1 **Island-specific evolution of a sex-primed autosome in the planarian *Schmidtea***

2 ***mediterranea***

3

4 Longhua Guo[1,6*], Joshua Bloom[1,6], Daniel Dols Serrate[2], Eyal Ben David[3], Olga T.

5 Schubert[1,6], Kaiya Kazuma[1], Katarina Ho[1], Yubao Wei[4], Daniel Leighton[1,6], James

6 Boocock[1,6], Tzitziki Lemus Vergara[1,6], Marta Riutort[2], Alejandro Sánchez Alvarado[5,6],

7 Leonid Kruglyak[1,6*]

8

9 [1] Department of Human Genetics, Department of Biological Chemistry, and University of

10 California Los Angeles, CA, USA

11 [2] Departament de Genètica, Microbiologia i Estadística, Universitat de Barcelona, and

12 Institut de Recerca de la Biodiversitat (IRBio), Barcelona, Catalonia, Spain

13 [3] Department of Biochemistry and Molecular Biology, The Institute for Medical Research

14 Israel-Canada, The Hebrew University-Hadassah Medical School, Ein Kerem, Jerusalem,

15 Israel

16 [4] Institute of Reproductive Medicine, Henan Provincial People's Hospital, Zhengzhou

17 University, Zhengzhou, China

18 [5] Stowers Institute for Medical Research, and, Kansas City, MO, USA

19 [6] Howard Hughes Institute for Medical Research, Chevy Chase, MD, USA

20

21 * To whom correspondence should be addressed. E-mail: longhuaguo@mednet.ucla.edu

22 (LG), lkruglyak@mednet.ucla.edu (LK).

23

24

## Abstract

26    The sexual biotype of the planarian *Schmidtea mediterranea* is a hermaphrodite

27    indigenous to Tunisia and several Mediterranean islands. Here, we isolated individual

28    chromosomes and used sequencing, Hi-C and linkage mapping to assemble a

29    chromosome-scale genome reference. The linkage map revealed an extremely low rate

30    of recombination on chromosome 1. We confirmed suppression of recombination on

31    chromosome 1 by quantifying recombination events in individual sperm and oocytes. We

32    showed that the extensive heterozygous regions, previously designated as J and V

33    haplotypes, comprise essentially all of chromosome 1. Genome sequencing of individuals

34    isolated in the wild indicated that this heterozygosity has evolved specifically in

35    populations from Sardinia and Corsica. We found that chromosome 1 acquired many

36    genes that determine the development of female and male reproductive systems, along

37    with haplotype-specific expression of such sex specific genes. These molecular

38    signatures and restricted recombination in a hermaphrodite led us to propose that

39    chromosome 1 is a sex-primed autosome, providing direct molecular evidence for the

40    traditional model of how sex chromosomes may have evolved from autosomes.

41

42

43    Sex chromosomes are thought to have evolved from homologous autosomes that

44    acquired sex-determining genes and lost their ability to recombine (*1-10*). As such, sex

45    chromosome evolution and recombination suppression are closely associated(*1-10*). Yet,

46    direct evidence of such homologous autosomes primed for evolution into sex

47    chromosomes is hard to capture, and little is presently known about the molecular

48    signatures associated with the evolution of recombination suppression.

49

50    The freshwater planarian *S. mediterranea* is an important model organism for studies of

51    regeneration (*11, 12*). It exists as asexual and sexual reproductive biotypes. The sexual

52    biotype is mostly distributed in Tunisia, and on the islands of Sardinia, Corsica, and Sicily

53    (*13*). The sexual biotype is a simultaneous hermaphrodite: it develops both male and

54    female reproductive systems in the same adult individual and obligately outcrosses to

55    fertilize other individuals (*14, 15*). We hypothesized that studying chromosome evolution

56    in a simultaneous hermaphrodite might provide insights into the early evolution of a

57    primitive sex chromosome.

58

59    *S. mediterranea* has 4 pairs of chromosomes, which are stably diploid. The genome is

60    reported to comprise 774 Mb assembled as 481 non-contiguous series of genomic

61    sequences or 'scaffolds' (*16-18*). In a previous study, we found that ~300 Mb of the

62    genome remains heterozygous even after extensive inbreeding of laboratory strains, and

63    that this phenomenon also occurs naturally in wild populations in Sardinia (*14*). The two

64    sets of heterozygous alleles were collectively named the J and V haplotypes. To define

65 the chromosomal locations of these alleles and investigate the reasons underpinning the

66 persistence of heterozygosity in *S. mediterranea*, a detailed assembly of all four

67 chromosomes was needed.

68

69 **Chromosome-Scale Genome Assembly**

70 To transform the 481 scaffolds (*16*) into a chromosome-scale genome reference, we

71 carried out chromosome sequencing (ChrSeq) (*19, 20*) of a lab strain, as well as

72 chromatin proximity sequencing by Hi-C (*21, 22*) (**Fig. 1A**). To do so, we dissected

73 individual chromosomes from mitotic cells by using laser capture and amplified and

74 sequenced each chromosome individually (**Fig. 1B**). We examined the sequencing depth

75 of each scaffold across multiple samples of the same chromosome to ensure

76 reproducibility and multiple samples of different chromosomes to ensure specificity (**fig.**

77 **S1**). Overall, we successfully amplified and confidently assigned to one of the four

78 chromosomes 740 Mb of the 774 Mb genome (**table S1**).

