To identify focal pH2Av signals, image objects were extracted from the stacked images by segmentation and labelling (objects that were smaller than 20 pixels were eliminated to remove noise and ambiguous signals). To calculate the enrichment of focal pH2Av signals on the MSL1-labelled chromosome (Fig. 3d), overlaps between focal pH2Av signals and MSL1 signals were obtained by image integration. To quantify the number of MSL1 signals with discrete focal shapes (Fig. 3h), objects over 30 pixels were extracted and counted. The data sets described above were also used to count the number of chromatin bridges (Fig. 3g). The 25 serial *z*-slice images were visually inspected to detect chromatin bridges in the Fiji software.

The brightness and contrast of the presented images in the manuscript were adjusted using the level tool in the GIMP 2.8 software package. The adjustment was performed uniformly on the entire images and only black/white input levels were modified.

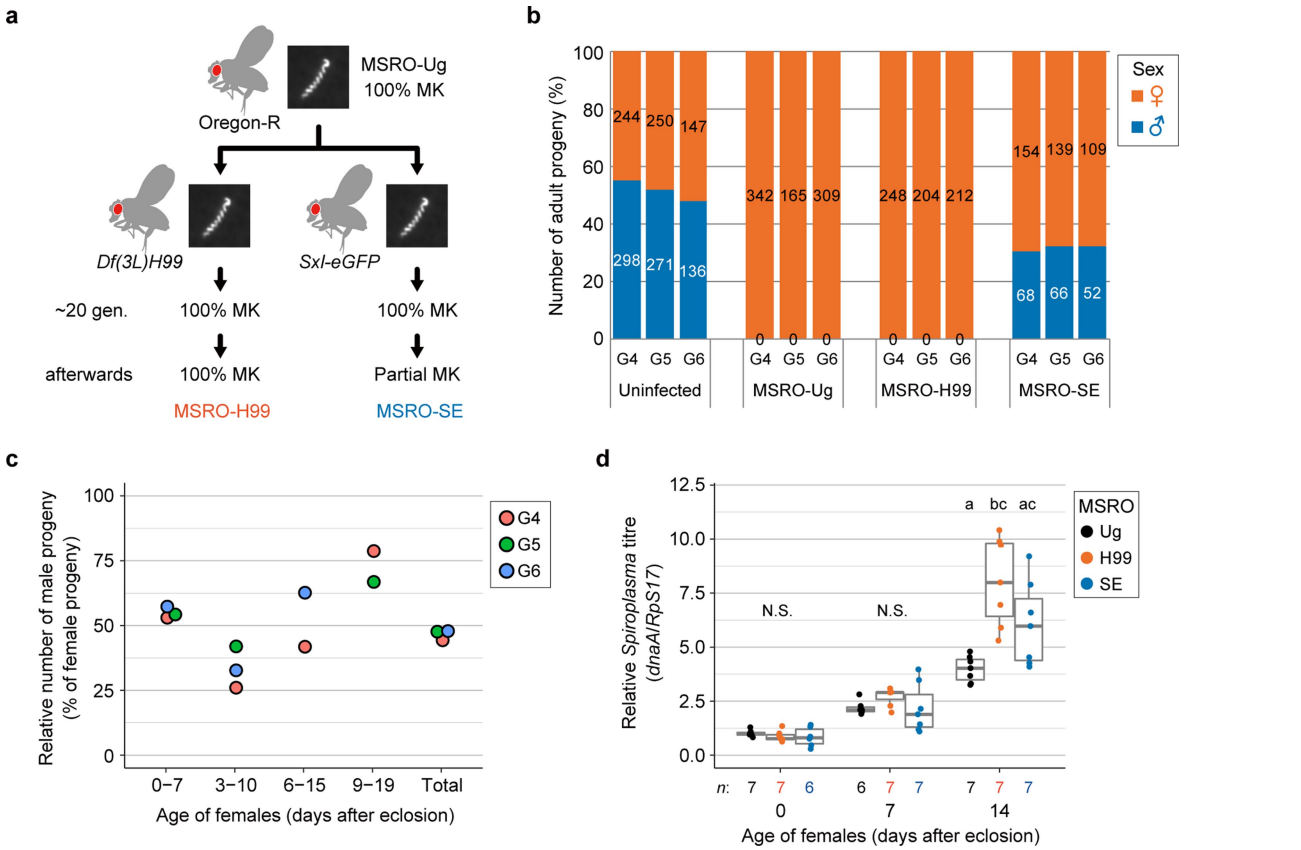
Statistics and reproducibility. No statistical methods were applied to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. We used the R v.3.4.1 software package (the R Foundation) for all statistical analyses. Multiple comparisons in Figs. 1c, 2c, 3c and Extended Data Fig. 1d, 5d were performed using the Steel–Dwass test of the pSDCFlig function in the NSM3 R package[62]. We performed the χ^2 test in Fig. 3d and the Mann–Whitney U test (two-tailed)

