

# Island-specific evolution of a sex-primed autosome in a sexual planarian

<https://doi.org/10.1038/s41586-022-04757-3>

Received: 15 September 2021

Accepted: 12 April 2022

Published online: 1 June 2022

Open access

 Check for updates

Longhua Guo<sup>1,2</sup>✉, Joshua S. Bloom<sup>1,2</sup>, Daniel Dols-Serrate<sup>3</sup>, James Boocock<sup>1,2</sup>, Eyal Ben-David<sup>4</sup>, Olga T. Schubert<sup>1,2</sup>, Kaiya Kozuma<sup>1</sup>, Katarina Ho<sup>1</sup>, Emily Warda<sup>1</sup>, Clarice Chui<sup>1</sup>, Yubao Wei<sup>5</sup>, Daniel Leighton<sup>1,2</sup>, Tzitziki Lemus Vergara<sup>1,2</sup>, Marta Riutort<sup>3</sup>, Alejandro Sánchez Alvarado<sup>2,6</sup> & Leonid Kruglyak<sup>1,2</sup>✉

The sexual strain of the planarian *Schmidtea mediterranea*, indigenous to Tunisia and several Mediterranean islands, is a hermaphrodite<sup>1,2</sup>. Here we isolate individual chromosomes and use sequencing, Hi-C<sup>3,4</sup> and linkage mapping to assemble a chromosome-scale genome reference. The linkage map reveals an extremely low rate of recombination on chromosome 1. We confirm suppression of recombination on chromosome 1 by genotyping individual sperm cells and oocytes. We show that previously identified genomic regions that maintain heterozygosity even after prolonged inbreeding make up essentially all of chromosome 1. Genome sequencing of individuals isolated in the wild indicates that this phenomenon has evolved specifically in populations from Sardinia and Corsica. We find that most known master regulators<sup>5–13</sup> of the reproductive system are located on chromosome 1. We used RNA interference<sup>14,15</sup> to knock down a gene with haplotype-biased expression, which led to the formation of a more pronounced female mating organ. On the basis of these observations, we propose that chromosome 1 is a sex-primed autosome primed for evolution into a sex chromosome.

Sex chromosomes evolve from homologous autosomes that acquire sex-determining genes and lose their ability to recombine<sup>16–21</sup>. As such, sex chromosome evolution and recombination suppression are closely associated<sup>16–21</sup>. However, because direct evidence of such homologous autosomes primed for evolution into sex chromosomes is difficult to capture, little is known about the molecular signatures associated with the evolution of recombination suppression.

The freshwater planarian *Schmidtea mediterranea*, an important model organism for studies of regeneration<sup>22,23</sup>, exists as asexual and sexual reproductive strains. The sexual strain is distributed mostly in Tunisia and on the islands of Sardinia, Corsica and Sicily<sup>1</sup>. The sexual strain is a simultaneous hermaphrodite that develops both male and female reproductive systems in the same adult individual and obligately outcrosses to fertilize other individuals<sup>24,25</sup>. Individuals in the asexual strain do not develop sexual reproductive systems. We considered that studying chromosome evolution in a simultaneous hermaphrodite might provide insight into the early evolution of a primitive sex chromosome.

*S. mediterranea* has four pairs of chromosomes, which are stably diploid. The genome is reported to comprise 774 Mb assembled as 481 non-contiguous series of genomic sequences, or scaffolds<sup>26–28</sup>. A previous study found that approximately 300 Mb of the genome remained heterozygous even after extensive inbreeding of laboratory strains, and that this phenomenon also occurs naturally in wild populations in Sardinia<sup>24</sup>. The two sets of heterozygous alleles were collectively named J and V haplotypes. To define the chromosomal locations of

these alleles and investigate the reasons underpinning the persistence of heterozygosity in *S. mediterranea*, a detailed assembly of all four chromosomes is needed.

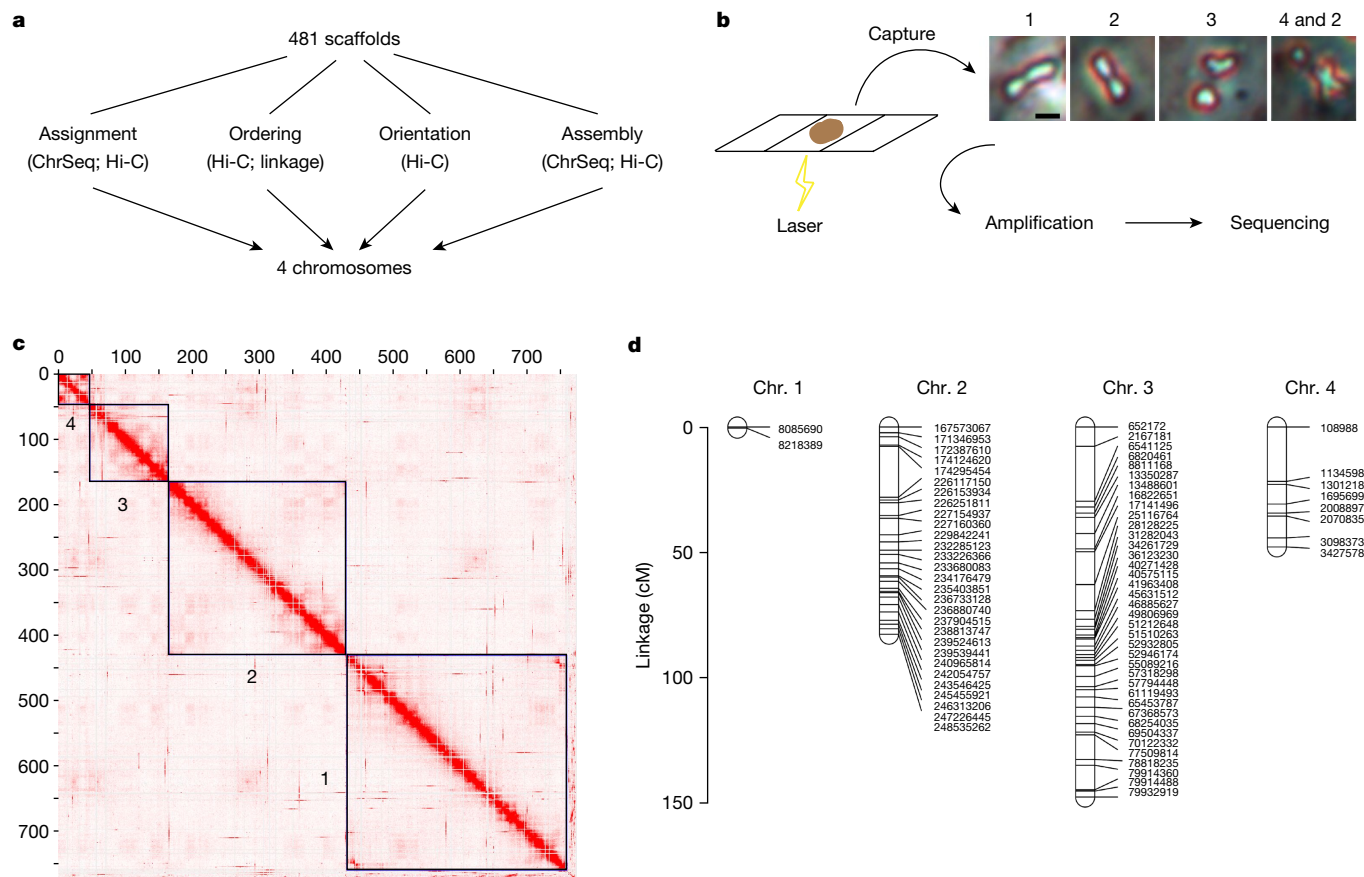
## Chromosome-scale genome assembly

To transform the 481 scaffolds<sup>26</sup> into a chromosome-scale genome reference, we carried out chromosome sequencing (ChrSeq)<sup>29,30</sup> of a laboratory strain, as well as chromatin proximity sequencing by Hi-C<sup>3</sup> (Fig. 1a). To do so, we dissected individual chromosomes from mitotic cells using laser capture and amplified and sequenced each chromosome individually (Fig. 1b). We examined the sequencing depth of each scaffold across multiple samples of the same and different chromosomes to ensure reproducibility and specificity, respectively (Extended Data Fig. 1). Overall, we successfully amplified and confidently assigned 740 Mb of the 774-Mb genome to one of four chromosomes (Supplementary Table 1).

We used ChrSeq information and chromatin interaction data generated by Hi-C to correct and connect the individual scaffolds within a chromosome into a chromosome-scale genome, hereafter referred to as Smed\_chr\_ref\_v1. The Hi-C data analysed using the SALSA scaffolding algorithm<sup>3</sup> resolved the 481 scaffolds into 57 super-scaffolds and 104 singletons. The ChrSeq data indicated that 3 of the 57 super-scaffolds were disjointed inter-chromosomal fragments, consistent with the Hi-C contact heat map (Fig. 1c). We split the disjointed super-scaffolds, merged the scaffolds into chromosomes, and ordered and oriented

<sup>1</sup>Department of Human Genetics and Department of Biological Chemistry, University of California, Los Angeles, Los Angeles, CA, USA. <sup>2</sup>Howard Hughes Medical Institute, Chevy Chase, MD, USA.

<sup>3</sup>Departament de Genètica, Microbiologia i Estadística, Institut de Recerca de la Biodiversitat, Universitat de Barcelona, Barcelona, Spain. <sup>4</sup>Department of Biochemistry and Molecular Biology, Institute for Medical Research Israel-Canada, Hebrew University of Jerusalem-Hadassah Medical School, Jerusalem, Israel. <sup>5</sup>Institute of Reproductive Medicine, Henan Provincial People's Hospital, Zhengzhou University, Zhengzhou, China. <sup>6</sup>Stowers Institute for Medical Research, Kansas City, MO, USA. ✉e-mail: longhuaguo@mednet.ucla.edu; lkruglyak@mednet.ucla.edu



**Fig. 1 | Chromosome-scale genome assembly.** **a**, Schematic of the use of Hi-C, ChrSeq and a linkage map to transform 481 scaffolds<sup>26</sup> into a chromosome-scale genome assembly. **b**, Chromosomes from mitotic cells spread onto membrane slides for laser capture and sequencing. Numbers denote the identity of chromosomes present in the representative samples. **c**, Contact

heat map of chromatin interactions indicated by Juicebox<sup>4</sup>; black boxes denote the four chromosomes in the final assembly. **d**, Linkage map generated through LinkageMapView<sup>68</sup>, showing a lack of recombination on chromosome 1. Tick marks and labels indicate genetic markers.

all scaffolds within the chromosomes using Juicebox visualization software<sup>4</sup> (Methods). Chromosome assignments by ChrSeq alone (Supplementary Table 1) and by Hi-C alone were inconsistent for only 3 of 384 (0.8%) scaffolds (Supplementary Table 2). We manually assigned these three scaffolds to chromosomes according to the Hi-C data. Hi-C also detected 26 inter-chromosomal or intra-chromosomal assembly errors (Supplementary Table 2) in the previous assembly<sup>26</sup>, 5 of which were confirmed by ChrSeq to be inter-chromosomal disjointed scaffolds (Supplementary Table 2). The final genome assembly, *Smed\_chr\_ref\_v1* (Fig. 1c), had four chromosomes with a total size of 764 Mb, which is 98.4% of that reported previously<sup>26</sup>. Of the 1.6% of the previous assembly not contained in *Smed\_chr\_ref\_v1* by Hi-C, about half (52.6%) of the scaffolds could not be assigned to a chromosome or were assigned to two chromosomes by ChrSeq, indicating lower assembly quality (Supplementary Table 3). Moreover, 33.0% and 12.4% of the scaffolds had chromatin interaction signals with structurally complex regions of chromosomes 1 and 2, respectively (Extended Data Fig. 2), suggesting that they might be alternative assemblies or repetitive sequences. In our new assembly, *Smed\_chr\_ref\_v1*, the two ends of chromosome 4 were capped by >1,000 copies of the telomere repeat TTAGGG, indicating high assembly quality.

