Published in final edited form as:
*Nature*. 2018 May ; 557(7704): 252–255. doi:10.1038/s41586-018-0086-2.

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

**Toshiyuki Harumoto**[*] and **Bruno Lemaitre**[*]
Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Station 19 CH-1015 Lausanne, Switzerland

## Abstract

Several lineages of symbiotic bacteria in insects selfishly manipulate host reproduction to spread in a population1, often by distorting host sex ratios. *Spiroplasma poulsonii*2,3 (hereafter *Spiroplasma*) is a helical and motile, Gram-positive symbiotic bacterium, that resides in a wide range of *Drosophila* species4. A striking feature of *Spiroplasma* is male killing, whereby the sons of infected female hosts are selectively killed during development1,2. Although male killing caused by *Spiroplasma* has been studied since the 1950s, the underlying mechanism is unknown. Here, we identify a *Spiroplasma* protein, designated *Sp*AID, whose expression induces male killing. Overexpression of *Sp*AID in *D. melanogaster* kills males but not females, and induces massive apoptosis and neural defects, recapitulating the pathology observed in *Spiroplasma*-infected male embryos5–11. Our data suggest that *Sp*AID targets the dosage compensation machinery on the male X chromosome to mediate its effects. *Sp*AID contains ankyrin repeats and a deubiquitinase domain, which are required for its subcellular localization and activity. Moreover, we found a laboratory mutant strain of *Spiroplasma* with reduced male-killing ability and a large deletion in the *SpAID* locus. Collectively, our study has uncovered a novel bacterial protein that affects host cellular machinery in a sex-specific way, which is likely to be the long-searched-for factor responsible for *Spiroplasma*-induced male killing.

Endosymbiotic bacteria have evolved sophisticated strategies to manipulate their hosts to increase their transmission, and sex ratio distorters of arthropods are perhaps the champions of 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



[*]Correspondence and requests for materials should be addressed to Toshiyuki Harumoto (toshiyuki.harumoto@epfl.ch) or Bruno Lemaitre (bruno.lemaitre@epfl.ch).

Toshiyuki Harumoto: ORCID 0000-0003-4394-5883
Bruno Lemaitre: ORCID 0000-0001-7970-1667

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

**Competing financial or non-financial interests**
The authors declare no competing financial or non-financial interests.


Europe PMC Funders Author Manuscripts

killing is one of the biggest longstanding mysteries in the field of insect symbiosis. Male killing by *Spiroplasma* (Fig. 1a) was described as early as the 1950s in *Drosophila*2. Past studies in *Spiroplasma* attributed the selective killing of male progeny to an unknown substance called “androcidin”, assumed to be secreted by the bacterium12. The identification of this toxin has been hampered by the lack of practical methods for molecular biology as with the case of other symbiotic bacteria.

*Spiroplasma* symbionts of *D. melanogaster* (strain MSRO for “*m*elanogaster sex ratio organism”) kill all male progeny (e.g., MSRO-H99; Extended Data Fig. 1a). We unexpectedly identified a *Spiroplasma* 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 due 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 an intriguing 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 *Spiroplasma* protein *Sp*AID (*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 bacterial effectors that manipulate host cellular processes13,14 and (*ii*) *SpAID* was located on a putative plasmid (Extended Data Fig. 2b) like other bacterial virulence factors15. 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 *Spiroplasma* MSRO genome16 (Supplementary Data), and we found no obvious homologous proteins in our BLAST searches.

To test whether *Sp*AID is responsible for male killing, we utilized the GAL4/UAS system to overexpress the gene as a C-terminal GFP fusion protein in *D. melanogaster*. Strikingly, *Sp*AID expression with the *Actin-GAL4* ubiquitous driver eliminated all male offspring, while it had no impact on female emergence (Fig. 1c). Thus *Sp*AID kills flies in a sex-dependent manner.

*Spiroplasma*-induced male killing in *Drosophila* is associated with abnormal apoptosis7,9 and neural disorganization5,8,9 during embryogenesis; the mechanism of neural defects is not known, but is suggested to be independent of apoptosis8,9. If *Sp*AID was the *bona fide* male-killing factor, its expression in embryos should phenocopy the above pathology. We employed the *nanos-GAL4* maternal driver to express *Sp*AID during early embryogenesis and monitored apoptosis (Fig. 2; with TUNEL staining) and neural organization (Extended Data Fig. 4; with the neuron-specific marker Elav17). The sex of embryos was determined by antibody staining for the Sex lethal (Sxl), a protein expressed only in females18. We found that *Sp*AID expression induces strong apoptosis and neural disorganization in male but not female embryos (Fig. 2 and Extended Data Fig. 4). The level of apoptosis in *Sp*AID-expressing male embryos increased as their development proceeded (Fig. 2c), similar to

*Spiroplasma*-infected embryos9. Although numerous cells underwent apoptosis in *Sp*AID-expressing male embryos, the neural cells did not appear to suffer ectopic cell death (Extended Data Fig. 4c). Thus, the expression of *Sp*AID is sufficient to induce the two different pathologies caused by male-killing *Spiroplasma*.

In *Drosophila* (XX female, XY male), the single male X chromosome is hyper-transcribed by two-fold 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 chromosome19. Prior studies have uncovered a link between the male-killing action of *Spiroplasma* and the host's dosage compensation machinery6,10,11. Genetic experiments revealed that *Spiroplasma* fails to kill males lacking the MSL components, while it can induce death in females ectopically expressing the MSL complex6,10,11. This suggests that *Spiroplasma* targets either the MSL complex directly or its downstream chromatin modifications (e.g. acetylation of histone H4 on lysine 16)19. Remarkably, we found that the expression of *Sp*AID triggers 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 *Spiroplasma* infection triggers DNA damage and segregation defects on the male X chromosome, thereby activating male-specific apoptosis11. To test whether *Sp*AID 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. In GFP-expressing control and *Sp*AID-expressing female embryos, a few H2Av foci were detected, whereas *Sp*AID-expressing male embryos exhibited numerous bright foci (Fig. 3a-c) that frequently overlapped with MSL1 signals (Fig. 3d). We also found that *Sp*AID expression in male embryos causes 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 *Sp*AID-expressing male embryos, 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 *Spiroplasma*-infected embryos (from stage 13 onward)10, probably because the high expression of *Sp*AID by the GAL4/UAS system results in much more severe effects than *Spiroplasma* infection. Finally, we examined the subcellular distribution of *Sp*AID in larval salivary glands, whose large cellular sizes facilitate detailed cytological analyses. In both sexes, *Sp*AID-GFP was found at plasma membranes, in the cytoplasm, and throughout nuclei, but was absent from the nucleolus (Fig. 4a, b). Strikingly, *Sp*AID-GFP was enriched on MSL1-labelled chromosomes in male nuclei (Fig. 4b). This result is consistent with its reliance on the dosage compensation machinery, even though equivalent localization analyses in embryos are needed.

To better characterize the action of *Sp*AID, 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 *Sp*AID 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 *Sp*AID (second instar larval stage) (Fig. 1c). Furthermore and in contrast to wild-type form *Sp*AID,

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. Interestingly, although its distribution pattern was otherwise indistinguishable from that of *Sp*AID-GFP in salivary gland cells, ΔANK-GFP was not enriched on MSL-labelled chromosomes (Fig. 4c, d). On the other hand, ΔOTU-GFP was only weakly detected within nuclei regardless of sex, and colocalization with MSL was not apparent (Fig. 4e, f). ΔOTU's male killing activity when overexpressed may be due to excess amounts of protein overriding its localization defect. All of these observations can be reconciled in a model of chromosomal targeting of *Sp*AID in which the OTU domain promotes nuclear localization and ankyrin repeats interact with the MSL complex or downstream histone modifications (Extended Data Fig. 7).

In summary, we have discovered a novel *Spiroplasma* gene that is likely to be responsible for male killing. Remarkably, expression of this single gene was sufficient to recapitulate the phenotypes associated with male killing. Our analysis revealed that *Sp*AID is the sole ankyrin repeat protein in the *Spiroplasma* genome (Supplementary Data). This contrasts with the case of *Wolbachia*, whose genomes encode more than 20 ankyrin repeat proteins21,22. Interestingly, the *Wolbachia* strain *w*Mel has a prophage-associated gene WD0633, that contains ankyrin repeats and the OTU domain23. Nevertheless, its transgenic expression in *D. melanogaster* produced no obvious phenotypes in a previous report24. Recent studies discovered *Wolbachia* genes that cause cytoplasmic incompatibility, a reproductive manipulation whereby symbiont-free females are unable to reproduce when mated with infected males25–27. Intriguingly, some of the genes that cause cytoplasmic incompatibility and *Sp*AID 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 *Sp*AID, thereby deciphering the principle of its sex-specific activity. A thorough understanding of symbiont-induced reproductive manipulations would not only provide novel insights into fundamental aspects of development, sex determination, and their evolution in insects, but could also provide significant clues to control insect populations.

## 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) for one generation to eliminate possible contamination with *Wolbachia* and *Spiroplasma.* After treatment, about 10 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 using them 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, found to be *Wolbachia*/*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* (ref. 31) flies were generously provided by Takehide Murata (RIKEN) and Mitzi Kuroda (Harvard Medical School), respectively.