in Fig. 3h and Extended Data Fig. 6. *P*-values less than 0.05 were considered as significant. Exact *P*-values are listed in Supplementary Table 2. For the zygotic expression of *spaid* and its deletion constructs (Fig. 1c and Extended Data Fig. 6), two independently established *UAS* transgenic lines were used to verify the reproducibility of the results. Other experiments were independently repeated at least two times, except for the qPCR analysis in Extended Data Fig. 1d (one experiment with 6–7 biological replicates).

Reporting summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

Data availability. Whole genomic sequence data have been deposited at GenBank under the BioProject number PRJNA416288. Sequence data for synthetic gene fragments have been deposited at GenBank under the accession numbers MG837001 and MG837002. Source Data for Figs. 1, 2, 3 and Extended Data Fig. 1, 5, 6 are available in the online version of the paper. All relevant data supporting the findings of this study are included within the article and its Supplementary Information Files or are available from the corresponding authors upon reasonable request.

28. Zhou, W., Rousset, F. & O'Neil, S. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. *Proc. Biol. Sci.* **265**, 509–515 (1998).
29. Fukatsu, T. & Nikoh, N. Two intracellular symbiotic bacteria from the mulberry psyllid *Anomoneura mori* (Insecta, Homoptera). *Appl. Environ. Microbiol.* **64**, 3599–3606 (1998).
30. Fukatsu, T. & Nikoh, N. Endosymbiotic microbiota of the bamboo pseudococcid *Antonina crawii* (Insecta, Homoptera). *Appl. Environ. Microbiol.* **66**, 643–650 (2000).
31. Kelley, R. L. et al. Expression of msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. *Cell* **81**, 867–877 (1995).
32. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
33. Ashburner, M., Golic, K. G. & Hawley, R. S. *Drosophila: a Laboratory Handbook. 2nd Edition*, (Cold Spring Harbor, New York, 2005).
34. Rørth, P. Gal4 in the *Drosophila* female germline. *Mech. Dev.* **78**, 113–118 (1998).
35. Pool, J. E., Wong, A. & Aquadro, C. F. Finding of male-killing *Spiroplasma* infecting *Drosophila melanogaster* in Africa implies transatlantic migration of this endosymbiont. *Heredity (Edinb)* **97**, 27–32 (2006).
36. Ye, F., Melcher, U., Rascoe, J. E. & Fletcher, J. Extensive chromosome aberrations in *Spiroplasma citri* Strain BR3. *Biochem. Genet.* **34**, 269–286 (1996).
37. Ku, C., Lo, W.-S., Chen, L.-L. & Kuo, C.-H. Complete genomes of two dipteran-associated spiroplasmas provided insights into the origin, dynamics, and impacts of viral invasion in *Spiroplasma*. *Genome Biol. Evol.* **5**, 1151–1164 (2013).
38. Anbutsu, H. & Fukatsu, T. Population dynamics of male-killing and non-male-killing spiroplasmas in *Drosophila melanogaster*. *Appl. Environ. Microbiol.* **69**, 1428–1434 (2003).