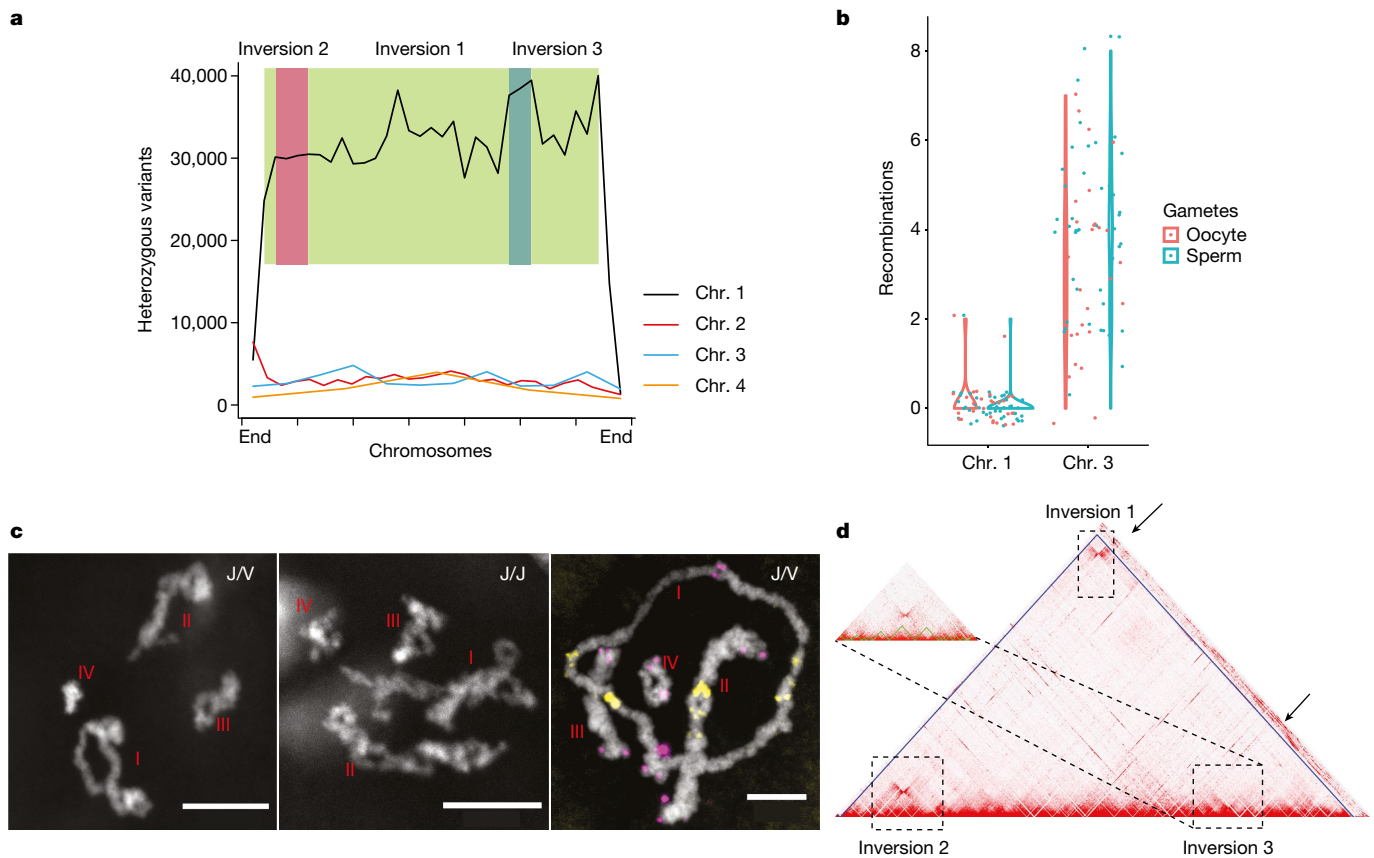
To validate the linearity of the chromosomes in *Smed\_chr\_ref\_v1*, we generated a linkage map. We crossed two divergent laboratory strains of *S. mediterranea*, S2F10b and D5, to produce an F<sub>2</sub> population, and we genotyped individual worms with RADseq<sup>31</sup>. Eighty markers that were evenly distributed and genotyped in at least 98% of the F<sub>2</sub> segregants (91 of 93) were used to establish four linkage groups (Fig. 1d

and Supplementary Table 4) representing the four chromosomes. The ordering of the 80 markers in the linkage map is consistent with *Smed\_chr\_ref\_v1*, independently supporting the quality of our chromosome-scale genome assembly. This highly contiguous and complete genome assembly and linkage map together facilitate further genetic and epigenetic investigations of the functions of the genome in this model planarian.

### Chromosome 1 recombination suppression

We next re-examined the heterozygous alleles that define the J and V haplotypes in the newly assembled genome. We found that 87.7% of the genetic markers that remained heterozygous were located on chromosome 1, spanning 333 Mb at a density of 30,148 variants per 10 Mb (Fig. 2a). The remaining 12.3% of the heterozygous markers were located on the other three chromosomes at a density of 3,274 variants per 10 Mb and probably correspond to differences between highly similar copies of repetitive elements rather than true polymorphisms (Supplementary Table 5). All F<sub>2</sub> worms ( $n = 93$ ) in this study were heterozygous for chromosome 1 but not for chromosomes 2–4 (Supplementary Table 6), which is consistent with the previous study<sup>24</sup>. Hence, we concluded that the J/V haplotypes were on chromosome 1.

Our linkage map revealed an extremely low rate of recombination on chromosome 1 of only 0.5 cM for the entire chromosome (Fig. 1d). This is particularly notable because, at 333 Mb, chromosome 1 was the largest of the four chromosomes, containing more than 40% of the genome.



**Fig. 2 | Recombination suppression and inversions on chromosome 1. a.** Distribution of heterozygous variants maintained in the S2 inbreeding pedigree along the four chromosomes. The y-axis shows variant counts per 10 Mb. Green, pink and blue boxes represent inversions 1, 2 and 3, respectively. **b.** Number of meiotic recombination events on chromosomes 1 and 3 in oocytes (red) and sperm cells (cyan). Dots represent individual gametes, summarized with violin plots. **c.** Chromosome 1 in late prophase I in oocytes from J/V worms (left and right panels) shows fewer crossovers than in oocytes

from J/J worms (middle panel). This experiment was repeated more than 10 times independently with similar results. The right panel shows FISH<sup>42</sup> results with probes for the telomeric repeat TTAGGG (magenta) and a repeat located near the centromere (yellow). This experiment was repeated more than three times independently with similar results. **d.** Chromatin contact heat map for chromosome 1. Dashed-line rectangles indicate the locations of potential inversions. Arrows indicate chromatin contact signals with inversion regions from unassigned scaffolds.

To directly examine whether chromosome 1 can recombine, we sequenced 45 single sperm cells and 28 single oocytes from J/V line S2 (Fig. 2d). Gamete sequencing is preferred because recombination events in hatchlings can be selected for by differential fertilization or embryonic lethality. We identified 3,197 single-nucleotide variants (SNVs) on chromosome 1 and 3,312 SNVs on chromosome 3, covering 99% of the length of each chromosome (Supplementary Table 7). The SNVs were distributed at a similar density across 20-Mb windows, with coefficients of variation of 0.38 and 0.31 for chromosomes 1 and 3, respectively (Supplementary Table 7). We observed that 98% of the sperm cells (44 of 45) and 93% of the oocytes (26 of 28) had no crossovers on chromosome 1. By contrast, most sperm cells and oocytes had crossovers on chromosomes 2, 3 and 4 (Fig. 2b, Extended Data Fig. 3a, b and Supplementary Table 7). We thus concluded that recombination on chromosome 1 was strongly suppressed.

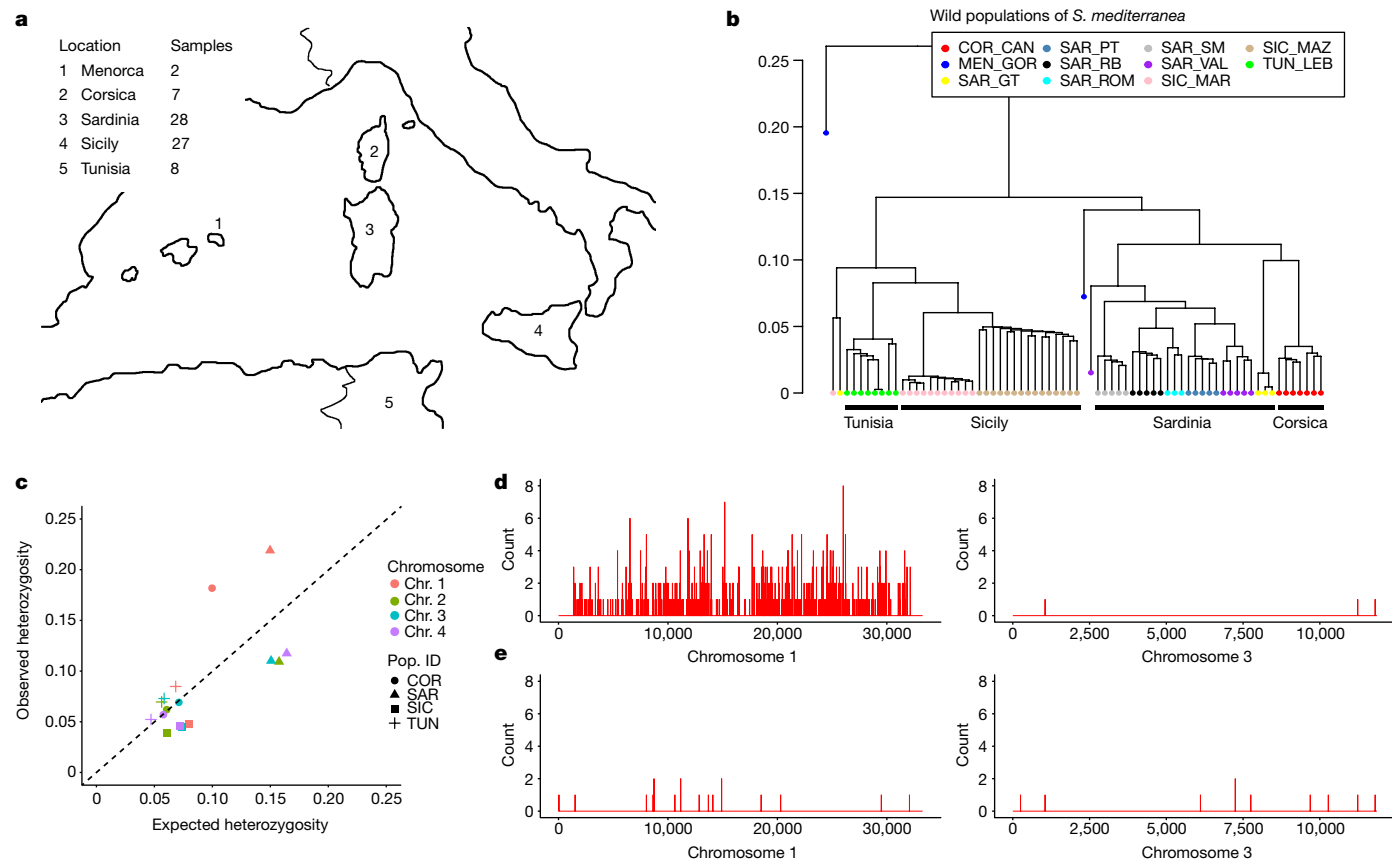
Consistent with this conclusion, we found that during prophase I, when other chromosomes had numerous crossovers, chromosome 1 formed a ring structure in the oocyte in J/V worms but not in J/J worms (Fig. 2c and Extended Data Fig. 4). Fluorescence in situ hybridization (FISH) with telomere probes suggested that crossovers between the two homologous pairs of chromosome 1 occurred only in regions close to the telomeres, leading to the observed ring conformation rather than the side-by-side pairing observed for chromosomes 2, 3 and 4. Furthermore, Hi-C analysis showed that chromosome 1 had three putative inversions, each >20 Mb in size (Fig. 2d); such inversions can cause

crossover suppression<sup>32–34</sup>. The rest of the genome had only one large inversion of approximately 10 Mb on chromosome 2 (Extended Data Fig. 2b).

### Island-specific evolution of chromosome 1

To investigate the genetic diversity of *S. mediterranea* in its natural environment and determine whether chromosome 1 J/V heterozygosity occurs throughout the entire species, we used RADseq<sup>31</sup> to sample the genomes of 70 sexual individuals from Sardinia, Corsica, Sicily and Tunisia and 2 asexual individuals from Menorca<sup>1</sup> (Fig. 3a). To look for genetic relationships between the individuals, we determined clustering of relatedness measured by identity-by-state pairwise distances<sup>35</sup> and identified two superclusters. Animals from Sicily were closely related to those from Tunisia, and animals from Sardinia were closely related to those from Corsica (Fig. 3b). Phylogenetic clustering (Extended Data Fig. 5), fixation index ( $F_{ST}$ ) values (Extended Data Fig. 6) and analyses with Structure (Extended Data Fig. 7) further supported this observation.

The relatedness of the populations from Sardinia and Corsica suggests that they may share genome characteristics that differ from those of the populations from Sicily and Tunisia. Indeed, animals in the populations from Sardinia and Corsica had greater heterozygosity on chromosome 1 than expected under Hardy–Weinberg equilibrium (Fig. 3c), whereas the heterozygosity on the other three chromosomes



**Fig. 3 | Island-specific evolution of chromosome 1 heterozygosity.**

**a**, Collection sites of 72 wild isolates from the Mediterranean islands and Tunisia. **b**, Maximum-likelihood tree of 2 asexual isolates from Menorca (MEN\_GOR) and 70 sexual isolates from Sardinia (SAR), Corsica (COR), Sicily (SIC) and Tunisia (TUN). Dots denote individual worms, with colours corresponding to their collection sites. **c**, Observed versus expected heterozygosity for each chromosome (denoted by colour) in each population (denoted by shape).

The dashed line corresponds to equality between observed values and those expected under Hardy–Weinberg equilibrium; the deviation of the points corresponding to chromosome 1 in the COR and SAR samples from the expectation is notable. **d**, **e**, Number of sites with heterozygosity in >80% of the population per 10-kb window along chromosome 1 (left) and chromosome 3 (right) in Sardinia (**d**) and Sicily (**e**).

in these populations and on all four chromosomes in the other populations closely followed expectations. By analysing the J/V haplotype markers in wild populations, we found that animals from Sardinia and Corsica were heterozygous J/V, whereas those from Sicily and Tunisia were homozygous J/J. Moreover, the animals from Sardinia ( $n = 28$ ) had many sites that were heterozygous in more than 80% of the individuals and were distributed along the length of chromosome 1 except near the ends; conversely, the animals from Sicily ( $n = 27$ ) had very few such heterozygous sites (Fig. 3d, e and Supplementary Table 8). Few such heterozygous sites were observed on chromosomes 2, 3 and 4 in either population (Fig. 3d, e and Extended Data Fig. 3c, d). These analyses suggest that chromosome 1 specifically evolved and diverged on the islands of Sardinia and Corsica.