To express the *SpAID*-encoding gene, we employed the GAL4/UAS system32 (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 larvae33. For maternal expression34, *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 Figs 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 *msl2* coding sequence (lacking Sxl-binding sites and resistant to translational repression) under the control of a heat shock-inducible promoter31, but we used its leaky basal expression. The resultant embryos were distinguished by GFP staining and only GFP-positive embryos (wild-type for *msl1*) were analysed (Extended Data Fig. 5a-d). For the expression of *Sp*AID 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-positive wandering third instar larvae were dissected.

## Construction of transgenic fly lines

For the gene synthesis of the *Sp*AID coding sequence by the GeneArt service (Thermo Fisher Scientific), TGA stop codons, which encode tryptophan in *Spiroplasma*, were mutated into TGG. The codon usage was optimization for the expression in *D. melanogaster*. Consequently, GC content was modified from 23.3% to 50.4%. We divided the *Sp*AID coding sequence into two parts and synthesized them separately: one was a 2,367-bp fragment that corresponds to *Sp*AID ΔC and the other was an 838-bp fragment for the remaining 3' portion of full length *Sp*AID. To obtain the full length *Sp*AID coding sequence (3,195 bp, without a stop codon), two fragments were fused by PCR with a 30-bp overlap and cloned into the pENTR vector by the pENTR/D-TOPO cloning kit (Thermo Fisher Scientific). To generate *Sp*AID deletion constructs, we amplified two PCR fragments (nucleotide positions 1-276 and 787-3,195 for ΔANK; nucleotide positions 1-789 and 1,396-3,195 for ΔOTU) from the synthetic *Sp*AID coding sequence and fused by PCR with a 24-bp overlap and cloned into the pENTR vector. We utilized PrimeSTAR HS DNA Polymerase (Takara Bio) for all PCR reactions. The Gateway cassette containing the *Sp*AID fragments was transferred into the pPWG destination vector (The *Drosophila* Genomics Resource Center #1078; The *Drosophila* Gateway Vector Collection by Terence Murphy) 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 *Spiroplasma* strain

Male-killing *Spiroplasma* can be easily transferred to other fly stocks by hemolymph injection2. *Spiroplasma* containing hemolymph was collected from a naturally infected *D. melanogaster* line Ug-SR (ref. 35) provided by John Jaenike (University of Rochester), and transferred to wild-type Oregon-R flies that were used as the source of hemolymph 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 (they were used in ref. 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 hemolymph from these fly stocks and injected them into Oregon-R flies to exclude the effect of host genetic background. We confirmed that only the *Spiroplasma* strain collected from the *Sxl-EGFP* line showed partial male killing (Extended Data Fig. 1b, c). The male-killing *Spiroplasma* strain in *D. melanogaster* is conventionally designated as MSRO (*melanogaster* sex ratio organism). Therefore, we call these *Spiroplasma* 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 study35.

We suspect that the partial male-killing strain might have resulted from genome rearrangements as reported in other *Spiroplasma* species maintained in the laboratory36. 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 (e.g. *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. Since the titre of the partial male-killing strain appears to be higher than the original strain in some aged females, this may have led the propagation of mutant bacteria and promoted their transmission (Extended Data Fig. 1d).

### Analysis of *Spiroplasma* 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, and 14 days and individually transferred to a 1.5 ml tube and stored at -80°C. Then, they were homogenized with plastic pestles by hand, and genomic DNA was purified by the DNeasy Blood & Tissue Kit (Qiagen). A portion of the *Spiroplasma dnaA* gene and the *D. melanogaster RpS17* gene were amplified with specific primer sets [dnaA_109F/246R (ref. 38) and rps17_615F/695R (ref. 39), respectively] (Supplementary Table 1) under the following PCR conditions: 95°C for 5 min followed by 45 cycles of 95°C for 10 sec, 55°C for 20 sec and 72°C for 20 sec. The 10-μl PCR mixture contained 2 μl of genomic DNA, 1x LightCycler 480 SYBR Green I master mix (Roche), 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 *Spiroplasma dnaA* gene against the host *RpS17* gene, we followed the Pfaffl method40. To estimate the PCR efficiency of each primer set, we performed quantitative PCR by using six 10-fold serial dilutions of genomic DNA purified from 10 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 *Spiroplasma*

To collect a sufficient amount of DNA for whole-genome sequencing of *Spiroplasma* strains MSRO-H99 and MSRO-SE, we recovered fly hemolymph by the centrifugal separation method developed by Laura Musselman (Washington University School of Medicine)41 with minor modification. 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 $CO_2$ anesthesia, 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 rpm (2,300 rcf) for 5 min at 4°C. We usually made 3-10 cartridges at one time and they were kept on ice until centrifugation. After removing cartridges, we checked precipitates that contain host hemolymph and *Spiroplasma* 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 12,000 rcf for 3 min at 4°C to eliminate large microorganisms including yeasts. The flow-through was centrifuged at the 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 sec, followed by 1 hr incubation at 56°C. Then, 4 μl RNase A (100 mg/ml; 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). During the extraction, samples were mixed by pulse-vortexing within 20 sec to prevent shearing of high molecular weight DNA. We collected hemolymph 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). 4 μg (MSRO-H99) or 5 μg (MSRO-SE) of the sheared DNA 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) in 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 genome16 and ordered by Mauve v2.4.0 (refs 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), respectively. These extra chromosomes could be plasmids because they contain several proteins, like Soj (the chromosome partitioning protein) and ARPs (adhesion-related proteins; P58, P12, P54, P123 and P18), which are located on plasmids in other *Spiroplasma* species45.

Whole-genome annotation was conducted by using Prokka v1.11 (ref. 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 by the NCBI Conserved Domain Database (NCBI CDD)51,52 with default parameters (Supplementary Data). Besides *Sp*AID (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 (#4; 87,892 bp) in MSRO-H99. It is likely that they are derived from the miss or partial assembly of *Sp*AID-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 10x) compared to the *Sp*AID locus (from 51,206 to 54,403 bp) with high coverage (over 330x). These genes were not found in MSRO-SE. Further domains were predicted by the InterPro protein sequence analysis and classification database53. The protein domain structure of *Sp*AID was drawn by the Illustrator for Biological Sequences (IBS) software v 1.0.2 (ref. 54) and modified manually.

## Homology searches

For homology searches, we utilized 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 *Sp*AID 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 *Sp*AID [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: $2.63x10^{-118}$, identities: 58.03%, positives: 69.63%, gaps: 14.07%), respectively].

### Sanger sequencing of the *SpAID* locus

10 adult Oregon-R females uninfected (as a negative control) or infected with MSRO-Ug, MSRO-H99, and MSRO-SE were collected and genomic DNA was purified (see quantitative PCR methods). The 3' portion of *SpAID* gene was amplified with forward and reverse primer sets (spaid_1568F with spaid.L_+136R for MSRO-Ug/MSRO-H99 and spaid.S_+362R for MSRO-SE, respectively) (Supplementary Table 1) under the following PCR conditions: 95°C for 2 min followed by 30 cycles of 95°C for 30 sec, 50°C for 30 sec and 60°C for 3 min. Due to the low GC content of the *SpAID* gene (23.3%), extension at 60°C instead of 72°C was essential for the successful PCR amplification55. The 50-μl PCR mixture contained 2 μl of genomic DNA, 1.25 U of GoTaq G2 DNA polymerase (Promega), 1x green reaction buffer (Promega), 0.5 μM primers, and 0.2 mM dNTP mixture. The PCR fragment was purified from agarose gel by the Wizard SV Gel and PCR Clean-Up System (Promega) and read by direct sequencing. To design sequencing primes, 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 *Spiroplasma* 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 according to refs 56,57. Embryos were collected from grape juice agar plates and dechorionated in 2.8% sodium hypochlorite solution, 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 4ºC 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-Sex lethal and anti-MSL1 antibodies were utilized for sexing embryos (see below).

The following primary antibodies were used: mouse anti-Sex lethal [1:20; Developmental Studies Hybridoma Bank (DSHB), M18]58, mouse and rabbit anti-MSL1 [1:200 and 1:500; kindly provided by John Lucchesi (Emory University)], 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; Sigma, D9542) for 10 min at room temperature after secondary antibody staining. TUNEL (terminal

deoxynucleotidyl transferase dUTP nick end labelling) 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 4ºC 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 *Spiroplasma* in host hemolymph (Fig. 1a), the abdomen of an adult female was dissected in 5 μl PBS with tweezers on a microscope slide. 1 μl 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 *Sp*AID, 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 by the Fiji software (Fiji Is Just ImageJ)60 with a custom macro. Image analysis was performed by custom R scripts with the EBImage package61. For the quantification of TUNEL signals of whole embryos (Figs 2c and Extended Data Fig. 5d) (acquired by a 20x/0.8 objective with 0.6x zoom scan; frame size: 512 x 512; 1.5 μm, two times optimal intervals), maximum projection images of DAPI and TUNEL staining were binarized and the former was utilized 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), stages 8-10 embryos were triply stained (pH2Av, MSL1, and DAPI) and two images were acquired per embryo (by a 63x/1.4 oil immersion objective; frame size: 1,024 x 1,024; 0.3 μm optimal intervals). We compiled 20 serial z-sections from the top (among 25 sections in total) to make projected images of pH2Av and MSL1 signals. These images were smoothed by Gaussian filter and binarized by moving average method, respectively. To identify focal pH2Av signals, image objects were extracted from the stacked images by segmentation and labelling, where objects not greater 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. Above datasets were also utilized 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 by the level tool in the GIMP 2.8 software (the GNU General Public License). The adjustment was performed uniformly on the entire images and only black/white input levels were modified.