39. Osborne, S. E., Leong, Y. S., O'Neill, S. L. & Johnson, K. N. Variation in antiviral protection mediated by different *Wolbachia* strains in *Drosophila simulans*. *PLoS Pathog.* **5**, e1000656 (2009).
40. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res.* **29**, e45 (2001).
41. Musselman, L. P. *Drosophila* hemolymph collection procedure. *YouTube* https://www.youtube.com/watch?v=im78OIBKIPA (2013).
42. Field, D. et al. Open software for biologists: from famine to feast. *Nat. Biotechnol.* **24**, 801–803 (2006).
43. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403 (2004).
44. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147 (2010).
45. Saillard, C. et al. The abundant extrachromosomal DNA content of the *Spiroplasma citri* GII3-3X genome. *BMC Genomics* **9**, 195 (2008).
46. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
47. Carle, P. et al. Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral invasion and important gene decay. *Appl. Environ. Microbiol.* **76**, 3420–3426 (2010).
48. Alexeev, D. et al. Application of *Spiroplasma melliferum* proteogenomic profiling for the discovery of virulence factors and pathogenicity mechanisms in host-associated spiroplasmas. *J. Proteome Res.* **11**, 224–236 (2012).
49. Lo, W.-S., Chen, L.-L., Chung, W.-C., Gasparich, G. E. & Kuo, C.-H. Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. *BMC Genomics* **14**, 22 (2013).
50. Davis, R. E. et al. Complete genome sequence of *Spiroplasma kunkelii* strain CR2-3X, causal agent of corn stunt disease in *Zea mays* L. *Genome Announc.* **3**, e01216–15 (2015).
51. Marchler-Bauer, A. et al. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–D229 (2011).
52. Marchler-Bauer, A. et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).
53. Finn, R. D. et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res* **45**, D190–D199 (2017).
54. Liu, W. et al. IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics* **31**, 3359–3361 (2015).
55. Su, X. Z., Wu, Y., Sifri, C. D. & Wellems, T. E. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Res.* **24**, 1574–1575 (1996).
56. Hartenstein, V. *Atlas of Drosophila Development*. (Cold Spring Harbor, New York, 1993).
57. Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster*. (Springer, Berlin, 1997).
58. Bopp, D., Bell, L. R., Cline, T. W. & Schedl, P. Developmental distribution of female-specific Sex-lethal proteins in *Drosophila melanogaster*. *Genes Dev.* **5**, 403–415 (1991).
59. O'Neill, E. M., Rebay, I., Tjian, R. & Rubin, G. M. The activities of two Ets-related transcription factors required for *Drosophila* eye development are modulated by the Ras/MAPK pathway. *Cell* **78**, 137–147 (1994).
60. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
61. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26**, 979–981 (2010).
62. Hollander, M., Wolfe, A. D. & Chicken, E. *Nonparametric Statistical Methods, 3rd Edition*, (John Wiley & Sons, New Jersey, 2013).