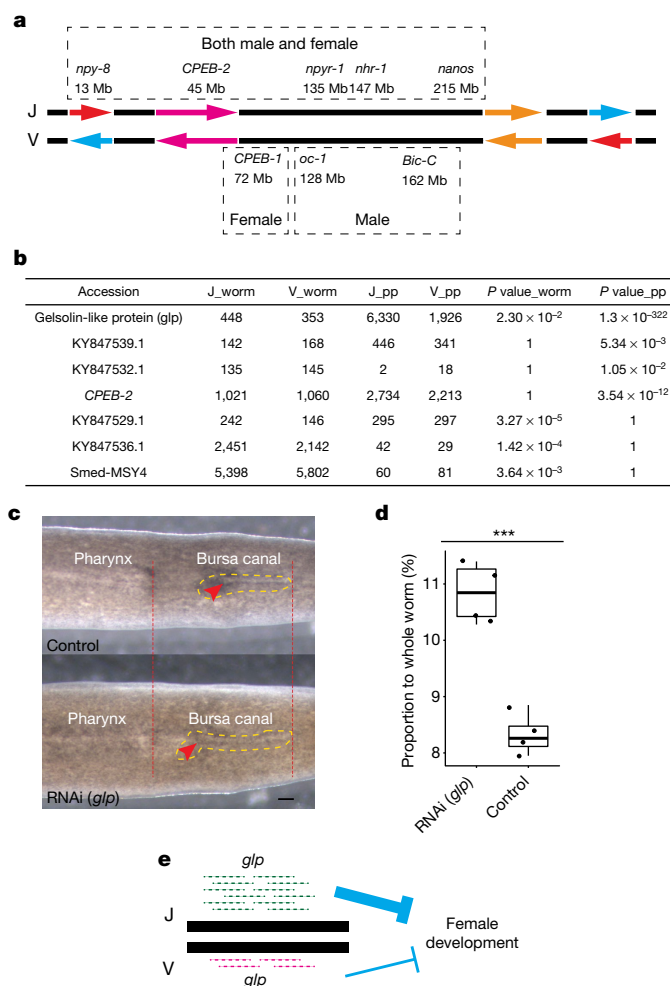
### A sex-primed chromosome

To gain insight into the island-specific suppression of recombination on chromosome 1, we examined the genes located on this chromosome, which contained 39% of all annotated genes. Five of the seven known master regulators of the reproductive system were found on chromosome 1, including *nanos*<sup>5</sup>, *nhr-1* (refs. <sup>8,9</sup>), *npv-8* (ref. <sup>6</sup>), *npvr-1* (ref. <sup>11</sup>) and *CPEB-2* (ref. <sup>12</sup>) (*ophis*<sup>11</sup> and *boule-2* (ref. <sup>10</sup>) were the two exceptions) (Fig. 4a). Knockdown of any one of these master regulators leads to depletion of both male and female reproductive tissues<sup>5,6,8–12</sup>. The presence of these genes with crucial roles in sexual development on chromosome 1 suggests that chromosome 1 integrity is important

for the maintenance of sexual reproduction. Indeed, the asexual lineage of *S. mediterranea*, which is devoid of any reproductive organs<sup>23</sup>, had a translocation from chromosome 1 to chromosome 3 and probably evolved through loss of function of one or more of these genes.

Chromosome 1 also contained three of the four known master regulators of male or female reproductive tissues (*CPEB-1* (ref. <sup>12</sup>), *onecut*<sup>13</sup> and *Bic-C*; *dmd-1* (ref. <sup>36</sup>) was the exception) (Fig. 4a). The *CPEB-1* gene is specifically required for the development of oocytes and yolk glands<sup>12</sup>. Loss of *onecut* and *Bic-C* expression leads to a ‘no-testes’ phenotype without affecting the ovary<sup>5,13</sup>. The presence of a female-determining gene on a chromosome that does not recombine provides an ideal foundation for the evolution of a sex chromosome. A loss-of-function mutation in the gene on one of the two homologous chromosomes would turn the chromosome with the mutation into a male-determining chromosome. Similar logic applies to a male-determining gene.

Haplotype-specific expression of sex-specific genes is a predicted signature of a sex-primed autosome. To test this prediction, we examined the expression of 20 genes in sexually mature adult worms and in the male copulatory organ, the penis papilla<sup>7</sup>, in a J/V line. The 20 genes were chosen because they were characterized as having sex-specific function and/or expression in the literature and include the 8 master regulators mentioned above<sup>5,12</sup> (Supplementary Table 9). Of these 20 genes, 13 contained at least one heterozygous variant in the coding sequence, which allowed us to determine whether these genes showed biased expression from the J or V haplotype (Extended Data Fig. 8). We found that a gelsolin-like protein<sup>12</sup>, *Smed-glp*, was expressed predominantly



**Fig. 4 | Acquisition and haplotype-specific expression of sex-related genes.**

**a**, Schematic diagram of the inferred structure of the J and V haplotypes on chromosome 1. Arrows denote putative inversions. Names, positions and sex specificity are shown for the eight genes on chromosome 1 with known key roles in the development of the reproductive system. **b**, Comparison of read counts for the J and V alleles of key genes in the transcriptomes of whole sexually mature worms and penis papillae (PP). Bonferroni-corrected *P* values from a two-sided binomial test of equal expression are shown. **c**, Dorsal view of control (top) and *Smed-glp* knockdown (bottom) sexually mature planarians. Red dashed lines and arrows indicate the posterior and anterior ends of the pharynx and the bursa canal, respectively. Scale bar, 400  $\mu$ m. Yellow dashed lines outline the bursa canal. **d**, Quantification of the length of the bursa canal relative to that of the whole worm. The y axis shows the percentage of whole-worm length spanned by the bursa canal as a box plot, with individual data points shown ( $n = 4$  each for control and knockdown worms). Statistical significance was assessed with a two-sided, two-sample Student's *t*-test ( $***P = 0.00029$ ). RNAi (*glp*): minimum, 10.28; maximum, 11.4; median, 10.845; first quartile, 10.375; third quartile, 11.31. Control: minimum, 7.95; maximum, 8.85; median, 8.26; first quartile, 8.06; third quartile, 8.6. **e**, Model of negative regulation of bursa canal development by *Smed-glp* alleles.

from the J haplotype in both whole worms and the penis papilla, with more haplotype-specific expression in the male copulatory organ. Four other genes showed biased expression from the J haplotype in whole worms (KY847529.1 and KY847536.1) or in the penis papilla (KY847539.1 and *CPEB-2*), and two additional genes showed biased expression from the V haplotype in the male copulatory organ (KY847532.1) or in whole worms (*Smed-MSY4*) (Fig. 4b and Supplementary Table 9). All seven of these genes were localized within the inversions or close to inversion breakpoints (Supplementary Table 5).

We used FISH with hybridization chain reaction to confirm that the expression of *Smed-glp* is highly enriched in both the penis papilla and the bursa canal<sup>17,37</sup> (Extended Data Fig. 9), which is a female organ used to receive sperm from mating partners. To examine the function of *Smed-glp*, we used RNA interference (RNAi)<sup>14,15</sup> to knock down its expression in young hatchlings. After 8 weeks of feeding, both the knockdown and control hatchlings reached sexual maturity. Notably, the bursa canal was much more pronounced in the knockdown animals (Fig. 4c, d). No obvious morphological differences were observed in the male copulatory organ. These results suggest that the J allele of *Smed-glp* may have a greater role in preventing over-development of the female bursa canal (Fig. 4e), probably by controlling the number of organ-specific muscle fibres<sup>38</sup>. This is consistent with a previous observation that J/J individuals show higher egg production, suggesting a larger contribution to female reproduction (Fig. 4b in ref. <sup>24</sup>; Supplementary Information).

We have thus identified a chromosome in the hermaphrodite planarian *S. mediterranea* that does not recombine, is enriched in master regulators of reproductive systems, shows allele-specific expression of sex-related genes, and contains genes orthologous to those on the sex chromosome of *Schistosoma mansoni* (Supplementary Tables 9 and 10, and Supplementary Information), which is the only known Platyhelminthes species with differentiated sex chromosomes<sup>39</sup>. These observations collectively led us to propose that chromosome 1 of *S. mediterranea* is primed for evolution into a sex chromosome.

Chromosome 1 of *S. mediterranea* marks an intriguing system for studying sex chromosome evolution<sup>40</sup>. Planarians from the islands of Sardinia and Corsica are heterozygous for the J and V haplotypes, whereas those in Sicily and Tunisia exist as J/J homozygotes. In laboratory crosses involving J/V lines, all hatchlings were J/V heterozygotes<sup>24</sup>. No V/V planarians have been identified either in nature or in laboratory crosses. We genotyped single zygotes from crosses between J/V individuals and determined that homozygous zygotes (J/J or V/V) exist (Extended Data Fig. 10a and Supplementary Information). We propose that early embryonic lethality leads to the loss of J/J and V/V adults in both nature and laboratory crosses. Such lethality may arise from degeneration of coding sequences on the J and V haplotypes as a consequence of crossover suppression. Indeed, chromosome 1 has an elevated rate of transposable elements and mutations introducing stop codons relative to the rest of the genome (Supplementary Table 11) as well as higher nonsynonymous substitution rates (Supplementary Table 13 and Extended Data Fig. 10b). As a consequence, the J and V haplotypes each carry unique sets of functional genes that are silenced or truncated on the other haplotype (Supplementary Table 11).

The presence of three nested inversions on chromosome 1 suggests the possibility that recombination suppression on this chromosome may have evolved in a stepwise manner similar to that reported for the human X and Y chromosomes<sup>17</sup>. In support of this possibility, we observed evidence for three evolutionary strata corresponding to the inverted regions. We used PacBio genome sequencing data<sup>26</sup> to identify long reads that bridged the inversion breakpoints, and we validated the three inversions identified by Hi-C (Supplementary Table 12 and Supplementary Information). We observed that the rates of heterozygous sites and synonymous substitutions were elevated within the three inversions in a pattern consistent with the evolutionary strata (Fig. 2a, Supplementary Tables 5 and 13, and Extended Data Fig. 10b, c). We used deep whole-genome sequencing data of individuals from laboratory crosses<sup>24</sup> to estimate the de novo mutation rate in sexual planarians to be approximately  $1.0 \times 10^{-8}$  mutations per nucleotide per generation. On the basis of this estimate, it is likely that inversions 3, 1 and 2 evolved approximately 450,000, 320,000 and 260,000 generations ago, respectively.

Although it is not possible to know how chromosome 1 will evolve, our findings provide a snapshot of an incipient stage that supports the

hypothesis that sex chromosomes evolve from homologous autosomes that acquire sex-specific roles and cease to recombine<sup>16–19</sup>. The locked J/V heterozygous system may facilitate this process by maintaining sex-specific alleles in the planarian population before the evolution of dioecy<sup>41</sup>. We propose that the planarian chromosome 1 haplotypes provide a unique opportunity to directly examine the molecular characteristics of a sex-primed autosome.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-022-04757-3>.

- Lazaro, E. M. et al. *Schmidtea mediterranea* phylogeography: an old species surviving on a few Mediterranean islands? *BMC Evol. Biol.* **11**, 274 (2011).
- Benazzi, M., Baguná, J., Ballester, R., Puccinelli, I. & Papa, R. D. Further contribution to the taxonomy of the «*Dugesia Lugubris-Polychroa Group*» with description of *Dugesia mediterranea* N.S.P. (Tricladida, Paludicola). *Boll. Zool.* **42**, 81–89 (1975).
- Burton, J. N. et al. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013).
- Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
- Wang, Y., Zayas, R. M., Guo, T. & Newmark, P. A. Nanos function is essential for development and regeneration of planarian germ cells. *Proc. Natl Acad. Sci. USA* **104**, 5901–5906 (2007).
- Collins, J. J. III et al. Genome-wide analyses reveal a role for peptide hormones in planarian germline development. *PLoS Biol.* **8**, e1000509 (2010).
- Chong, T., Stary, J. M., Wang, Y. & Newmark, P. A. Molecular markers to characterize the hermaphroditic reproductive system of the planarian *Schmidtea mediterranea*. *BMC Dev. Biol.* **11**, 69 (2011).
- Tharp, M. E., Collins, J. J. III & Newmark, P. A. A lophotrochozoan-specific nuclear hormone receptor is required for reproductive system development in the planarian. *Dev. Biol.* **396**, 150–157 (2014).
- Zhang, S. et al. A nuclear hormone receptor and lipid metabolism axis are required for the maintenance and regeneration of reproductive organs. Preprint at *bioRxiv* <https://doi.org/10.1101/279364> (2018).
- Iyer, H., Issigonis, M., Sharma, P. P., Extavour, C. G. & Newmark, P. A. A premeiotic function for boule in the planarian *Schmidtea mediterranea*. *Proc. Natl Acad. Sci. USA* **113**, E3509–E3518 (2016).
- Saberi, A., Jamal, A., Beets, I., Schoofs, L. & Newmark, P. A. GPCRs direct germline development and somatic gonad function in planarians. *PLoS Biol.* **14**, e1002457 (2016).
- Rouhana, L., Tasaki, J., Saberi, A. & Newmark, P. A. Genetic dissection of the planarian reproductive system through characterization of *Schmidtea mediterranea* CPEB homologs. *Dev. Biol.* **426**, 43–55 (2017).
- Li, P. et al. Single-cell analysis of *Schistosoma mansoni* identifies a conserved genetic program controlling germline stem cell fate. *Nat. Commun.* **12**, 485 (2021).
- Newmark, P. A., Reddien, P. W., Cebrià, F. & Alvarado, A. S. Ingestion of bacterially expressed double-stranded RNA inhibits gene expression in planarians. *Proc. Natl Acad. Sci. USA* **100**, 11861–11865 (2003).
- Rouhana, L. et al. RNA interference by feeding in vitro-synthesized double-stranded RNA to planarians: methodology and dynamics. *Dev. Dynam.* **242**, 718–730 (2013).
- Bachtrog, D. A dynamic view of sex chromosome evolution. *Curr. Opin. Genet. Dev.* **16**, 578–585 (2006).
- Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
- Rice, W. R. Evolution of the Y sex chromosome in animals. *BioScience* **46**, 331–343 (1996).
- Charlesworth, B. The evolution of sex chromosomes. *Science* **251**, 1030–1033 (1991).