Europe PMC Funders Author Manuscripts

## Statistics and reproducibility

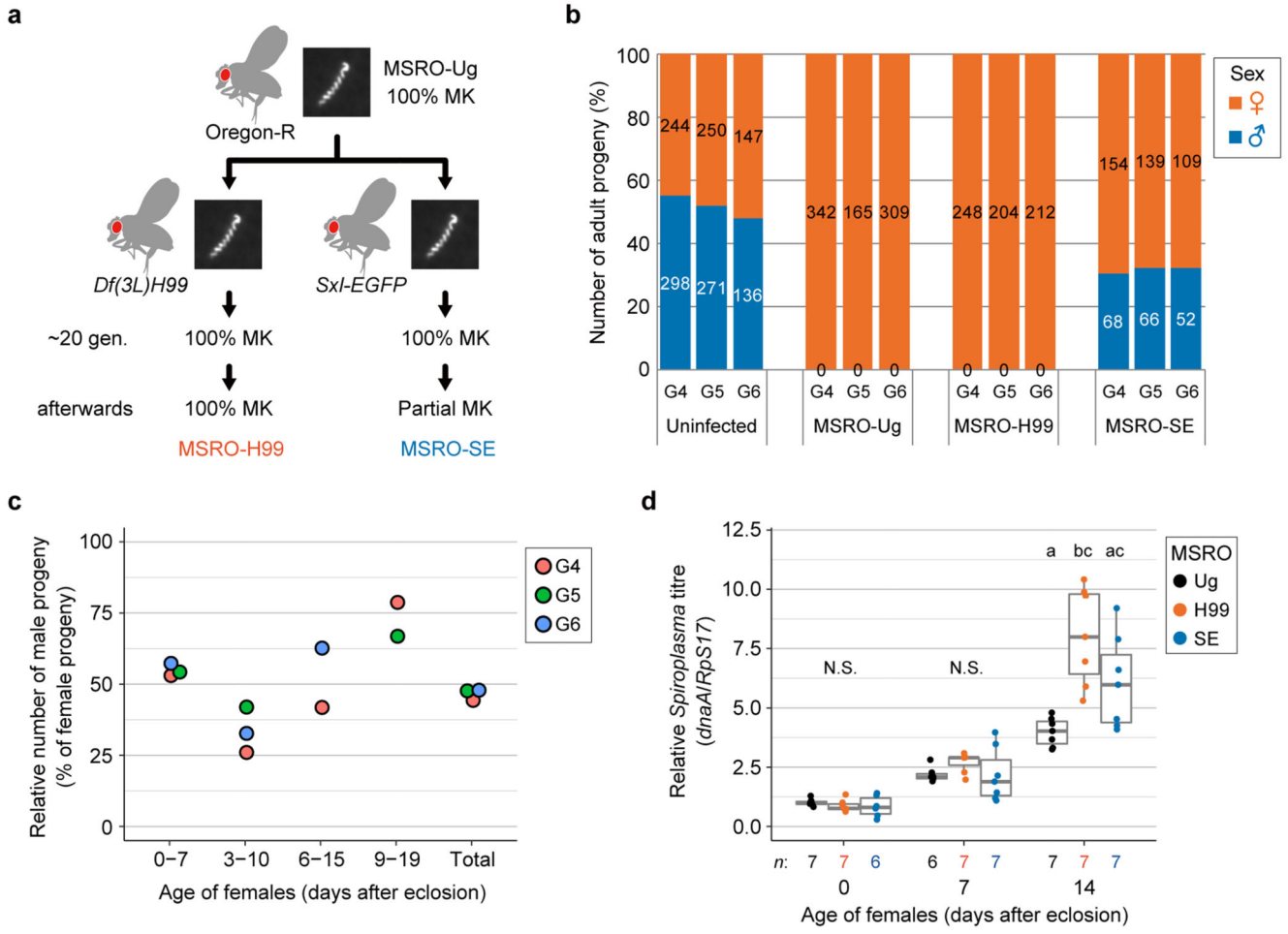
No statistical methods were applied to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. We used the R software v3.4.1 (the R Foundation) for all statistical analyses. Multiple comparisons in Figs 1c, 2c, 3c and Extended Data Figs 1d, 5d were performed using the Steel-Dwass test by the pSDCFlig function in the NSM3 R package62. We performed the $\chi^2$ 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).

## 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 Figs 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 available upon reasonable request.

Europe PMC Funders Author Manuscripts

## Extended Data




**Extended Data Figure 1. Identification and characterization of the partial male-killing *Spiroplasma* strain.**

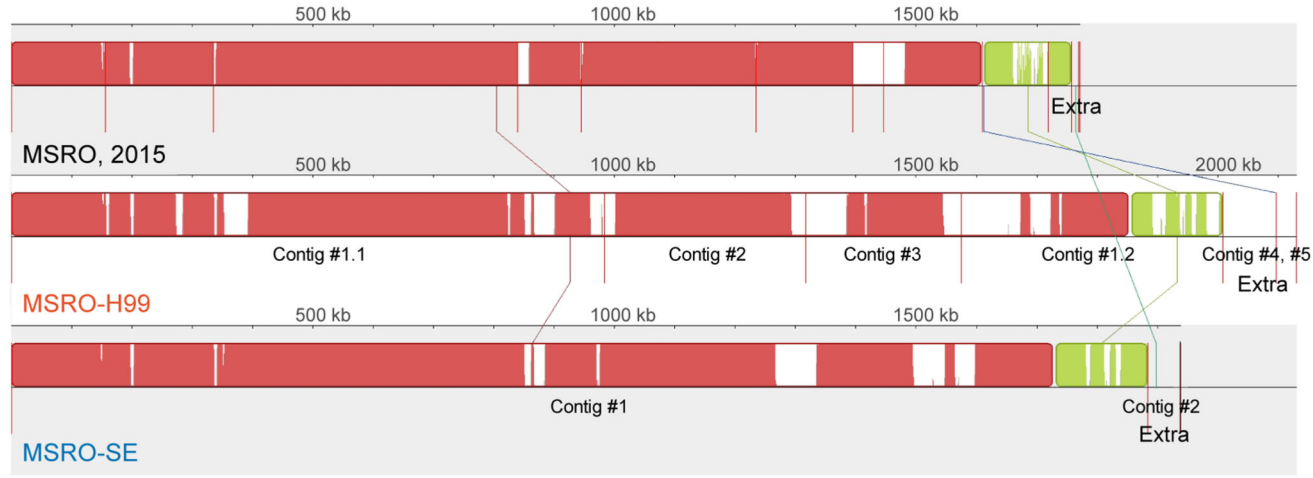
**a,** An illustration showing the origin of the *Spiroplasma* strains analysed in this study. Pictures show male-killing (MK) *Spiroplasma* of *D. melanogaster*. MSRO-Ug is the original male-killing strain maintained in the Oregon-R wild-type fly. Fly stocks [*Df(3L)H99* and *Sxl-EGFP*] artificially 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, while the other (MSRO-H99) kept the ability to induce complete male killing. See Methods 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 for three times at fourth- to sixth-generations (G4-6) after the establishment of infection. In **c**, the relative number of male offspring (% of females) obtained from Oregon-R female flies infected with MSRO-SE were plotted. The data point was excluded if the total count of flies was below 10. **d,** Relative titre of *Spiroplasma* within individual female fly. Adult females infected with three MSRO strains were aged for 0, 7, 14 days after eclosion and analysed by qPCR. Data were normalized with respect to 0 day

females infected with MSRO-Ug. Different characters indicate statistically significant differences ($P < 0.01$; N.S., not significant, $P > 0.05$; Steel-Dwass test; see Supplementary Table 2). Please note that the titres of 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**