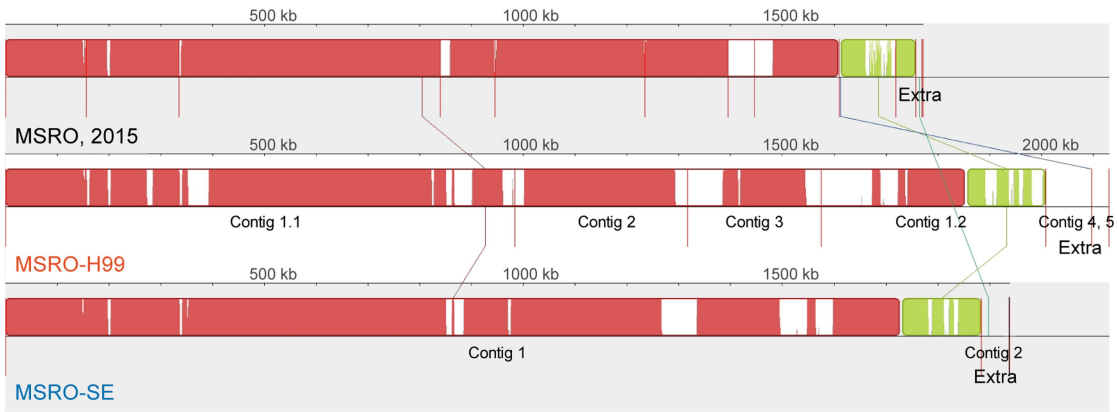


Extended Data Fig. 1 | Identification and characterization of the partial male-killing *S. poulsonii* strain. a, An illustration showing the origin of the *S. poulsonii* strains analysed in this study. Pictures show male-killing (MK) *S. poulsonii* of *D. melanogaster*. MSRO-Ug is the original male-killing strain maintained in the Oregon-R wild-type fly. Fly stocks (*Df(3L)H99* and *Sxl-eGFP*) artificially infected with this original strain showed complete male killing (100% MK) for the first 20 generations. Afterwards, one strain (MSRO-SE) started to show the partial male-killing phenotype, whereas the other (MSRO-H99) kept the ability to induce complete male killing. See Methods section 'Identification and characterization of the partial male-killing *S. poulsonii* strain' for more detail. **b**, **c**, Sex-ratio analysis of the adult progeny obtained from Oregon-R flies infected with MSRO-Ug, MSRO-H99, and MSRO-SE. We repeated experiments three times on the fourth, fifth and sixth generations (G4–6) after the establishment of infection. In **c**, the relative number of male offspring (percentage of females) obtained from Oregon-R female flies infected with MSRO-SE are plotted. Data points were excluded if the total count of flies was below 10. **d**, Relative titre of *S. poulsonii* within individual female flies. Adult females infected with three MSRO strains were kept for 0, 7 and 14 days after eclosion and analysed by qPCR. Data were normalized with respect to females from day 0 that were infected with MSRO-Ug. Different letters indicate statistically significant different groups ($P < 0.01$; N.S., not significant, $P > 0.05$; Steel–Dwass test; see Supplementary Table 2). Please note that the titres of the three strains are comparable and even the higher titre in old females (see 14 days in **d**) fails to induce complete male killing in MSRO-SE (**c**). Box and dot plots are as in Fig. 1c and sample sizes (*n*, number of adult flies) are shown at the bottom.

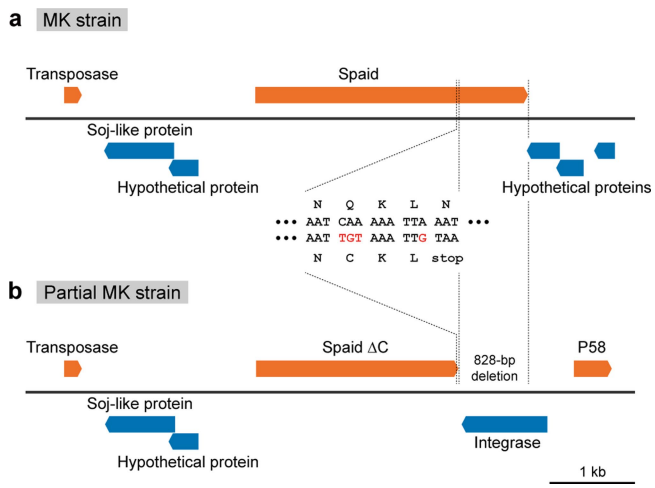
a

	MSRO, 2015	MSRO-H99	MSRO-SE
Reference	Paredes *et al.*, 2015	This work	This work
Male killing	Complete	Complete	Partial
Contigs	12	5	2
Largest contig (bp)	504,367	1,417,292	1,883,572
Contig 1	-	1,417,292	1,883,572
Contig 2	-	333,653	55,268
Contig 3	-	257,938	-
Contig 4	-	87,892	-
Contig 5	-	34,188	-
Total length (bp)	1,771,859	2,130,963	1,938,840
GC (%)	27	26	26
N50	179,219	1,417,292	1,883,572
N75	155,942	333,653	1,883,572
L50	3	1	1
L75	6	2	1
Total genes	1,976	2,749	2,516
CDS	1,942	2,715	2,482
tRNA	31	31	31
rRNA	3	3	3