- Muller, H. J. Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. *Genetics* **3**, 422–499 (1918).
- Charlesworth, D. Evolution of recombination rates between sex chromosomes. *Philos. Trans. R. Soc. Lond. B* **372**, 20160456 (2017).
- Reddien, P. W. & Sanchez Alvarado, A. Fundamentals of planarian regeneration. *Annu. Rev. Cell Dev. Biol.* **20**, 725–757 (2004).
- Newmark, P. A. & Sanchez Alvarado, A. Not your father's planarian: a classic model enters the era of functional genomics. *Nat. Rev. Genet.* **3**, 210–219 (2002).
- Guo, L., Zhang, S., Rubinstein, B., Ross, E. & Alvarado, A. S. Widespread maintenance of genome heterozygosity in *Schmidtea mediterranea*. *Nat. Ecol. Evol.* **1**, 19 (2016).
- Zayas, R. M. et al. The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: analysis of ESTs from the hermaphroditic strain. *Proc. Natl Acad. Sci. USA* **102**, 18491–18496 (2005).
- Grohme, M. A. et al. The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature* **554**, 56–61 (2018).
- Brandl, H. et al. PlanMine—a mineable resource of planarian biology and biodiversity. *Nucleic Acids Res.* **44**, D764–D773 (2016).
- Robb, S. M., Ross, E. & Sanchez Alvarado, A. SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Res.* **36**, D599–D606 (2008).
- Weise, A. et al. High-throughput sequencing of microdissected chromosomal regions. *Eur. J. Hum. Genet.* **18**, 457–462 (2010).
- Ma, L. et al. Direct determination of molecular haplotypes by chromosome microdissection. *Nat. Methods* **7**, 299–301 (2010).
- Bayona-Vasquez, N. J. et al. Adapterama III: quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). *PeerJ* **7**, e7724 (2019).
- Dobzhansky, T. & Epling, C. The suppression of crossing over in inversion heterozygotes of *Drosophila pseudoobscura*. *Proc. Natl Acad. Sci. USA* **34**, 137–141 (1948).
- Miller, D. E. et al. The molecular and genetic characterization of second chromosome balancers in *Drosophila melanogaster*. *G3* **8**, 1161–1171 (2018).
- Sun, Y., Svedberg, J., Hiltunen, M., Corcoran, P. & Johannesson, H. Large-scale suppression of recombination predates genomic rearrangements in *Neurospora tetrasperma*. *Nat. Commun.* **8**, 1140 (2017).
- Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
- Chong, T., Collins, J. J. III, Brubacher, J. L., Zarkower, D. & Newmark, P. A. A sex-specific transcription factor controls male identity in a simultaneous hermaphrodite. *Nat. Commun.* **4**, 1814 (2013).
- Hyman, L. H. in *The Invertebrates*, vol. II (ed. Boell, E. J.) 52–458 (McGraw-Hill Book Company, 1951).
- Bertin, B. et al. *Gelsolin* and *dCryAB* act downstream of muscle identity genes and contribute to preventing muscle splitting and branching in *Drosophila*. *Sci. Rep.* **11**, 13197 (2021).
- Buddenborg, S. et al. Assembled chromosomes of the blood fluke *Schistosoma mansoni* provide insight into the evolution of its ZX sex-determination system. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.08.13.456314> (2021).
- Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997 (1978).
- Charlesworth, D. & David, S. in *Sex Determination in Plants*, 1st ed. (ed. Ainsworth, C. C.) 25–50 (Garland Science, 1999).
- Guo, L. et al. An adaptable chromosome preparation methodology for use in invertebrate research organisms. *BMC Biol.* **16**, 25 (2018).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

## Methods

### Planarian husbandry, RNAi and phenotyping

Sexual planarians were fed organic beef liver once a week. All animals used for experiments were selected randomly. For RNAi, hatchlings around 2 weeks old from strain S2F10 were used. RNAi food was prepared by mixing 1 µg of double-stranded RNA with 10 µl of liver paste<sup>45</sup>. To examine the RNAi phenotypes, photographs were taken with a stereomicroscope (Zeiss) when the animals were freely swimming. The lengths of the worms or bursa canal were measured with ImageJ software by multiple researchers in a double-blind experimental design. The ratio of the bursa canal to the whole worm was used to mitigate potential variations in worm size and their degree of relaxation caused by their soft bodies.

### Chromosome sequencing

Chromosomes were collected from multiple animals of one clonal line, S2, maintained in the laboratory by amputation and regeneration. Chromosome spreads were prepared on nuclease-free membrane slides (Zeiss) according to a previously developed protocol except that, at the last step, the tissues were dissociated into single nuclei and placed onto the slides without squashing with a coverslip<sup>42</sup>. Single chromosomes were identified under a ×40 lens and were collected into the caps of single PCR tubes by PALM MicroBeam laser microdissection (Zeiss). The collected chromosomes were spun down with a tabletop centrifuge in 4 µl PBS, and the DNA in the pellets was amplified with a REPLI-g Single Cell Kit (Qiagen) for sequencing on a MiSeq or HiSeq 3000 sequencing system (Illumina).

### Chromosome-scale genome assembly

A Hi-C sequencing library was prepared from multiple animals of the S2 strain using the enzyme DpnII. Sequenced reads were aligned to *dd\_Smes\_g4.fasta*<sup>26</sup> with *bwa mem* (version 0.7.17)<sup>43</sup>. An assembly file was prepared from the SALSA<sup>3</sup> output *FINAL.fasta* with *juicebox\_scripts* (Phase Genomics). The *.hic* file was prepared by *run-assembly-visualizer.sh* from three-dimensional de novo assembly<sup>44</sup>. The two files were loaded into *Juicebox*<sup>4</sup> for scaffold manipulation using *split*, *merge*, *order* and *orient* commands and for chromosome assembly. The modified assembly file (chromosome-scale) was converted to *fasta* (*Smed\_chr\_ref\_v1*) with *juicebox\_assembly\_converter.py*.

### Hi-C library construction

A Hi-C library was generated using Phase Genomics Proximo Animal Kit version 3.0. Approximately four worms were finely chopped and were then cross-linked for 20 min at room temperature with end-over-end mixing in 1 ml of Proximo cross-linking solution. The cross-linking reaction was terminated with quenching solution for 15 min at room temperature, again with end-over-end mixing. The quenched tissue was rinsed once with 1× chromatin rinse buffer (CRB), transferred to a liquid nitrogen-cooled mortar and ground to a fine powder. The powder was resuspended in 700 µl Proximo lysis buffer 1 and lysed with glass beads for 20 min at room temperature on a vortex mixer. A low-speed spin was used to clear the large debris, and the chromatin-containing supernatant was transferred to a new tube. After a second spin at higher speed, the supernatant was removed, and the pellet containing the nuclear fraction of the lysate was washed with 1× CRB. After removal of the 1× CRB wash, the pellet was resuspended in 100 µl Proximo lysis buffer 2 and incubated at 65 °C for 15 min. Chromatin was bound to the recovery beads for 10 min at room temperature. The beads were placed on a magnetic stand and washed with 200 µl of 1× CRB.

The chromatin bound on the beads was resuspended in 150 µl Proximo fragmentation buffer, and 2.5 µl of Proximo fragmentation enzyme was added. The reaction was incubated for 1 h at 37 °C, cooled to 12 °C and then incubated with 2.5 µl of finishing enzyme for 30 min. After the addition of 6 µl of Stop Solution, the beads were washed with 1× CRB and

were resuspended in 100 µl of Proximo ligation buffer supplemented with 5 µl of proximity ligation enzyme. The proximity ligation reaction was incubated at room temperature for 4 h with end-over-end mixing. To this, 5 µl of reverse cross-linking enzyme was added, and the reaction was incubated at 65 °C for 1 h.

After reversing the cross-links, the free DNA was purified with recovery beads, and the Hi-C junctions were bound to streptavidin beads and washed to remove unbound DNA. The washed beads were used to prepare paired-end deep sequencing libraries using Proximo library preparation reagents.

### Oocyte and sperm cell sequencing

Sperm cells were released from sexually mature S2 strain animals into calcium- and magnesium-free buffer (1% BSA). The cell dissociation solution was placed onto a slide and examined under a phase-contrast microscope to identify single sperm cells. Oocytes were released from egg capsules. Single sperm cells or oocytes were transferred into single PCR tubes for amplification with a REPLI-g Single Cell Kit (Qiagen). RAD-seq libraries and whole-genome libraries were prepared and sequenced on a HiSeq 3000 or NovaSeq S2 sequencing system (Illumina). RADseq sequencing data were analysed as described in the linkage map section. Whole-genome sequencing data were analysed as described in the recombination section.

### Synteny analysis

For candidate genes, the protein sequences of genes of interest were obtained from *Schmidtea* specimens. Such genes were aligned to *S. mansoni* genomes<sup>39</sup> using the Protein to Nucleotide BLAST (*tblastn*) tool. To systematically examine the synteny of the whole genome, we used *SonicParanoid* (version 1.3.8)<sup>45</sup> to identify one-to-one protein orthologues between *S. mediterranea* and *S. mansoni*. The *S. mansoni* genome assembly (V9) and its protein annotations are available at <https://zenodo.org/record/5149023#.Ybk9jn3MK3I>.

### Linkage map

AJ/J line, D5, was crossed to a J/V line, S2F8b, to build an F<sub>2</sub> population of 93 animals. Genomic DNA was extracted from a fragment of each animal using an Easy-DNA gDNA Purification Kit (K180001, ThermoFisher). Sequencing libraries for RADseq were prepared according to the procedures of Adapterama III<sup>31</sup> with a few modifications. Genetic variants were identified using *Stacks* (version 2.41)<sup>46,47</sup>. All variants were filtered with *VCFtools* (version 0.1.14)<sup>48</sup> to remove insertions and deletions and to select biallelic SNVs. Clusters of markers located within 200 bp were removed because they were likely to correspond to repetitive elements. Markers disobeying Mendelian segregation were also removed. Only markers that were homozygous in both parents were used. The linkage map was built with *R/QTL*<sup>49</sup>.

### Quantifying recombination

Sequencing reads from S2 and its oocytes and sperm cells were aligned to *Smed\_chr\_ref\_v1* with *bwa mem* (version 0.7.17). Genetic variants were jointly called by using the Genome Analysis Toolkit (GATK, version 4.1.4.1) with *GenomicsDB* and *GenotypeGVCFs*<sup>50</sup>. Biallelic heterozygous markers in S2 were further filtered by removing abnormal markers, including those with no segregation of two alleles in the gametes, clusters of markers in close proximity (<200 bp apart) and markers of heterozygosity in sperm cells. The J and V haplotypes were manually phased for oocytes without recombination and were phased by *MPR.genotyping*<sup>51</sup> for all gametes. The *MPR.genotyping* package was also used to impute or correct missing or erroneous genotypes. The final genotype bins were used to identify and visualize recombination with customized R code. Quantification of recombination was focused on crossovers between long tracks of haplotypes along a chromosome. Putative gene conversion events such as short tracks of haplotype switches encompassing <1% of the chromosome length were not included.

## Gene expression analysis

To examine the gene content on chromosome 1, transcriptome data were downloaded from the NCBI Sequence Read Archive (SRA). To examine genes related to sexual development, transcriptomes from sexual adults<sup>9,12,52</sup>, sexual juveniles<sup>9</sup> and sexual adults with *nhr-1* RNAi<sup>9</sup> were used. To examine stem cell-enriched genes, transcriptomes from sorted X1 cells and CIW4 were used<sup>53,54</sup>.

All sequencing data were aligned to *dd\_Smed\_v6* (ref. <sup>27</sup>) with *bwa mem* (version 0.7.17). Differential gene expression was analysed with DESeq2 (version 1.26.0)<sup>55</sup>. Expression was quantified at the transcript level with kallisto (version 0.44.0)<sup>56</sup> and was imported and summarized to gene-level count matrices by tximport<sup>57</sup>.

## Haplotype-specific expression

To examine the haplotype-specific expression of critical regulators of the reproductive system, mRNA was extracted from six sexually mature animals of the S2J/V line and was analysed as three biological replicates, with two animals pooled into each replicate. Nine penis papillae were dissected from nine sexual adult animals of the same line and were analysed as three biological replicates, with penis papillae from three animals pooled into each replicate. All mRNA samples were extracted on the same day and were processed at the same time for library preparation and sequencing to minimize technical variation. Libraries for RNA-seq were prepared with a Clontech SMARTer Stranded Total RNA-seq (Pico) Kit. The workflow consisted of converting total RNA to cDNA and then adding adaptors for Illumina sequencing through PCR. The PCR products were purified, and ribosomal DNA was depleted. The cDNA fragments were further amplified with primers universal to all libraries. Lastly, the PCR products were purified again to yield the final cDNA library. Different adaptors were used to multiplex samples in one lane. Sequencing was performed on an Illumina NovaSeq 6000 with a 150-bp paired-end-read run. Data quality checks were conducted using the Illumina Sequencing Analysis Viewer. Demultiplexing was performed using Illumina Bcl2fastq2 version 2.17.

All sequencing data were aligned to *dd\_Smed\_v6* (ref. <sup>27</sup>) with *bwa mem* (version 0.7.17). To ensure accuracy, haplotype-specific expression of the 14 genes of interest was manually examined using the Integrative Genomics Viewer. J or V allele counts were identified for each biallelic variant in the exons. For a particular gene, the allele counts were aggregated from all variants on all exons for the J and V haplotypes. The allele counts were then subjected to binomial testing and Bonferroni correction<sup>58</sup> to determine whether the observed allele bias was statistically significant.  $P > 1$  after Bonferroni correction was set to 1.