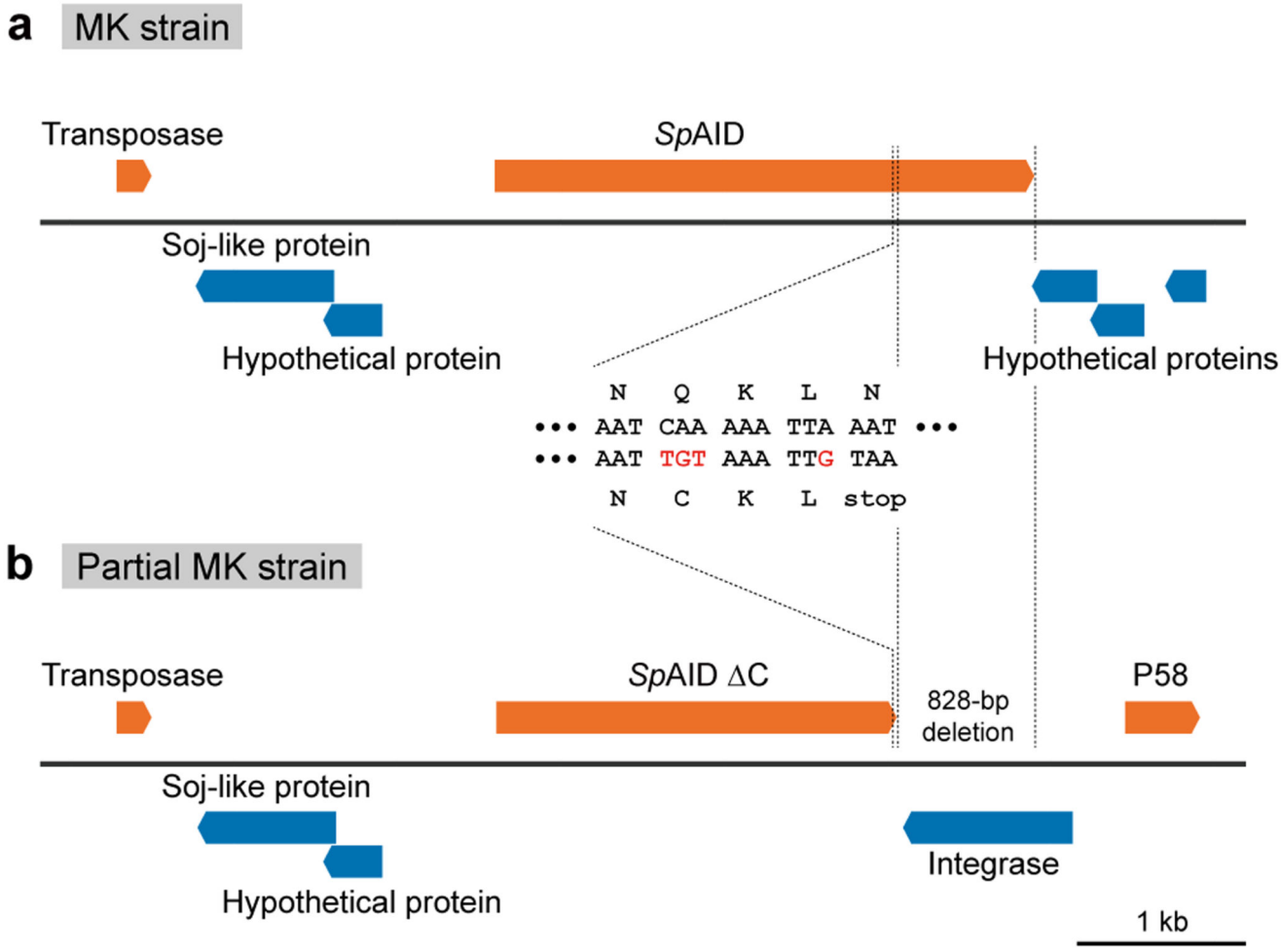



**Extended Data Figure 2. Whole-genome sequencing studies of the male-killing *Spiroplasma* variants.**




Europe PMC Funders Author Manuscripts

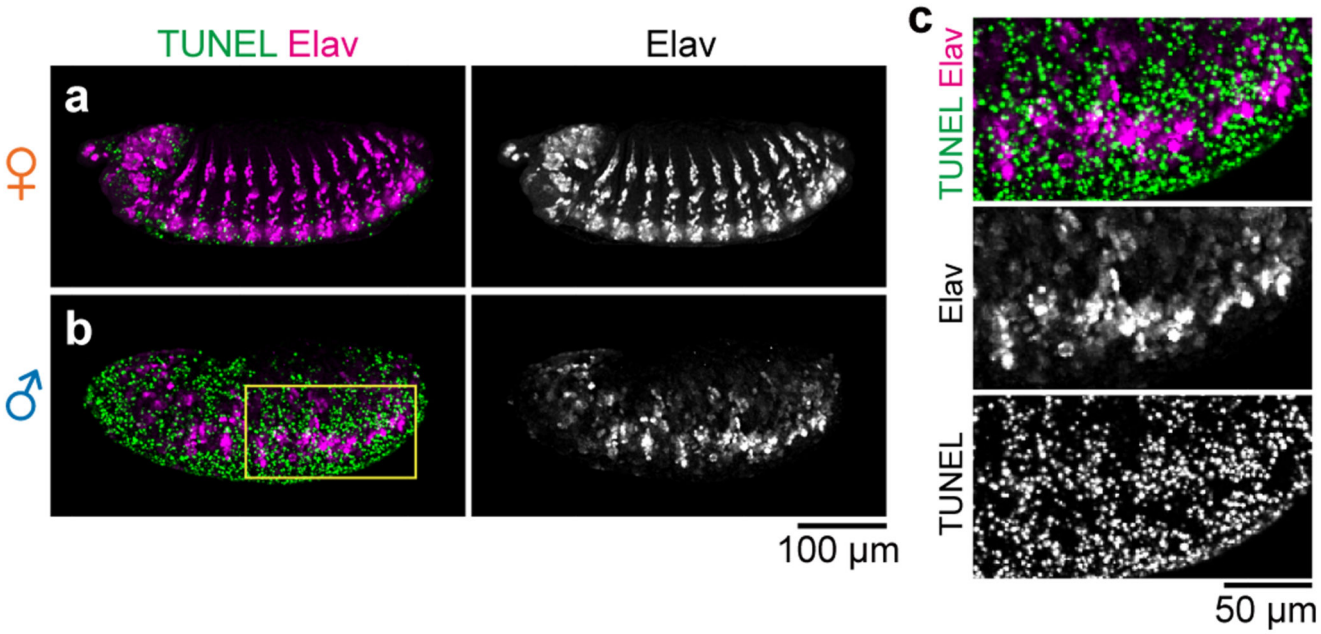
**a,** The comparison of genomic features of three *Spiroplasma* strains. MSRO-H99 and MSRO-SE are newly obtained variants isolated in this study (Extended Data Fig. 1a; see also Methods). As a control, the data of the previously reported male-killing *Spiroplasma* genome16 is also indicated (MSRO, 2015). **b**, Whole-genome alignment of the three *Spiroplasma* strains. To start the alignment from the *dnaA* gene, the 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) are shown as "extra". *Sp*AID (gene ID: SMH99_26490) and *Sp*AID ΔC (gene ID: SMSE_25110) are located on these extra chromosomes in MSRO-H99 (contig #4) and MSRO-SE (contig #2), respectively.




**Extended Data Figure 3. Genetic alterations of the *SpAID* locus in the partial male-killing *Spiroplasma* 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) *Spiroplasma* 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 by the Sanger method (see Methods).

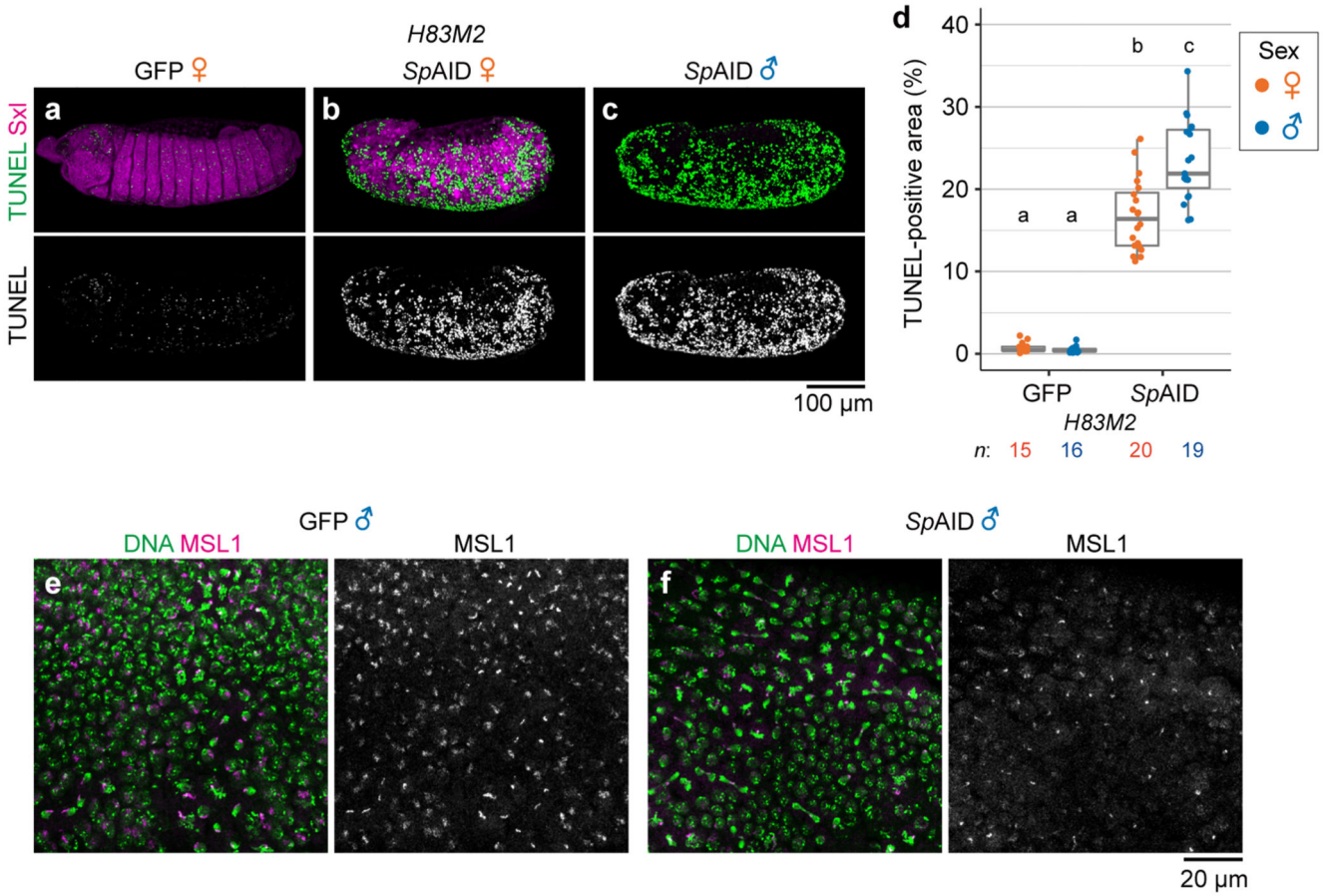



**Extended Data Figure 4. Neural defects of *Sp*AID-expressing embryos.**

Representative images of stage 13-14 female (**a**, $n = 14$) and male (**b**, $n = 16$) embryos maternally expressing *Sp*AID, stained for TUNEL (green) and neural cells (Elav, magenta). Single-channel images of Elav are also shown. The boxed 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.