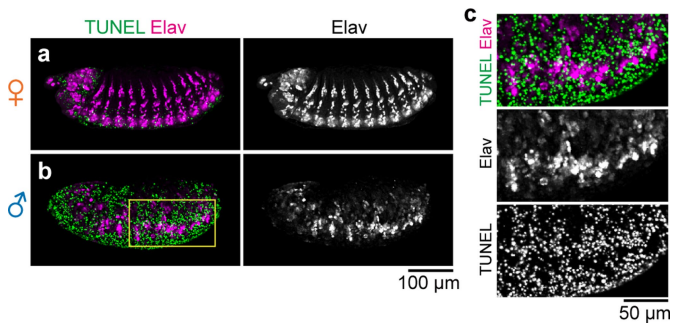
b



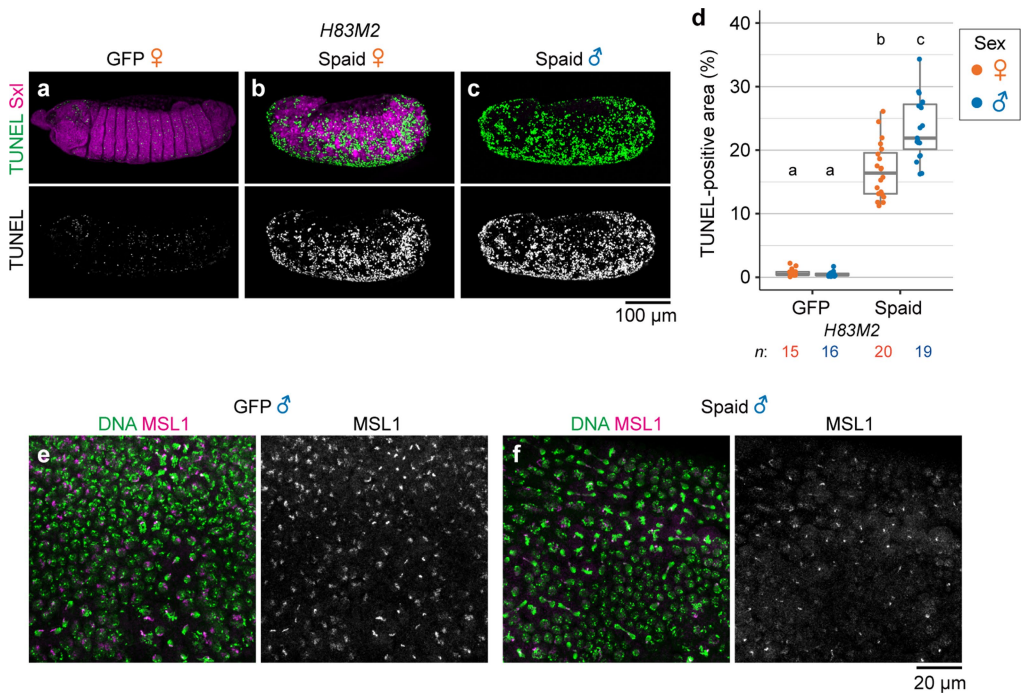
Extended Data Fig. 2 | Whole-genome sequencing studies of the male-killing *S. poulsonii* variants. a, Comparison of the genomic features of three *S. poulsonii* strains. MSRO-H99 and MSRO-SE are newly obtained variants isolated in this study (Extended Data Fig. 1a). As a control, data from the previously reported male-killing *S. poulsonii* genome[16] are also indicated. **b**, Whole-genome alignment of the three *S. poulsonii* strains. To start the alignment from the *dnaA* gene, contig 1 of MSRO-H99 was split into two fragments (contig 1.1 and 1.2). The locations of the contigs corresponding to extra chromosomes (putative plasmids; see Methods section 'Genomic data analyses and protein domain searches') are shown as 'extra'. *spaid* (gene ID from GenBank BioProject PRJNA416288: SMH99_26490) and *spaid* ΔC (gene ID: SMSE_25110) are located on these extra chromosomes in MSRO-H99 (contig 4) and MSRO-SE (contig 2), respectively.