## De novo map, parameter optimization and phylogenetic inference

All datasets were run through the de novo pipeline as implemented in Stacks version 2.52 (refs. <sup>46,47</sup>). First, paired-end reads were demultiplexed and filtered for quality using PROCESS\_RADTAGS and were truncated to a length of 135 bp. Individual reads with Phred scores below 30 or uncalled bases were discarded (96.9% of reads passed the quality filters). Optimal parameters were identified following the guidelines of ref. <sup>59</sup> by running multiple iterations of the de novo pipeline and varying just one parameter with each new iteration on a subset of 13 samples from the same population, [Sic\_mar], following recommendations<sup>60</sup>. We varied the minimum stack depth (-m) between 1 and 6 (m1–m6), the number of mismatches allowed between stacks (-M) between 1 and 10 (M1–M10), and the number of mismatches allowed to merge catalogue loci (-n) while keeping all other parameters constant (m3, M2 and n0). We then compared the number of polymorphic assembled loci across samples using a sample representation cut-off of 80% (r80) and gain or loss of polymorphic loci with each new iteration. Once -m and -M were optimized, we assessed -n by evaluating the change in the number of polymorphic loci for  $n = M - 1$ ,  $n = M$  and  $n = M + 1$ . RAD loci

were then assembled using the *denovo\_map.pl* wrapper in Stacks, and the following parameters were set:  $m = 3$ ,  $M = 2$  and  $n = 3$ .

Assembled loci that were present in 75% of all individuals were kept from POPULATIONS, a minor allele frequency (MAF) filter of 0.04 (--min-maf) was used to filter out singleton SNPs that could mask population structure and a maximum observed heterozygosity (--max-obs-het) filter of 0.99 was used to remove potentially paralogous loci<sup>61</sup>. Additionally, to build the phylogenetic tree, we concatenated all RADseq loci after filtering (--phylip-var-all). Phylogenetic trees were built by maximum likelihood using RAXML-NG version 0.9.0 (ref. <sup>62</sup>), starting from a random seed and applying a GTR+G substitution model and 1,000 bootstrap replicates. The sample from Menorca (Sme7-5\_men) was used as an outgroup to root the tree.

## Population structure analysis

For this analysis, we excluded the outgroup sequence and ran POPULATIONS to retain loci present in all populations (-p 10) and 75% of the individuals present in each population (-r 0.75). On the basis of the loci that passed our filtering criteria, a random whitelist of 1,000 loci was generated and again run through POPULATIONS with the same criteria but retaining the first SNP at each locus (--write-single-snp). The output was exported in Structure format, and Structure version 2.3.4 (ref. <sup>63</sup>) was used to infer population structure with 10,000 chains as burn-in and 100,000 MCMC chains with 20 iterations for  $K = [1-11]$ . The resulting files were run through Structure Harvester<sup>64</sup>, and the optimal  $K$  was determined<sup>65</sup>.

## DNA FISH

For telomere FISH on oocyte chromosomes, ovaries were dissected as reported previously<sup>66,67</sup>. Hybridization was carried out as reported with chromosome spreads<sup>42</sup> except that the ovaries were kept in suspension in washing buffer or hybridization buffer. Before hybridization, the dissected ovaries were treated with a digestion buffer (0.1% SDS and 10  $\mu\text{g ml}^{-1}$  proteinase K (Qiagen) in 0.3% Triton X-100 in PBS) for 10 min at room temperature. The repeat located near centromeres in Fig. 2c had the following sequence: TCTGGACGGAATTTTTAATCTTTATAGGCTTGATCTCTGTC AATTTTTATTTGTTTTCATAATCTTTGATATATTTCTCGATAACTTTTGATTCTCTACATGATAGCATTTTAAAAATTGCAAAAATCATAACGGGCTCGTCAACACAAGTCAT.

## Hybridization chain reaction and RNA FISH

To examine tissue expression of the genes *gfp* and *smedwi-1*, we used probes and buffers for third-generation ISH chain reaction purchased from Molecular Instruments. Sexually mature planarians were treated with 7.5% *N*-acetyl-L-cysteine (Sigma-Aldrich) for 10 min and were then fixed in 4% paraformaldehyde (Electron Microscopy Sciences; 16% solution diluted 1:4 in PBS) for 20 min. The copulatory apparatus was dissected into a 1.5-ml RNase-free tube. The rest of the procedures followed the hybridization chain reaction RNA FISH protocol of Molecular Instruments for whole-mount mouse embryos.

## Reporting summary

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

## Data availability

The authors confirm that all data underlying the findings are fully available without restriction. All sequencing data are available from the NCBI SRA database (accession number PRJNA731187). The chromosome-scale genome assemblies for sexual *S. mediterranea* (including phased genomes and genomes with repetitive elements) are openly available on the Planosphere (<https://planosphere.stowers.org/>), PlanMine (<https://planmine.mpibpc.mpg.de/planmine/begin.do>), GenBank (GCA\_022537955.1) and Zenodo (<https://doi.org/10.5281/>



zenodo.5807415) databases. We used publicly available NCBI PacBio sequencing data (accession numbers SRX2700681–SRX2700684) and planarian transcriptome data (accession numbers SRR2658118–SRR2658125, SRR2658134–SRR2658141, SRR3473955–SRR3473957, SRR3629945–SRR3629952, SRR6351185–SRR6351188, SRR6351201–SRR6351204, SRR6351213–SRR6351216, SRR6363910–SRR6363927 and SRR6364586–SRR6364588).

## Code availability

Custom code for data analysis can be openly accessed on the Zenodo database at <https://doi.org/10.5281/zenodo.5807415>.

43. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
44. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
45. Cosentino, S. & Iwasaki, W. SonicParanoid: fast, accurate and easy orthology inference. *Bioinformatics* **35**, 149–151 (2019).
46. Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* **22**, 3124–3140 (2013).
47. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: building and genotyping loci de novo from short-read sequences. *G3* **1**, 171–182 (2011).
48. Danecek, P. et al. The variant call format and VCF tools. *Bioinformatics* **27**, 2156–2158 (2011).
49. Broman, K. W., Wu, H., Sen, S. & Churchill, G. A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
50. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
51. Xie, W. et al. Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. *Proc. Natl Acad. Sci. USA* **107**, 10578–10583 (2010).
52. Davies, E. L. et al. Embryonic origin of adult stem cells required for tissue homeostasis and regeneration. *eLife* **6**, e21052 (2017).
53. Zeng, A. et al. Prospectively isolated tetraspanin<sup>+</sup> neoblasts are adult pluripotent stem cells underlying planaria regeneration. *Cell* **173**, 1593–1608 (2018).
54. Duncan, E. M., Chitsazan, A. D., Seidel, C. W. & Sanchez Alvarado, A. Set1 and MLL1/2 target distinct sets of functionally different genomic loci in vivo. *Cell Rep.* **13**, 2741–2755 (2015).
55. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
56. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
57. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
58. Etymologia: Bonferroni correction. *Emerg. Infect. Dis.* **21**, 289 (2015).
59. Paris, J. R., Stevens, J. R. & Catchen, J. M. Lost in parameter space: a road map for STACKS. *Methods Ecol. Evol.* **8**, 1360–1373 (2017).
60. Rochette, N. C. & Catchen, J. M. Deriving genotypes from RAD-seq short-read data using Stacks. *Nat. Protoc.* **12**, 2640–2659 (2017).
61. Stobie, C. S., Oosthuizen, C. J., Cunningham, M. J. & Bloomer, P. Exploring the phylogeography of a hexaploid freshwater fish by RAD sequencing. *Ecol. Evol.* **8**, 2326–2342 (2018).
62. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
63. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
64. Earl, D. A. & vonHoldt, B. M. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).
65. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software Structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
66. Guo, L. et al. Subcellular analyses of planarian meiosis implicates a novel, double-membraned vesiculation process in nuclear envelope breakdown. Preprint at *bioRxiv* <https://doi.org/10.1101/620609> (2019).
67. Guo, F. et al. Planarian ovary dissection for ultrastructural analysis and antibody staining. *J. Vis. Exp.* **175**, e62713 (2021).
68. Ouellette, L. A., Reid, R. W., Blanchard, S. G. & Brouwer, C. R. LinkageMapView—rendering high-resolution linkage and QTL maps. *Bioinformatics* **34**, 306–307 (2018).

**Acknowledgements** We thank L. Rouhana, G. Bruni and S. Zdravljic for discussions; Z. Kashif, E. Belfer and B. Pakfar for planarian husbandry; and X. Li and the UCLA Technology Center for Genomics & Bioinformatics for next-generation sequencing. We thank Y. Zeng, E. Noskova and D. Pasha for discussions on demographic history. We acknowledge Life Science Editors for professional editing services. Confocal laser-scanning microscopy was performed at the Advanced Light Microscopy/Spectroscopy Laboratory and the Leica Microsystems Center of Excellence at the California Nano Systems Institute at UCLA with funding from NIH Shared Instrumentation grant S10OD025017 and NSF Major Research Instrumentation grant CHE-0722519. This work was supported by grants from the Howard Hughes Medical Institute (L.K.) and the Helen Hay Whitney Foundation (L.G.).

**Author contributions** L.K. and L.G. conceptualized the study. L.G. coordinated the project, conducted experiments and wrote the original draft. L.G., E.B.-D. and T.V.L. developed methods for single-sperm sequencing. L.G., K.K., K.H., E.W., C.C. and Y.W. collected the data. L.G., J.S.B., D.D.-S., M.R., J.B. and D.L. analysed the data. K.K., K.H., E.W., C.C. and Y.W. maintained planarian colonies. L.K., L.G., O.T.S. and A.S.A. edited the manuscript. A.S.A. provided the D5 and S2 strains, and sequencing data of a dozen oocytes. M.R. provided wild isolate DNA. L.K. acquired funding and supervised the project.

**Competing interests** The authors declare no competing interests.

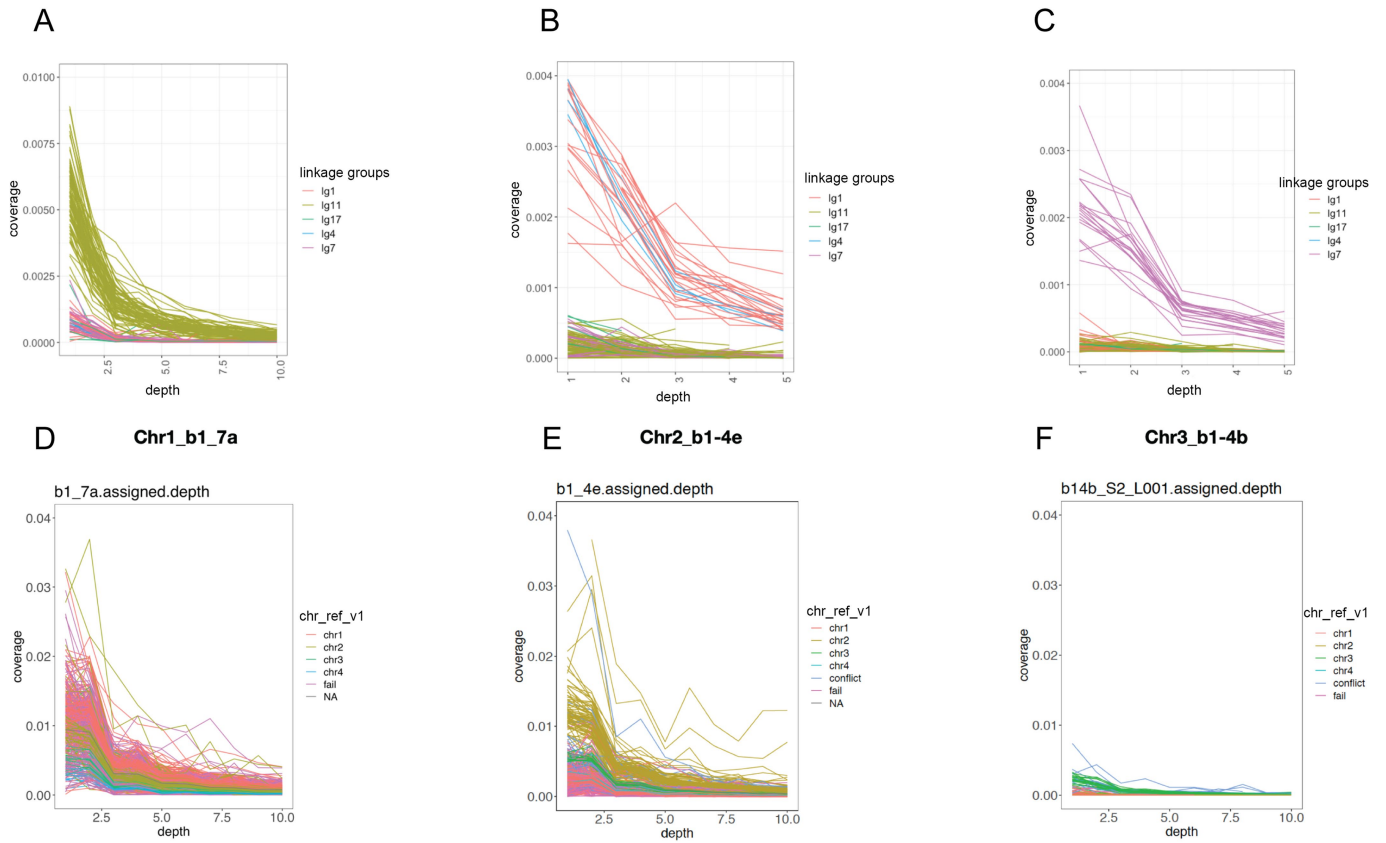
### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-022-04757-3>.