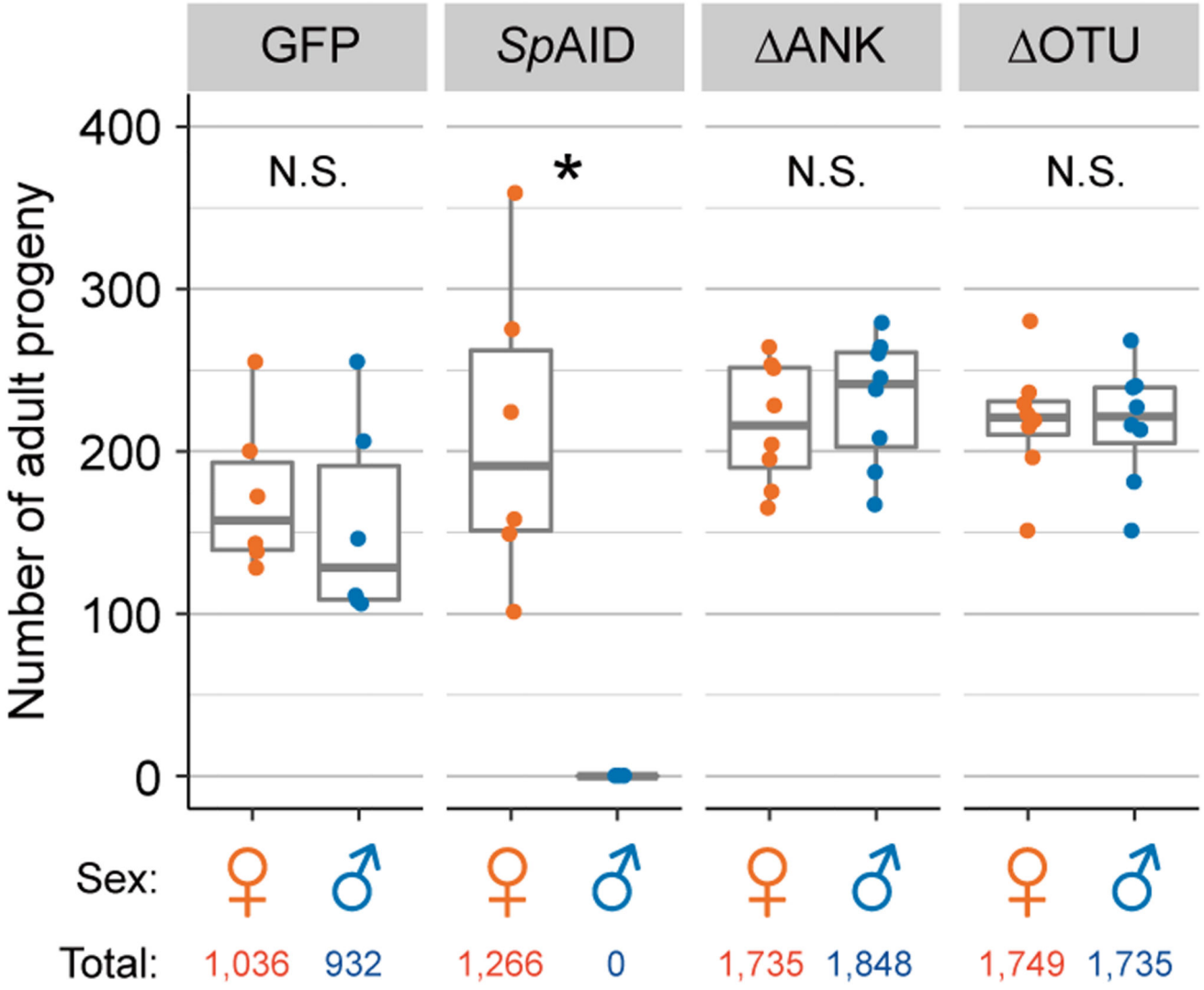



Europe PMC Funders Author Manuscripts



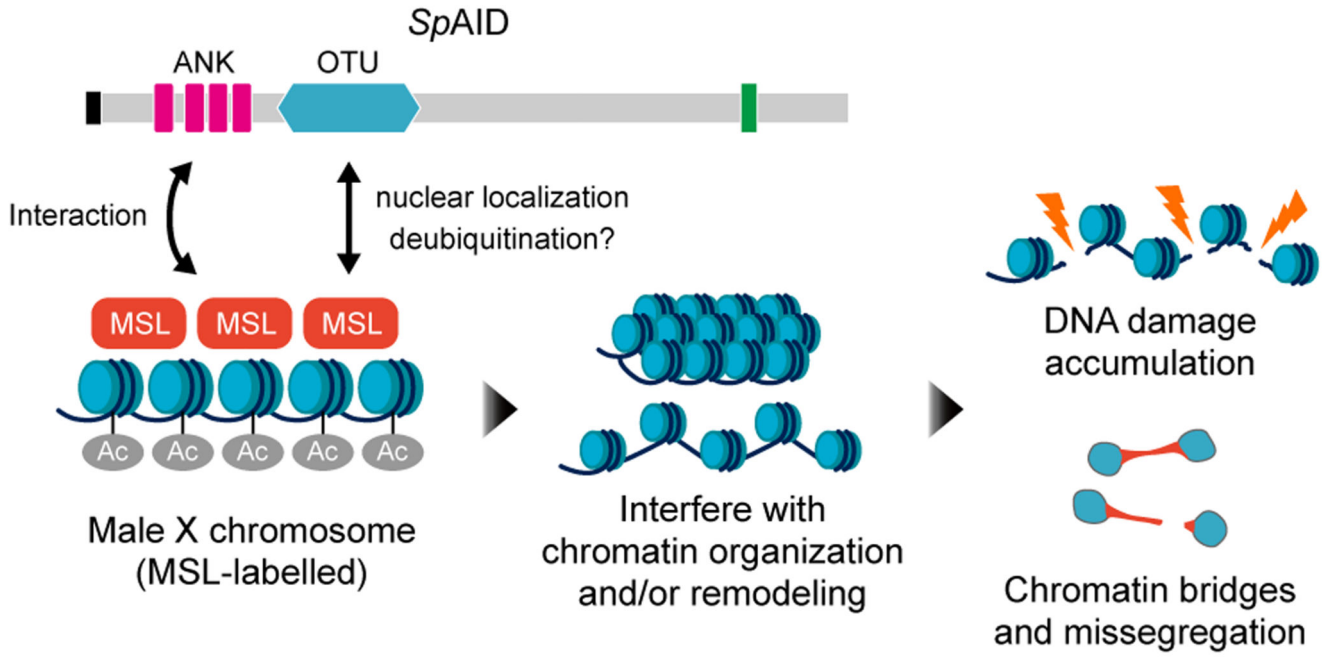

**Extended Data Figure 5. *Sp*AID acts through the dosage compensation machinery.**
**a-c,** Representative images of stages 13-14 embryos ectopically expressing the MSL complex by the *H83M2* transgene, stained for TUNEL (green) and Sxl (magenta). GFP-expressing control female (**a**, $n = 15$), *Sp*AID-expressing 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 significant differences ($P < 0.001$; 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 stages 8-10 male embryos expressing GFP (**e**, $n = 25$) and *Sp*AID (**f**, $n = 25$), stained for DNA (green) and MSL1 (magenta) from the datasets analysed in Fig. 3. All *UAS* transgenes were expressed maternally.

Europe PMC Funders Author Manuscripts




**Extended Data Figure 6. Expression of *Sp*AID by 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, *Sp*AID, ΔANK, and ΔOTU; $n = 6$ independent crosses for GFP and *Sp*AID, $n = 8$ independent crosses for ΔANK and ΔOTU). The *UAS-GFP* line was used as a negative control. With this weak GAL4 driver, *Sp*AID 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 Figure 7. A proposed model for *Sp*AID-induced male-killing phenotypes.**

*Sp*AID 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.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank the Bloomington Stock Center in USA and the Department of *Drosophila* Genomics and Genetic Resources at Kyoto Institute of Technology in Japan (DGGR) for fly stocks. We also thank the Developmental Studies Hybridoma Bank at the University of Iowa in USA for monoclonal antibodies, John Jaenike, Takehide Murata, and Mitzi Kuroda for providing fly strains, John Lucchesi for providing antibodies, the BioImaging & Optics Platform (BIOP) in EPFL for confocal microscopy, Keith Harshman and the Lausanne Genomic Technologies Facility (GTF) in the University of Lausanne (UNIL) for whole-genome sequencing and draft genome assemblies, Christoph Vorburger and Steve Perlman for comments and suggestions on the manuscript, and the members of the laboratory for discussion and support. We also thank Takema Fukatsu for the *Spiroplasma*-infected fly stocks that were originally established in his laboratory by T.H. 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.