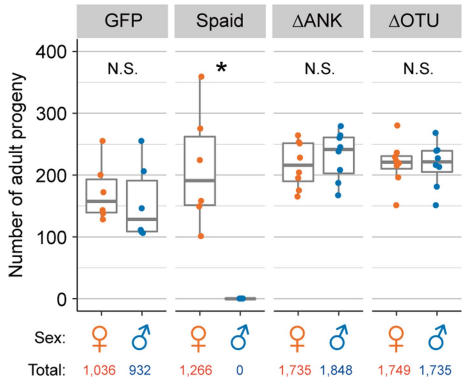
Extended Data Fig. 3 | Genetic alterations of the *spaid* locus in the partial male-killing *S. poulsonii* strain. The genome structures around the *spaid* loci in the male-killing (**a**, MSRO-Ug and MSRO-H99) and the partial male-killing (**b**, MSRO-SE) *S. poulsonii* strains. Genes encoded on opposite strands are shown in different colours (red and blue, respectively). An 828-bp deletion and nucleotide substitutions (coloured in red; corresponding amino acid sequences are presented in one-letter code) in the 3′ region of the *spaid* gene are indicated. These sequence alterations were confirmed using Sanger sequencing (see Methods).



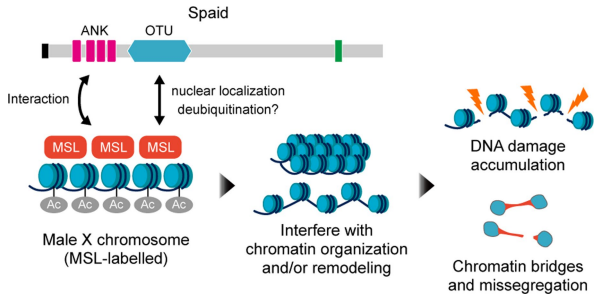
Extended Data Fig. 4 | Neural defects of Spaid-expressing embryos. Representative images of stage 13–14 female (**a**, $n = 14$) and male (**b**, $n = 16$) embryos maternally expressing Spaid and stained for TUNEL (green) and neural cells (Elav, magenta). Single-channel images of Elav are also shown. The outlined region in **b** is magnified in **c** with single-channel images of Elav and TUNEL. Embryos were co-stained for Elav, TUNEL, Sxl and DNA, and selected channels are shown in **a–c** and Fig. 2a, b, respectively.



Extended Data Fig. 5 | Spaid acts through the dosage compensation machinery. a–c, Representative images of stage 13–14 embryos ectopically expressing the MSL complex (*H83M2* transgene), stained for TUNEL (green) and Sxl (magenta). GFP-expressing control female (**a**, $n = 15$), Spaid-expressing female (**b**, $n = 20$), and male (**c**, $n = 19$) embryos are shown. **d**, Quantification of TUNEL signals in *H83M2* embryos at stages 13–14. Different characters indicate significantly different groups ($P < 0.001$; Steel–Dwass test; see Supplementary Table 2). The box and dot plot (females, red; males, blue) is as in Fig. 1c and sample sizes (n, number of embryos) are shown at the bottom. **e**, **f**, Representative images of epithelial cells of stage 8–10 male embryos expressing GFP (**e**, $n = 25$) and Spaid (**f**, $n = 25$), stained for DNA (green) and MSL1 (magenta) from the datasets analysed in Fig. 3. All *UAS* transgenes were expressed maternally.