**Correspondence and requests for materials** should be addressed to Longhua Guo or Leonid Kruglyak.

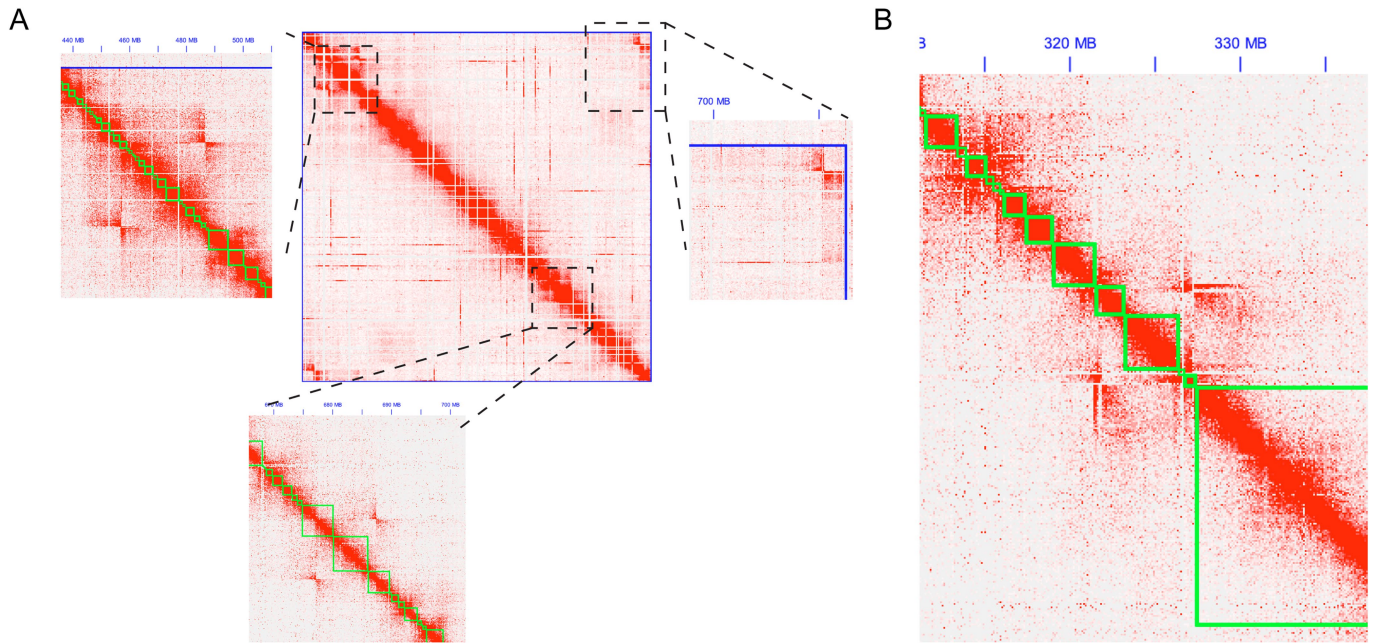
**Peer review information** *Nature* thanks Bo Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



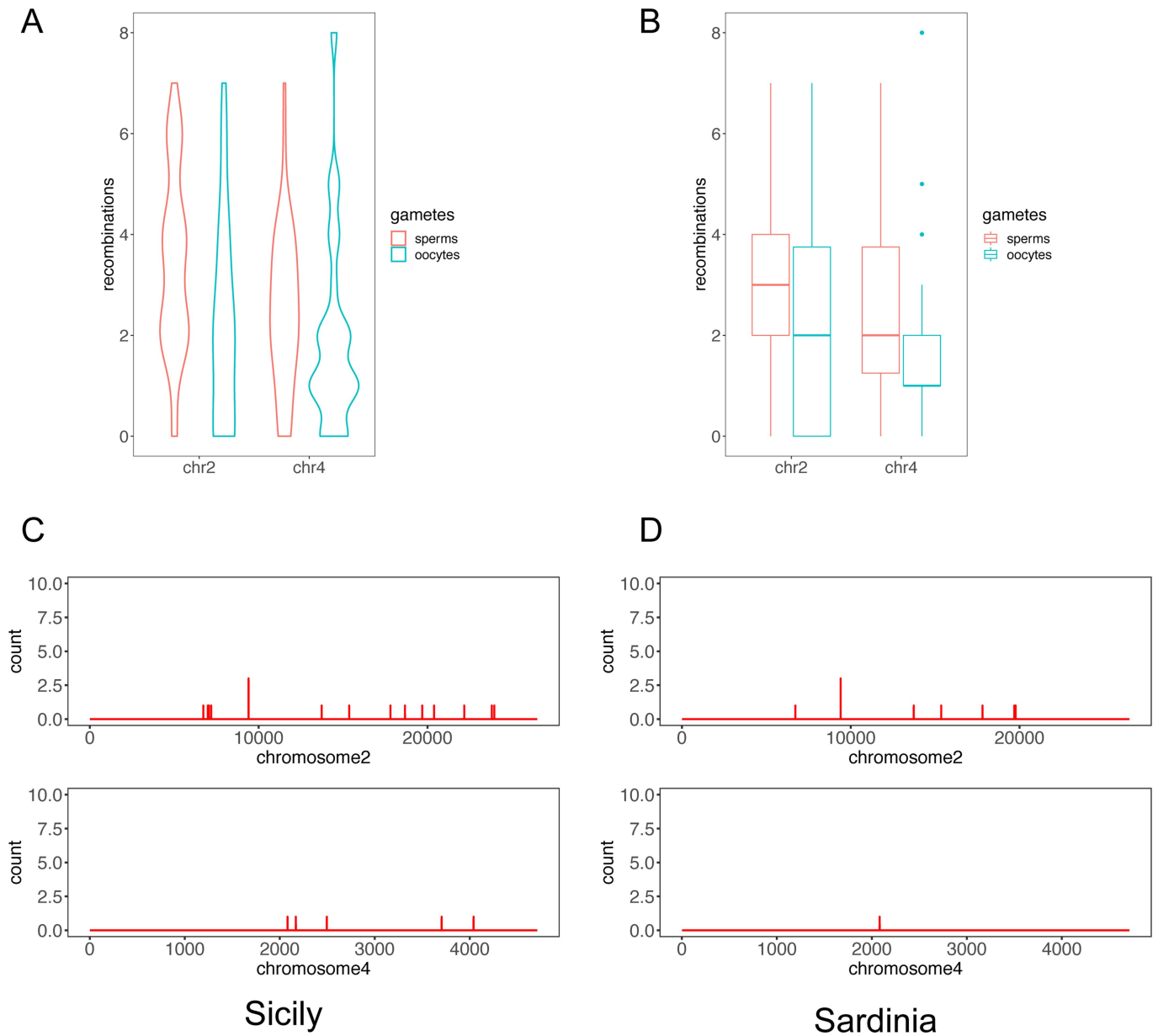
**Extended Data Fig. 1 | Chromosome assignment by ChrSeq.** (A-C) Chromosome sequencing depth of different scaffolds is consistent with linkage group assignment. Linkage groups: lg11 = chr1, lg1+lg4 = chr2, lg7 = chr3, lg17 = chr4. (D-F) Chromosome sequencing depth of different

scaffolds is consistent with chromosome assignment in the genome assembly *smed\_chr\_ref\_v1*. Chromosome sequencing samples: chromosome 1 (A, D); chromosome 2 (B, E); chromosome 3 (C, F).



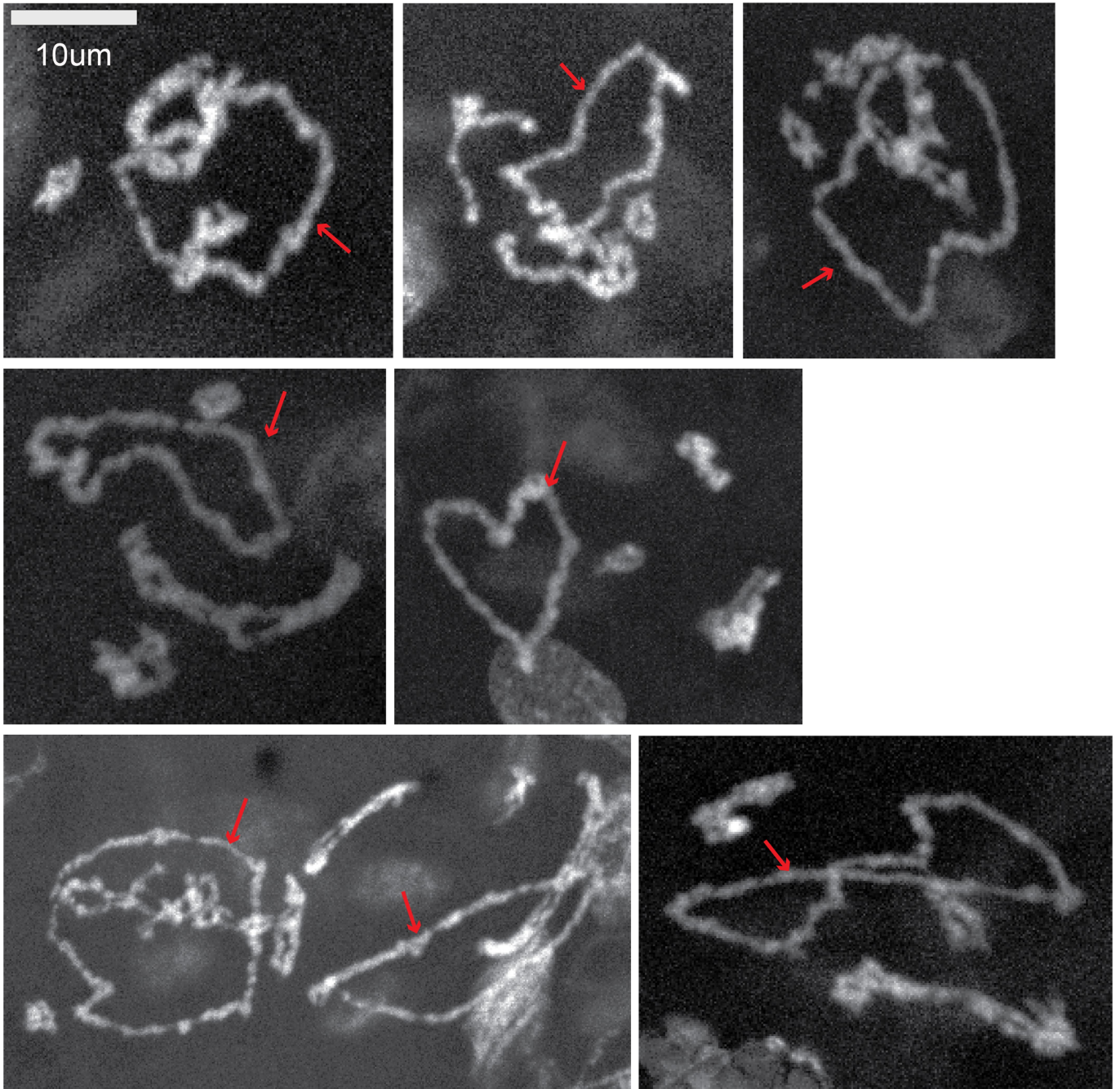
**Extended Data Fig. 2 | Putative inversions in the planarian genome detected by chromatin contact heatmap.** (A) chromosome 1 chromatin contact heatmap showing 3 potential inversions (dashed-line rectangles)

visualized in Juicebox. (B) A complex structural variant on chromosome 2. Green outline: dd\_Smes\_g4 scaffolds. Blue outline: chromosome boundaries.



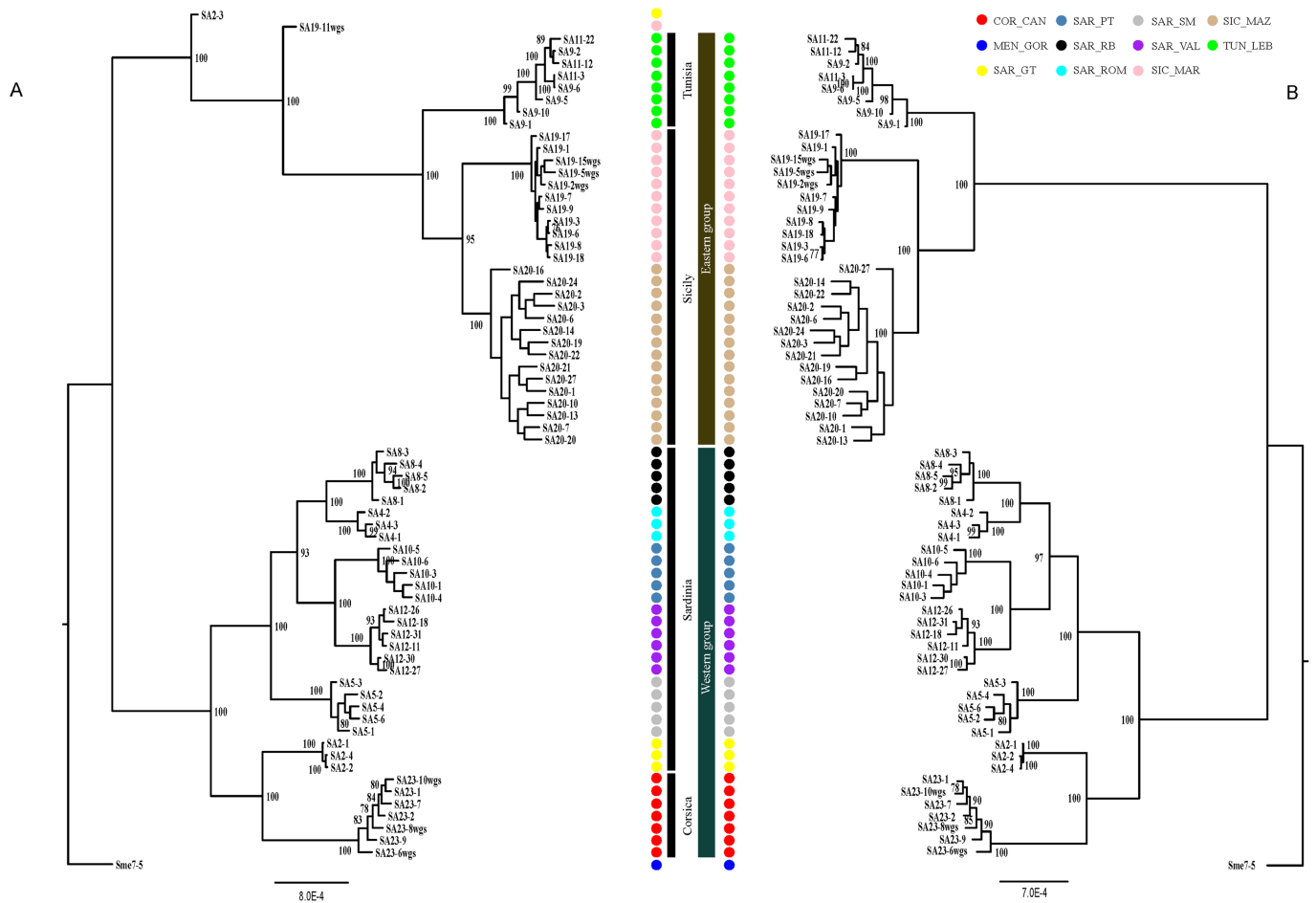
**Extended Data Fig. 3 | Recombination in gametes and heterozygosity distribution in the wild on chromosomes 2 and 4.** (A-B) Number of recombinations in oocytes and sperm on chromosomes 2 and 4 summarized as a violin plot (A) and a box plot (B). The median number of recombinations on chromosome 2 is 3 for sperm and 2 for oocytes, derived from a total of 49 sperm and 38 oocytes. The median number of recombinations on chromosome 4 is 2 for sperm and 1 for oocytes, derived from a total of 50 sperm and 26 oocytes.