## References

1. Hurst GDD, Frost CL. Reproductive parasitism: maternally inherited symbionts in a biparental world. Cold Spring Harb Perspect Biol. 2015; 7:a017699. [PubMed: 25934011]
2. Williamson, DL., Poulson, DF. Sex Ratio Organisms (Spiroplasmas) of *Drosophila*. The Mycoplasmas, Volume III: Plant and Insect Mycoplasmas. Whitcomb, RF., Tully, JG., editors. Vol. Ch. 6. Academic Press; 1979. p. 175-208.

3. Williamson DL, 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. 1999; 49:611–618. [PubMed: 10319483]
4. Haselkorn TS. The *Spiroplasma* heritable bacterial endosymbiont of *Drosophila*. Fly (Austin). 2010; 4:80–87. [PubMed: 20081357]
5. Tsuchiyama-Omura S, Sakaguchi B, Koga K, Poulson DF. Morphological features of embryogenesis in *Drosophila melanogaster* infected with a male-killing *Spiroplasma*. Zoolog Sci. 1988; 5:375–383.
6. Veneti Z, Bentley JK, Koana T, Braig HR, Hurst GDD. A functional dosage compensation complex required for male killing in *Drosophila*. Science. 2005; 307:1461–1463. [PubMed: 15746426]
7. Bentley JK, Veneti Z, Heraty J, Hurst GDD. The pathology of embryo death caused by the male-killing *Spiroplasma* bacterium in *Drosophila nebulosa*. BMC Biol. 2007; 5:9. [PubMed: 17362512]
8. Martin J, Chong T, Ferree PM. Male killing *Spiroplasma* preferentially disrupts neural development in the *Drosophila melanogaster* embryo. PLoS ONE. 2013; 8:e79368. [PubMed: 24236124]
9. Harumoto T, Anbutsu H, Fukatsu T. Male-killing *Spiroplasma* induces sex-specific cell death via host apoptotic pathway. PLoS Pathog. 2014; 10:e1003956. [PubMed: 24550732]
10. Cheng B, Kuppanda N, Aldrich JC, Akbari OS, Ferree PM. Male-killing *Spiroplasma* alters behavior of the dosage compensation complex during *Drosophila melanogaster* embryogenesis. Curr Biol. 2016; 26:1339–1345. [PubMed: 27161498]
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. 2016; 7:12781. [PubMed: 27650264]
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. 1971; 18:45–56. [PubMed: 5003804]
13. Makarova KS, Aravind L, Koonin EV. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and *Chlamydia pneumoniae*. Trends Biochem Sci. 2000; 25:50–52. [PubMed: 10664582]
14. Al-Khodor S, Price CT, Kalia A, Abu Kwaik Y. Functional diversity of ankyrin repeats in microbial proteins. Trends Microbiol. 2010; 18:132–139. [PubMed: 19962898]
15. Dobrindt U, Hochhut B, Hentschel U, Hacker J. Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol. 2004; 2:414–424. [PubMed: 15100694]
16. Paredes JC, et al. Genome sequence of the *Drosophila melanogaster* male-killing *Spiroplasma* strain MSRO endosymbiont. MBio. 2015; 6:e02437–14. [PubMed: 25827421]
17. Campos AR, Rosen DR, Robinow SN, White K. Molecular analysis of the locus *elav* in *Drosophila melanogaster*: a gene whose embryonic expression is neural specific. EMBO J. 1987; 6:425–431. [PubMed: 3107982]
18. Salz HK, Erickson JW. Sex determination in *Drosophila*: the view from the top. Fly (Austin). 2010; 4:60–70. [PubMed: 20160499]
19. Lucchesi JC, Kuroda MI. Dosage compensation in *Drosophila*. Cold Spring Harb Perspect Biol. 2015; 7:a019398. [PubMed: 25934013]
20. Madigan JP, Chotkowski HL, Glaser RL. DNA double-strand break-induced phosphorylation of *Drosophila* histone variant H2Av helps prevent radiation-induced apoptosis. Nucleic Acids Res. 2002; 30:3698–3705. [PubMed: 12202754]
21. Siozios S, et al. The diversity and evolution of *Wolbachia* ankyrin repeat domain genes. PLoS ONE. 2013; 8:e55390. [PubMed: 23390535]
22. Jernigan KK, Bordenstein SR. Ankyrin domains across the Tree of Life. PeerJ. 2014; 2:e264. [PubMed: 24688847]
23. Wu M, et al. Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements. PLoS Biol. 2004; 2:E69. [PubMed: 15024419]
24. Yamada R, Iturbe-Ormaetxe I, Brownlie JC, O'Neill SL. Functional test of the influence of *Wolbachia* genes on cytoplasmic incompatibility expression in *Drosophila melanogaster*. Insect Mol Biol. 2011; 20:75–85. [PubMed: 20854481]

Europe PMC Funders Author Manuscripts


25. Beckmann JF, Fallon AM. Detection of the *Wolbachia* protein WPIP0282 in mosquito spermathecae: implications for cytoplasmic incompatibility. Insect Biochem Mol Biol. 2013; 43:867–878. [PubMed: 23856508]
26. Beckmann JF, Ronau JA, Hochstrasser M. A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. Nat Microbiol. 2017; 2:17007. [PubMed: 28248294]
27. LePage DP, et al. Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. Nature. 2017; 543:243–247. [PubMed: 28241146]
28. Zhou W, Rousset F, O'Neil S. Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc Biol Sci. 1998; 265:509–515. [PubMed: 9569669]
29. Fukatsu T, Nikoh N. Two intracellular symbiotic bacteria from the mulberry psyllid *Anomoneura mori* (Insecta, Homoptera). Appl Environ Microbiol. 1998; 64:3599–3606. [PubMed: 9758773]
30. Fukatsu T, Nikoh N. Endosymbiotic microbiota of the bamboo pseudococcid *Antonina crawii* (Insecta, Homoptera). Appl Environ Microbiol. 2000; 66:643–650. [PubMed: 10653730]
31. Kelley RL, et al. Expression of Msl-2 causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. Cell. 1995; 81:867–877. [PubMed: 7781064]
32. Brand AH, Perrimon N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development. 1993; 118:401–415. [PubMed: 8223268]
33. Ashburner, M., Golic, KG., Hawley, RS. Drosophila: A Laboratory Handbook. 2nd Edition. Cold Spring Harbor Laboratory Press; 2005.
34. Rørth P. Gal4 in the *Drosophila* female germline. Mech Dev. 1998; 78:113–118. [PubMed: 9858703]
35. Pool JE, Wong A, Aquadro CF. Finding of male-killing *Spiroplasma* infecting *Drosophila melanogaster* in Africa implies transatlantic migration of this endosymbiont. Heredity (Edinb). 2006; 97:27–32. [PubMed: 16685282]
36. Ye F, Melcher U, Rascoe JE, Fletcher J. Extensive chromosome aberrations in *Spiroplasma citri* Strain BR3. Biochem Genet. 1996; 34:269–286. [PubMed: 8894049]
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. 2013; 5:1151–1164. [PubMed: 23711669]
38. Anbutsu H, Fukatsu T. Population dynamics of male-killing and non-male-killing spiroplasmas in *Drosophila melanogaster*. Appl Environ Microbiol. 2003; 69:1428–1434. [PubMed: 12620825]
39. Osborne SE, Leong YS, O'Neill SL, Johnson KN. Variation in antiviral protection mediated by different *Wolbachia* strains in *Drosophila simulans*. PLoS Pathog. 2009; 5:e1000656. [PubMed: 19911047]
40. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001; 29:e45. [PubMed: 11328886]
41. Musselman, LP. *Drosophila* hemolymph collection procedure. YouTube. 2013. Available at: https://www.youtube.com/watch?v=im78OIBKlPA
42. Field D, et al. Open software for biologists: from famine to feast. Nat Biotechnol. 2006; 24:801–803. [PubMed: 16841067]
43. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004; 14:1394–1403. [PubMed: 15231754]
44. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE. 2010; 5:e11147. [PubMed: 20593022]
45. Saillard C, et al. The abundant extrachromosomal DNA content of the *Spiroplasma citri* GII3-3X genome. BMC Genomics. 2008; 9:195. [PubMed: 18442384]
46. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; 30:2068–2069. [PubMed: 24642063]
47. Carle P, et al. Partial chromosome sequence of *Spiroplasma citri* reveals extensive viral invasion and important gene decay. Appl Environ Microbiol. 2010; 76:3420–3426. [PubMed: 20363791]
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. 2012; 11:224–236. [PubMed: 22129229]