Extended Data Fig. 6 | Expression of Spaid using a weak *GAL4* driver. The number of adult progeny (females, red; males, blue) obtained from crosses between the *armadillo-GAL4* driver line (weak and ubiquitous expression) and four *UAS* transgenic lines (GFP, Spaid, ΔANK, and ΔOTU; $n = 6$ independent crosses for GFP and Spaid, $n = 8$ independent crosses for ΔANK and ΔOTU). The *UAS-GFP* line was used as a negative control. With this weak *GAL4* driver, Spaid still eliminated all male progeny, while both ΔANK and ΔOTU had no impact on male viability. An asterisk indicates the statistically significant difference ($P < 0.01$; N.S., not significant, $P > 0.05$; two-tailed Mann–Whitney U test; see Supplementary Table 2). Box and dot plots are as in Fig. 1c. The total numbers of adult counts for each genotype and sex are shown at the bottom.



Extended Data Fig. 7 | A proposed model for Spaid-induced male-killing phenotypes. Spaid utilizes the OTU domain and ankyrin repeats (ANK) to target the host nucleus and the male X chromosome. 'MSL' and 'Ac' indicate the dosage compensation complex and resultant histone acetylation, respectively. See text for other explanations.

Corresponding author(s): Toshiyuki Harumoto
Bruno Lemaitre

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

   Describe how sample size was determined.

   No statistical methods were applied to determine sample sizes; however, they were determined empirically on the basis of observed effects.

2. Data exclusions

   Describe any data exclusions.

   Extended Data Fig. 1c: we excluded data points if the total count of flies was below 10 (2 out of 12 data points in total).
   Extended Data Fig. 1d: we excluded samples if their duplicates had a difference in Cp values > 0.5 cycles or if their Cp values were more than 30 cycles (2 out of 63 samples in total).

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   For the zygotic expression of SpAID and its deletion constructs (Fig. 1c and Extended Data Fig. 6), two independently established UAS transgenic lines were used and obtained similar results. In other experiments, at least two replicates were successful. For the qPCR analysis in Extended Data Fig. 1d, only one experiment with 6-7 biological replicates was performed, because we obtained consistent results at three independent time points.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   This is not relevant to our study, because we compared samples with different genetic backgrounds.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   This is not relevant to our study, because samples were not allocated to groups.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

R v3.4.1 for statistical analysis and data plot
The EBImage package for R and Fiji (ImageJ 1.51r) for imaging analysis
HGAP v2 and v3 for genome assembly
Bio-Linux 8 for genomic data analysis
Mauve v2.4.0 for genome alignment
Prokka v1.11 for genome annotation

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

We used antibodies already validated by the research community. Below is the list:
mouse anti-Sex lethal [1:20; Developmental Studies Hybridoma Bank (DSHB), M18]
mouse anti-MSL1 [1:200; kindly provided by John Lucchesi (Emory University)]
rabbit anti-MSL1 [1:500; kindly provided by John Lucchesi (Emory University)]
rabbit anti-Histone H2AvD pS137 (1:1,000; Rockland, 600-401-914)
chicken anti-GFP (1:500; Aves Labs, GFP-1020)
rat anti-Elav (1:20; DSHB, 7E8A10)
Alexa Fluor 488 goat anti-rabbit (1:1,000; Molecular Probes, A11034)
Alexa Fluor 488 goat anti-rat (1:1,000; Molecular Probes, A11006)
Alexa Fluor 488 goat anti-chicken (1:1,000; Molecular Probes, A11039)
Alexa Fluor 555 goat anti-rabbit (1:1,000; Molecular Probes, A21429)
Alexa Fluor 555 goat anti-mouse (1:1,000; Molecular Probes, A21424)
Alexa Fluor 647 goat anti-rabbit (1:1,000; Molecular Probes, A21245)
Alexa Fluor 647 goat anti-mouse (1:1,000; Molecular Probes, A21236)

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used. — No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used. — No eukaryotic cell lines were used.

c. Report whether the cell lines were tested for mycoplasma contamination. — No eukaryotic cell lines were used.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. — No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study. — We used Drosophila melanogaster laboratory stocks. Adults, larvae, and stages 8-10, 11-12, and 13-14 embryos (both females and males) were analyzed.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants. — The study did not involve human research participants.