Chr2 sperm: min = 0, max = 7, median = 3, first quartile = 2, third quartile = 4.  
 Chr2 oocytes: min = 0, max = 7, median = 2, first quartile = 0, third quartile = 4;  
 Chr4 sperm: min = 0, max = 7, median = 2, first quartile = 1, third quartile = 4;  
 Chr4 oocytes: min = 0, max = 3, median = 1, first quartile = 1, third quartile = 2.  
 (C-D) Very few sites are consistently heterozygous in more than 80% of the wild isolates on chromosomes 2 and 4 in Sicily (C) and in Sardinia (D).



**Extended Data Fig. 4 | Representative images of late prophase I oocytes with ring-shaped chromosome 1.** More than 100 oocytes from > 40 ovaries and > 20 sexual strains were examined for the presence of a ring-shaped chromosome 1. The experiment was repeated >10 times independently with

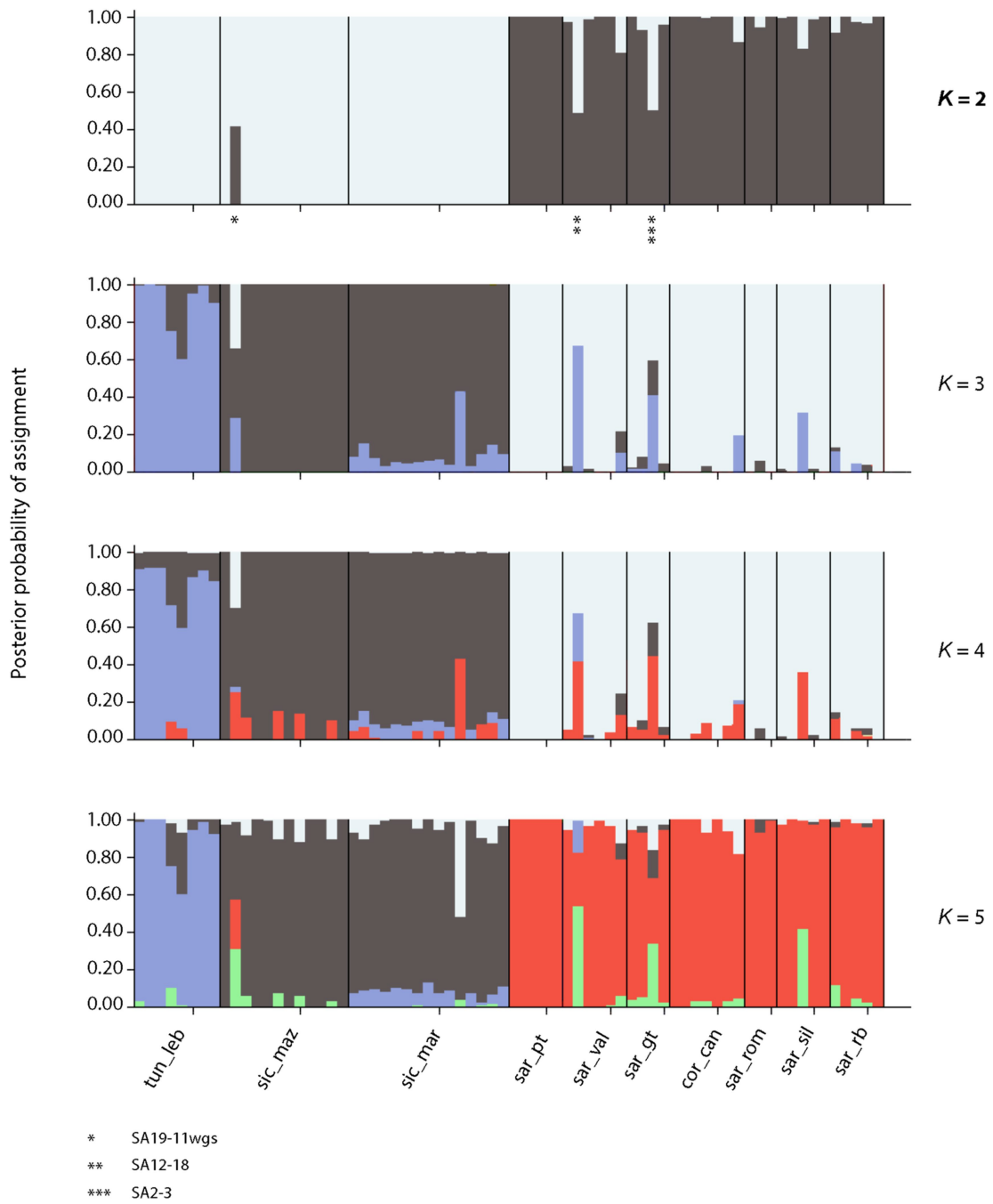
similar results. Chromosome 1 always forms a ring in J/V strains in late prophase I oocytes. Shown in the figure are chromosomes from 8 different oocytes at late prophase I. Arrows point to the ring-shaped chromosome 1.



**Extended Data Fig. 5 | Phylogenetic tree of wild isolates of *Schmidtea mediterranea*.** The tree is rooted with one asexual animal from Menorca (Sme7-5). The tree on the left (A) includes two animals from SAR and SIC with

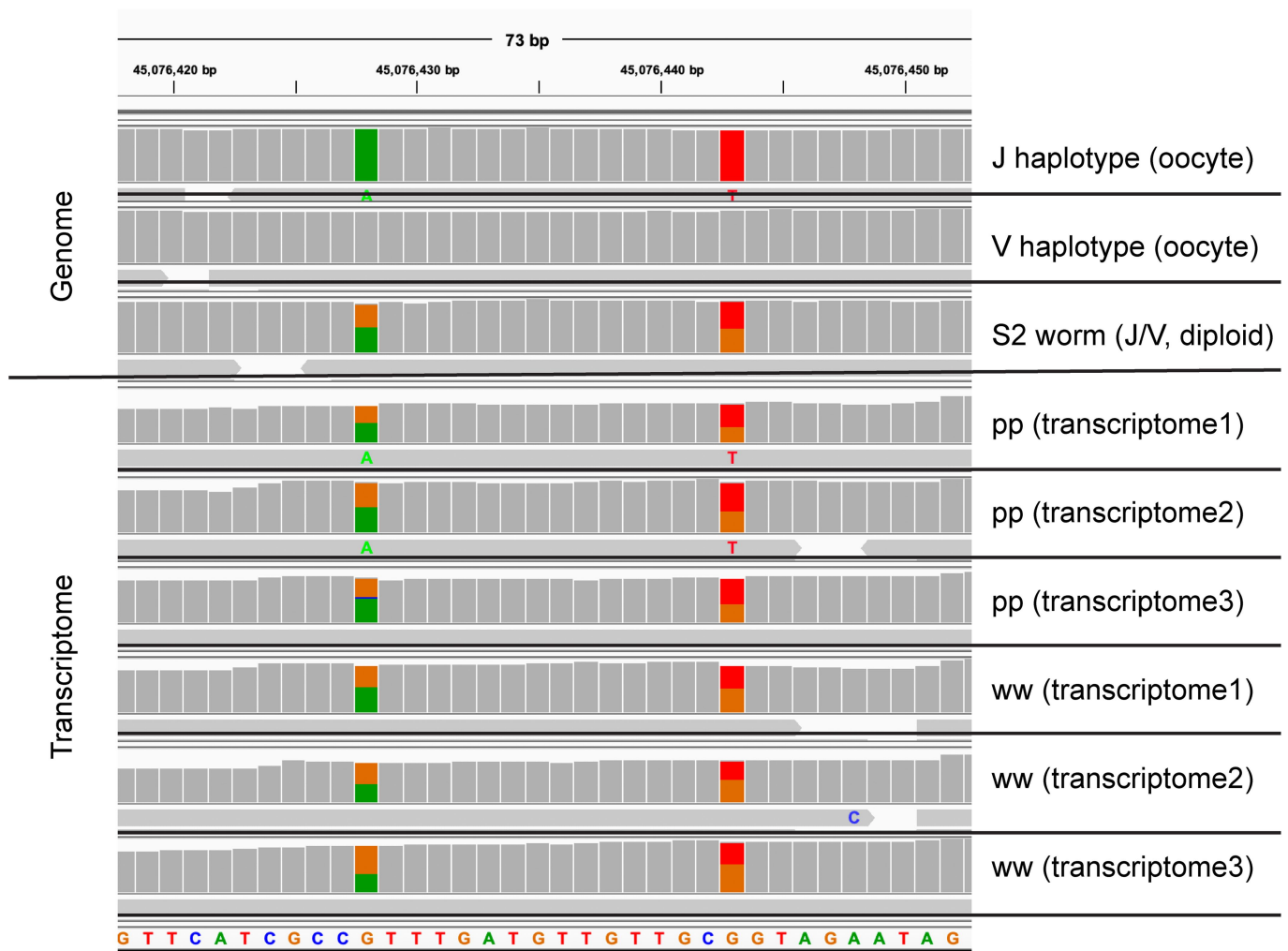
higher genome heterozygosity than the cohorts from the same collection sites. Phylogenetic relations agree with the previous designation of eastern and western populations<sup>1</sup>. SAR: Sardinia; COR: Corsica; SIC: Sicily; TUN: Tunisia.





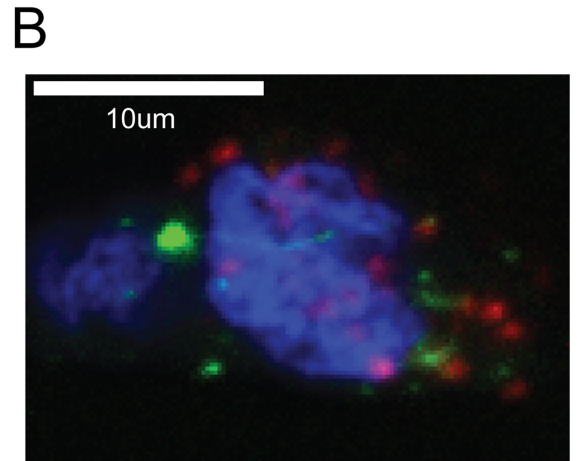
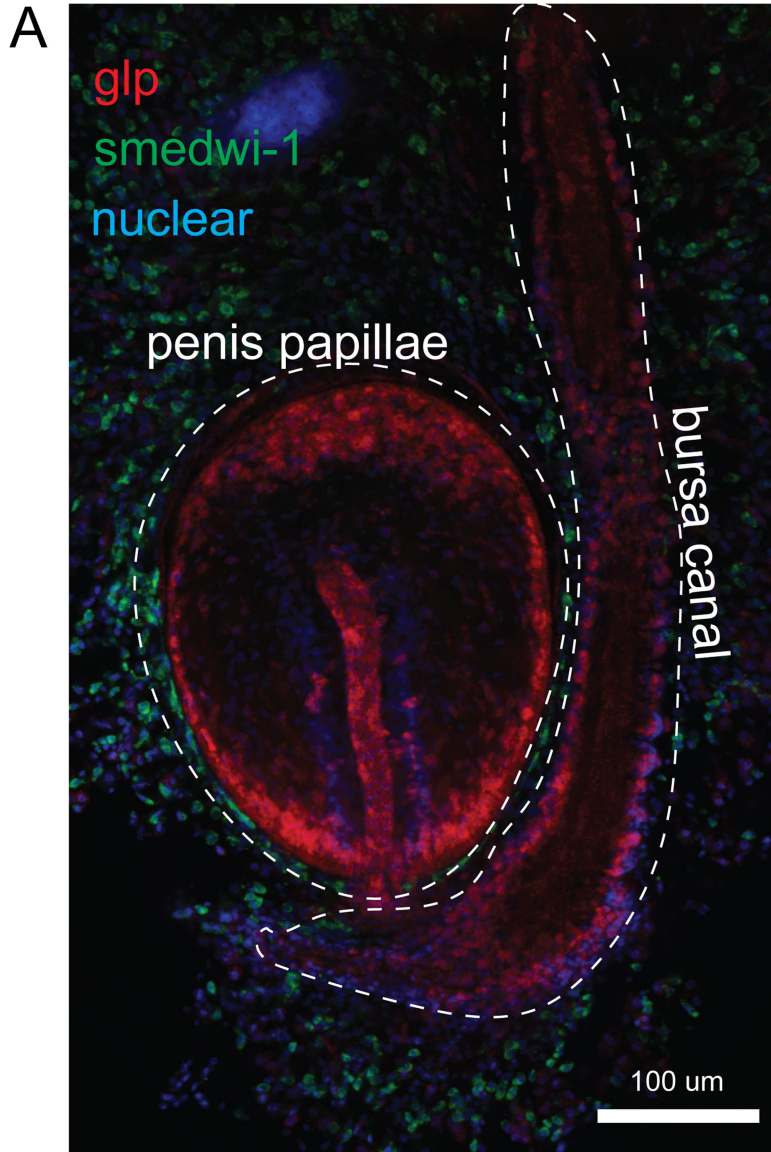
**Extended Data Fig. 7 | STRUCTURE analysis of wild isolates.** Evanno analysis supports a K = 2 split, dividing the samples geographically into Eastern and Western groups. With K = 3, a Tunisian signal was found to be present in all other 3 groups (light blue).