49. Lo W-S, Chen L-L, Chung W-C, Gasparich GE, Kuo C-H. Comparative genome analysis of *Spiroplasma melliferum* IPMB4A, a honeybee-associated bacterium. BMC Genomics. 2013; 14:22. [PubMed: 23324436]
50. Davis RE, et al. Complete genome sequence of *Spiroplasma kunkelii* strain CR2-3x, causal agent of corn stunt disease in *Zea mays* L. Genome Announc. 2015; 3
51. Marchler-Bauer A, et al. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 2011; 39:D225–229. [PubMed: 21109532]
52. Marchler-Bauer A, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015; 43:D222–226. [PubMed: 25414356]
53. Finn RD, et al. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 2017; 45:D190–D199. [PubMed: 27899635]
54. Liu W, et al. IBS: an illustrator for the presentation and visualization of biological sequences. Bioinformatics. 2015; 31:3359–3361. [PubMed: 26069263]
55. Su XZ, Wu Y, Sifri CD, Wellems TE. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. Nucleic Acids Res. 1996; 24:1574–1575. [PubMed: 8628694]
56. Hartenstein, V. Atlas of *Drosophila* Development. Cold Spring Harbor Laboratory Press; 1993.
57. Campos-Ortega, JA., Hartenstein, V. The Embryonic Development of *Drosophila* melanogaster. Springer-Verlag; 1997.
58. Bopp D, Bell LR, Cline TW, Schedl P. Developmental distribution of female-specific Sex-lethal proteins in *Drosophila melanogaster*. Genes Dev. 1991; 5:403–415. [PubMed: 1900493]
59. O'Neill EM, Rebay I, Tjian R, Rubin GM. The activities of two Ets-related transcription factors required for *Drosophila* eye development are modulated by the Ras/MAPK pathway. Cell. 1994; 78:137–147. [PubMed: 8033205]
60. Schindelin J, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012; 9:676–682. [PubMed: 22743772]
61. Pau G, Fuchs F, Sklyar O, Boutros M, Huber W. EBImage—an R package for image processing with applications to cellular phenotypes. Bioinformatics. 2010; 26:979–981. [PubMed: 20338898]
62. Hollander, MA., Wolfe, D., Chicken, E. Nonparametric Statistical Methods. 3rd Edition. John Wiley & Sons, Inc; 2013.


Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

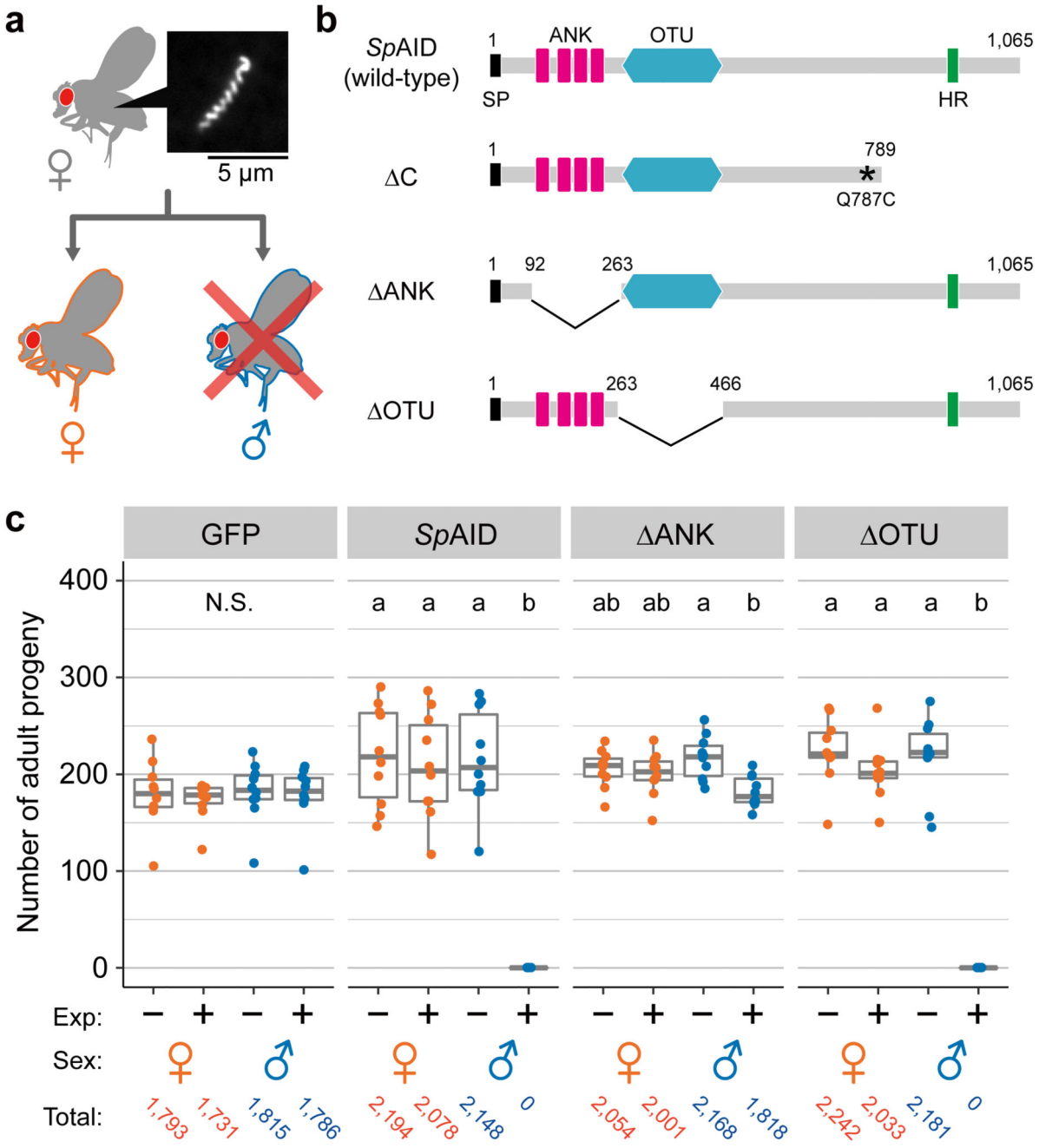



**Figure 1. Expression of *Sp*AID selectively eliminates male offspring.**

**a,** *Spiroplasma*-induced male killing in *Drosophila*. Infected females (top) produce only female offspring (bottom). A picture shows a male-killing *Spiroplasma* of *D. melanogaster* detected by DNA staining. **b,** Protein structure of *Sp*AID, 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). *Sp*AID ΔC of the partial male-killing strain encodes a protein with an amino acid substitution (Q787C) and C-terminal truncation. The structures of two deletion constructs of *Sp*AID (ΔANK and ΔOTU) are also

indicated. The numbers represent amino acid (aa) residues. **c,** The number of adult progeny obtained from crosses between the *Actin-GAL4* line and four *UAS* transgenic lines (GFP, *Sp*AID, 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) expressing the transgenes (+, having both *Actin-GAL4* and *UAS* transgenes) and siblings not expressing the transgenes (-, having only *UAS* transgenes) as internal controls. Different characters indicate statistically significant differences ($P < 0.0001$, $P < 0.05$ for ΔANK; N.S., not significant, $P > 0.05$; Steel-Dwass test; see Supplementary Table 2). 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.



Europe PMC Funders Author Manuscripts

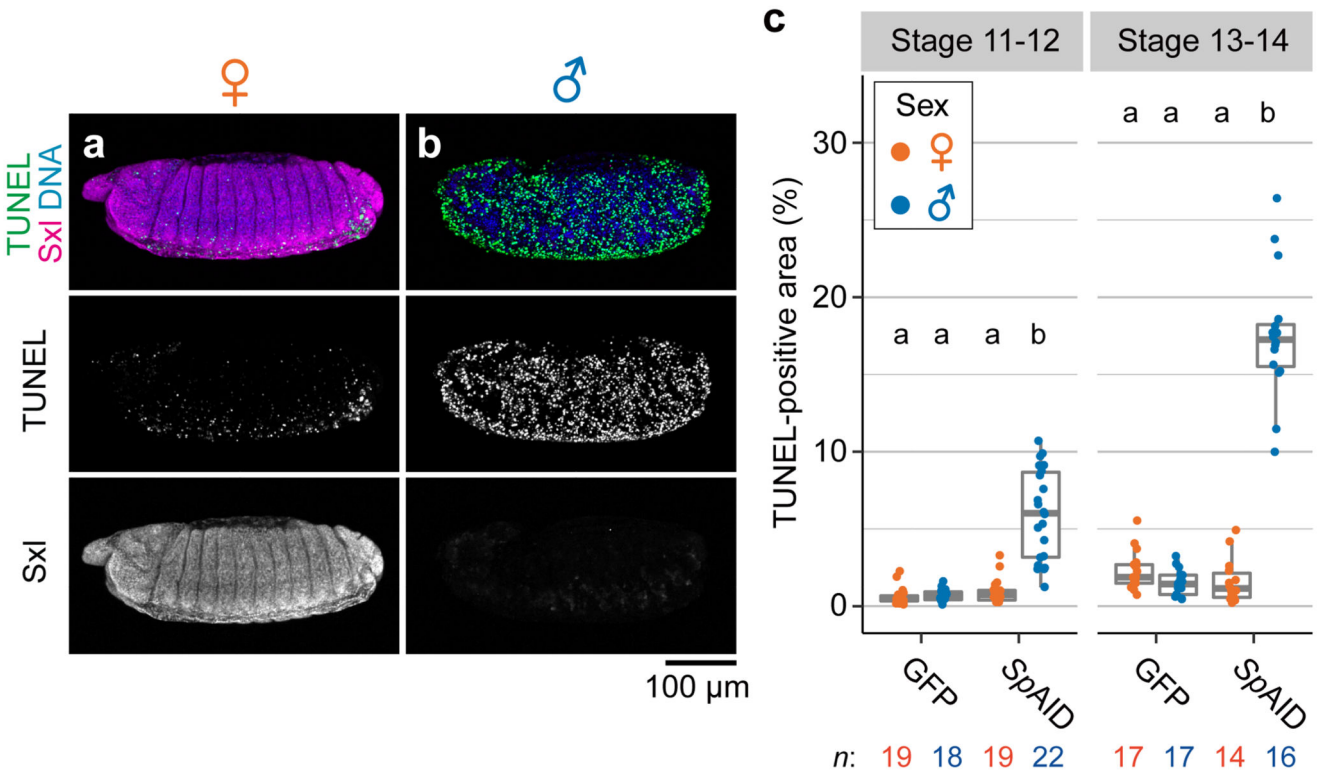



**Figure 2. Expression of *Sp*AID reproduces male-killing phenotypes during embryogenesis.**

**a, b,** Representative images ofstages 13-14 female (**a**, $n = 14$) and male (**b**, $n = 16$) embryos maternally expressing *Sp*AID, 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 stages 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 are 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 Figure 4.