**Extended Data Fig. 8 | Datasets for the examination of haplotype-specific expression.** The sequencing reads from genomes (J-haplotype oocytes, V-haplotype oocytes, diploid S2 worms) and transcriptomes (pp, ww from the diploid S2 worms) were aligned to Smed\_chr\_ref\_v1. Shown in the figure are reads for the gene CPEB-2. Two variants with J and V alleles are marked by colored bars. In the genome of the adult worm S2 (i.e., DNA), both J and V alleles are present at equal ratios. In the genome of oocytes (J or V haplotype, i.e.,

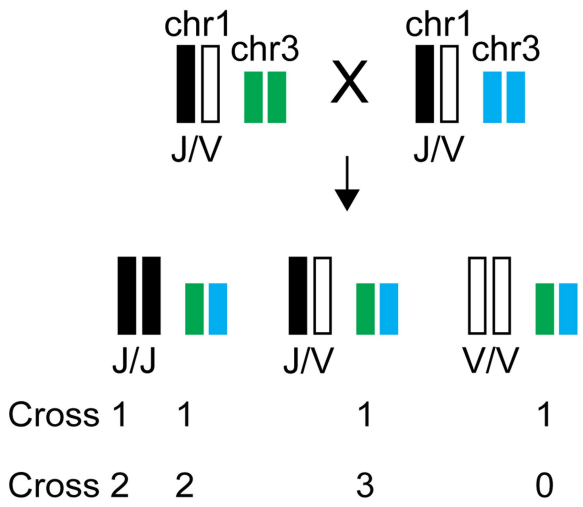
DNA), the variants have only one allele. These observations validate the heterozygous variants. In the transcriptomes (i.e., RNA) of penis papillae (pp) and whole worms (ww) from S2 adults, both J and V alleles are present. In the three transcriptomes of pp, J alleles were more highly expressed. In the three transcriptomes of ww, the J:V allele ratios vary and do not significantly depart from equal expression.



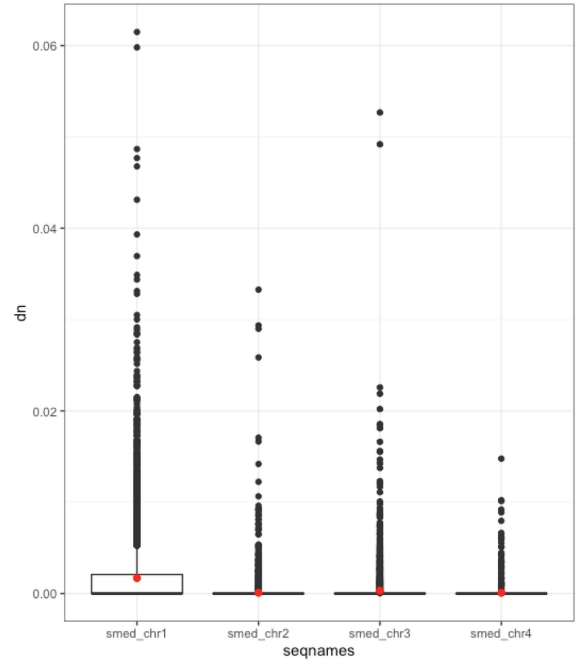
**Extended Data Fig. 9 | Specific expression of *smed-glp* in the male reproductive organ (penis papillae) and female reproductive organ (bursa canal).** Expression of the *glp* gene (in red color) and the stem cell marker *smedwi-1* (in green color) in the dissected copulatory apparatus was visualized by Fluorescence *in situ* hybridization chain reaction. (A) White dashed lines outline penis papillae and bursa canal. Nuclei were stained by DAPI (in blue

color). No *glp* expression in tissues surrounding penis papillae or bursa canal. Less than 1% of *glp*<sup>+</sup> cells express the stem cell marker *smedwi-1*. (B) Cells expressing the two markers do exist in both bursa canal and penis papillae. This experiment was repeated 3 times independently with 6 sexually mature worms, producing similar results.

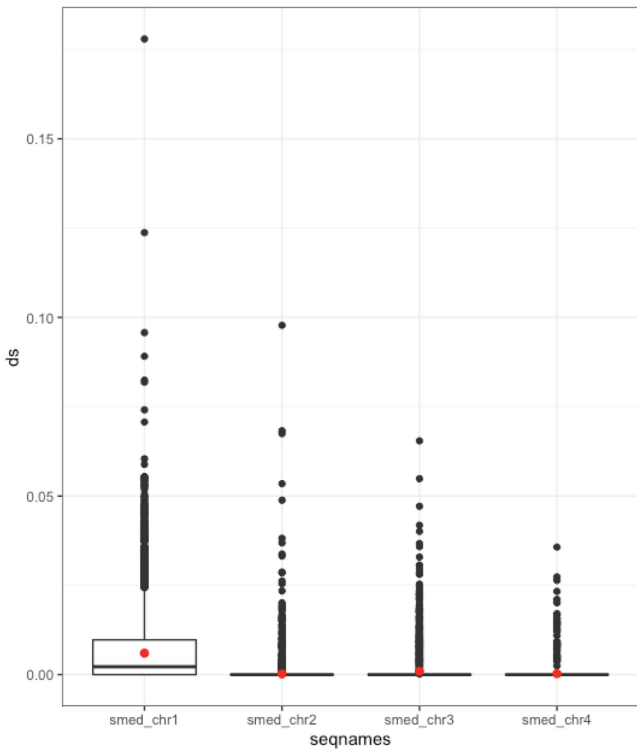
A



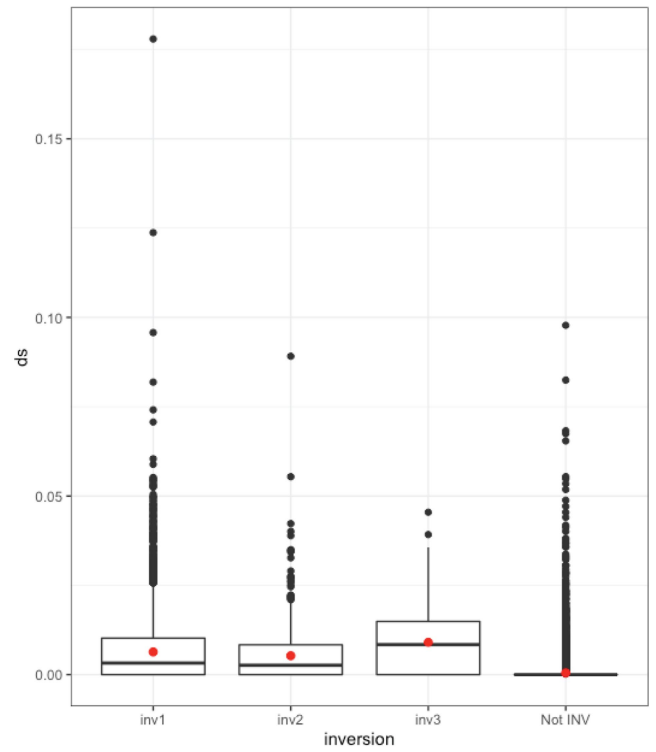
B



C



D



Extended Data Fig.10 | See next page for caption.

# Article

**Extended Data Fig. 10 | Embryonic lethality and synonymous divergence in *Schmidtea mediterranea*.** (A) Two J/V lines with differential SNP markers on chromosome 3 were crossed. Zygotes were collected for genotyping. Heterozygous chromosome 3 markers validate the occurrence of fertilization. All three genotypes of chromosome 1 (J/J, J/V and V/V) were observed in zygotes. (B-C) Non-synonymous divergence (B) and Synonymous divergence (C) for all heterozygous sites in the J/V strain S2 and their distribution in different chromosomes summarized as boxplots. Chromosome 1 has the highest mean (red dots) and median dN or dS values. The min, max, median, mean, first quartile, and third quartile of dN (B) for each chromosome are as following: chr1 = (0, 0.0052, 0, 0.0017, 0, 0.0021), chr2 = (0, 0, 0, 0.000092, 0, 0), chr3 = (0, 0, 0, 0.00031, 0, 0), and chr4 = (0, 0, 0, 0.000088, 0, 0). The min, max, median, mean, first quartile, and third quartile of dS (C) for each chromosome are as following: chr1 = (0, 0.024, 0.0022, 0.006, 0, 0.0098), chr2 = (0, 0, 0, 0.00021, 0, 0), chr3 = (0, 0, 0, 0.00093, 0, 0), and chr4 = (0, 0, 0, 0.00026, 0, 0). (D) Synonymous divergence for all heterozygous sites in the J/V

strain S2, and their distribution in the three chromosome 1 inversions and the rest of the genome, summarized as boxplots. Mean (red dots) and median dS decreases in the following order: inversion 3 (inv3), inversion 1 (inv1), inversion 2 (inv2), and the rest of the genome (Not INV). One way ANOVA and Turkey Honest Significant Differences tests showed that dS were significantly different between all 3 inversions (inv2-inv1: p.adj = 0.005; inv3-inv1: p.adj < 1e-06; inv3-inv2: p.adj < 1e-06). The min, max, median, mean, first quartile, and third quartile of dS for each inversion are as following: inversion 1 = (0, 0.026, 0.0034, 0.0064, 0, 0.01), inversion 2 = (0, 0.021, 0.0026, 0.0053, 0, 0.0083), inversion 3 = (0, 0.036, 0.0084, 0.0091, 0, 0.015), and non-inversion = (0, 0, 0, 0.0005, 0, 0). A total of 8317 genes from chromosome 1, 6568 genes from chromosome 2, 3394 genes from chromosome 3, and 1834 genes from chromosome 4 were plotted for non-synonymous divergence (B) and synonymous divergence (C). A total of 6633 genes from inversion 1, 754 genes from inversion 2, 298 genes from inversion 3, and 12428 genes from the rest of the genome were plotted for synonymous divergence (D).

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- |     |           |
|-----|-----------|
| n/a | Confirmed |
|-----|-----------|
- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
  - A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
  - 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 all covariates tested
  - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
  - A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
  - For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
  - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
  - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
  - Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The authors confirm that all data underlying the findings are fully available without restriction. All sequencing data are available from the NCBI SRA database (accession number PRJNA731187). The chromosome-scale genome assemblies for sexual *Schmidtea mediterranea* (including phased genomes and genomes with repetitive elements) are openly available at: planosphere(planosphere.stowers.org); planmine (planmine.mpibpc.mpg.de/planmine/begin.do); Zenodo (10.5281/zenodo.5807415). Publicly available PacBio sequencing data from NCBI were used (accessions: SRX2700681-SRX2700684). Publicly available planarian transcriptome data from NCBI were used (accessions: SRR2658118-SRR2658125, SRR2658134-SRR2658141, SRR3473955-SRR3473957, SRR3629945-SRR3629952,

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences     Behavioural & social sciences     Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The numbers of chromosomes, sperm, oocytes, wild isolates and animals used in each experiment range from 3 to more than 20. The numbers were chosen to ensure representative samples; the specific choices for each experiment were guided by practical considerations of reproducibility and feasibility.
Data exclusions	Failed sequencing samples were excluded if the DNA amplification failed or the sequencing resulted with little data. No data exclusions were performed on successful samples.
Replication	All experiments were repeated at least three times. All attempts to replicate the experiments were successful.
Randomization	All chromosomes, sperm, oocytes and animals were selected randomly. For experimental and control groups, animals were allocated randomly.
Blinding	The measurements of the length of female bursa canal or worms were double-blinded and repeated at least twice. In other experiments, blinding is not essential. For example, in the analysis of recombination or wild diversity, all gametes, animals, chromosomes were processed together as one dataset. The researchers were essentially blinded to the relevance of a particular variant, chromosome or animal.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	The animals used in the study are the sexual biotype of <i>Schmidtea mediterranea</i> , which are simultaneous hermaphrodites. Sexually mature adults (4 month old to 1 year old) were used.
Wild animals	No live wild animals were used. DNA samples previously extracted from wild animals were sequenced and analyzed.
Field-collected samples	Wild isolates of <i>Schmidtea mediterranea</i> were collected from the field and immediately fixed in 70% ethanol. DNA were extracted from these animals in earlier studies. Such DNA samples were used in the current study.
Ethics oversight	As <i>Schmidtea mediterranea</i> is a free-living invertebrate and harmless to humans/environments, no university ethical approval is needed. We were mindful to give the animals best housing conditions (e.g., constant optimal temperature, clean water, consistent feeding schedules, ample space) and best euthanasia practice (e.g., quick treatment with 10% bleach). We followed guidelines of ARRIVE in our reporting.

Note that full information on the approval of the study protocol must also be provided in the manuscript.