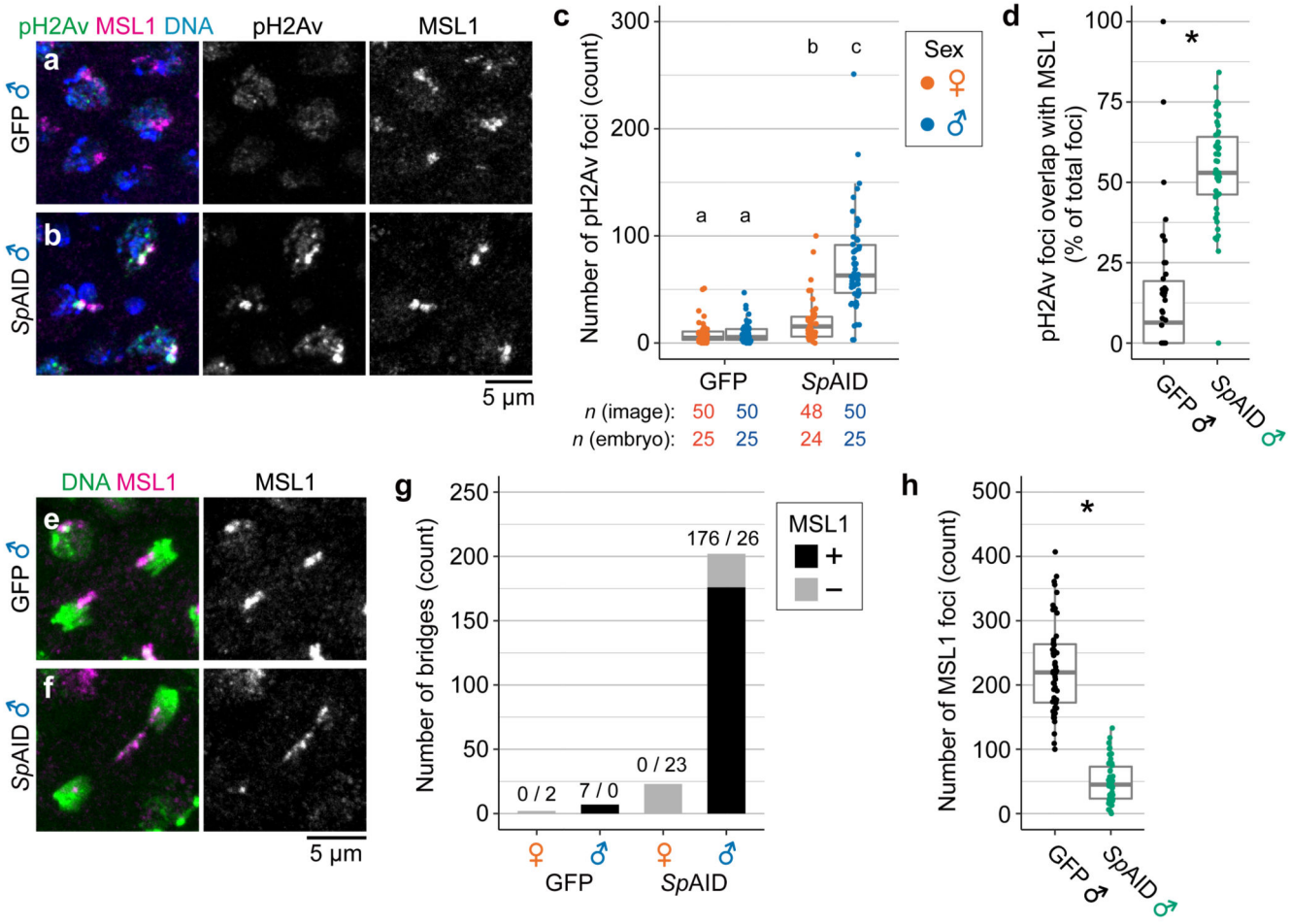



**Figure 3. *Sp*AID acts through the MSL complex.**

**a, b,** Epithelial cells in stage 9 male embryos expressing GFP (**a**) and *Sp*AID (**b**), stained for pH2Av (green), MSL1 (magenta), and DNA (blue). **c,** Quantification of pH2Av foci in embryos expressing GFP and *Sp*AID. **d,** Percentage of pH2Av foci overlapping with MSL1 signals in male embryos expressing GFP [median (interquartile range): 6.3% (0-19.3%)] and *Sp*AID [52.9% (46.2-64.2%)]. **e, f,** Dividing cells in stage 9 male embryos expressing GFP (**e**, proper segregation) and *Sp*AID (**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 *Sp*AID. **h,** The number of MSL1 focal signals in male embryos expressing GFP (black) and *Sp*AID (green). The same datasets of stages 8-10 embryos were analysed in **a-h** ($n = 50$ or 48 images per condition). Different characters or asterisks indicate significant differences in **c** ($P < 0.001$; Steel-Dwass test), **d** ($P < 0.0001$; $\chi^2$ 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.

Europe PMC Funders Author Manuscripts

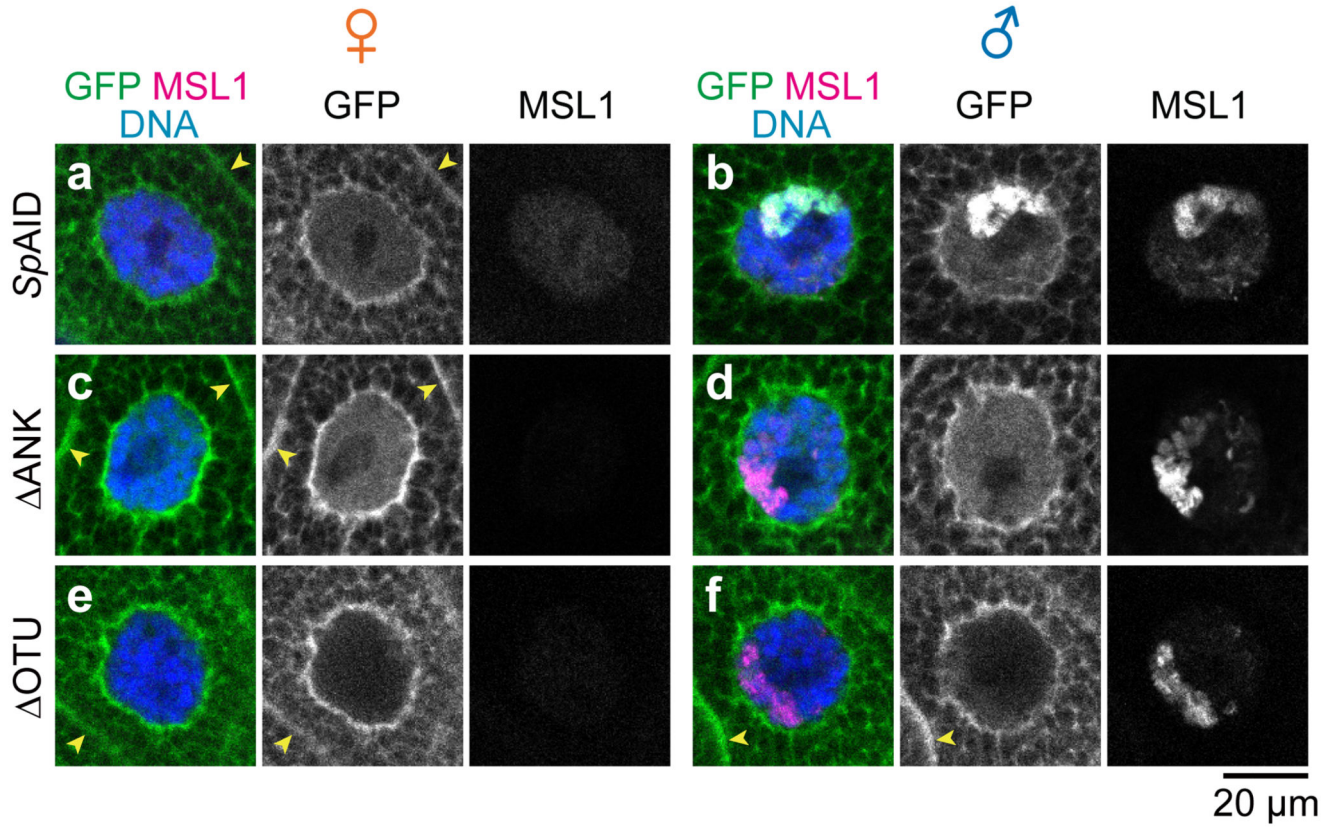



**Figure 4. Subcellular localization of *Sp*AID.**

**a-f,** Larval salivary glands expressing *Sp*AID-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 fluorescent 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.