79

80 We used ChrSeq information and chromatin interaction data generated by Hi-C to correct

81 and connect the individual scaffolds within a chromosome into a chromosome-scale

82 genome, hereafter named Smed_chr_ref_v1. Hi-C data analyzed by the scaffolding

83 algorithm SALSA (*21*) resolved 481 scaffolds into 57 super-scaffolds and 104 singletons.

84 ChrSeq revealed that 3 of the 57 super-scaffolds were disjointed inter-chromosomal

85 fragments, consistent with the Hi-C contact heatmap (**Fig. 1C**). We split the disjointed

86 super-scaffolds, merged the scaffolds into chromosomes, and ordered and oriented all

87 scaffolds within the chromosomes by using the visualization software Juicebox (*23*)

4

88    (**Methods**). Chromosome assignments by ChrSeq alone (**table S1**) and by Hi-C alone

89    were inconsistent for only 3 of 384 (0.8%) scaffolds (**table S2**). We assigned these 3

90    scaffolds to chromosomes manually based on the Hi-C data. Hi-C also detected 26 inter-

91    chromosomal or intra-chromosomal assembly errors (**table S2**) in the previous assembly

92    (*16*), 5 of which were confirmed by ChrSeq to be inter-chromosomal disjointed scaffolds

93    (**table S2**). The final genome assembly (Smed_chr_ref_v1, **Fig.1C**) has 4 chromosomes

94    with a total size of 764 Mb, 98.4% of that reported previously (*16*). Of the 1.6% of the

95    previous assembly that are not contained in Smed_chr_ref_v1 by Hi-C, around half

96    (52.6%) of the scaffolds could not be assigned to a chromosome or were assigned to 2

97    chromosomes by ChrSeq, indicating their lower assembly quality (**table S3**); 33.0% and

98    12.4% of them had chromatin interaction signals with structurally complex regions of

99    chromosome 1 and 2 (**fig. S2**), respectively, suggesting that they might be alternative

100   assemblies or repetitive sequences. In our new assembly, Smed_chr_ref_v1, the two

101   ends of chromosome 4 are capped by >1,000 copies of the telomere repeat TTAGGG,

102   indicating high assembly quality.

103

104   To validate the linearity of the chromosomes in Smed_chr_ref_v1, we generated a linkage

105   map. We crossed two divergent laboratory strains of *S. mediterranea*, S2F10b and D5,

106   to produce an F2 population and genotyped individual worms with RADseq (*24*). Eighty

107   markers that are evenly distributed and genotyped in at least 98% of the F2 segregants

108   (91 of 93) were used to establish 4 linkage groups (**Fig. 1D, table S4**). These 4 linkage

109   groups represent the 4 chromosomes. The ordering of the 80 markers in the linkage map

110   is consistent with Smed_chr_ref_v1, independently supporting the quality of our

111   chromosome-scale genome assembly. This highly contiguous and complete genome

112   assembly and linkage map together facilitate further genetic and epigenetic investigations

113   of the functions of the genome in this model planarian.

114

115   **Chromosome 1 Recombination Suppression**

116   We reexamined the heterozygous alleles that define the J and V haplotypes in the newly

117   assembled genome. We found that 87.7% of the genetic markers that remained

118   heterozygous were located on chromosome 1, spanning 333 Mb at a density of 30,148

119   variants per 10 Mb (**Fig. 2A**). The remaining 12.3% of heterozygous markers were located

120   on the other 3 chromosomes at a density of 3,274 markers per 10 Mb; these likely

121   correspond to differences between highly similar copies of repetitive elements rather than

122   to true polymorphisms (**table S5**). All F2 worms (n=93) in this study were heterozygous

123   for chromosome 1 but not for chromosomes 2–4 (**table S6**), consistent with our previous

124   study (*14*). Hence, we conclude that the J/V haplotypes are on chromosome 1.

125

126   Our linkage map revealed an extremely low rate of recombination on chromosome 1 (only

127   0.5 cM for the entire chromosome; **Fig. 1D**). This is particularly remarkable because at

128   333 Mb, chromosome 1 is the largest of the 4 chromosomes, containing more than 40%

129   of the genome.

130

131   To examine directly whether chromosome 1 can recombine, we sequenced 45 single

132   sperm and 28 single oocytes from a J/V line, S2 (**Fig. 2B**). Gamete sequencing is

133   preferred because recombination events in hatchlings can be selected for by differential

6

134    fertilization or embryonic lethality. We identified 3,197 single nucleotide variants (SNVs)

135    on chromosome 1 and 3,312 on chromosome 3, covering 99% of the length of each

136    chromosome (**table S7**). SNVs were distributed at a similar density across 20 Mb

137    windows (coefficient of variation 0.38 and 0.31 for chromosome 1 and 3) (**table S7**). We

138    observed that 98% (44/45) of the sperm and 93% (26/28) of the oocytes had no

139    crossovers on chromosome 1; by contrast, only one sperm and two oocytes had no

140    crossovers on chromosome 3 (**Fig. 2B**). We conclude from this data that recombination

141    on chromosome 1 is strongly suppressed.

142

143    Consistent with our conclusion that recombination is suppressed on chromosome 1, we

144    found that during prophase I, when other chromosomes had numerous crossovers,

145    chromosome 1 formed a ring structure in the oocyte in J/V worms but not in J/J worms

146    (**Fig. 2C**). Fluorescence *in situ* hybridization with telomere probes showed that crossovers

147    between the two homologous pairs of chromosome 1 happened only in regions close to

148    the telomeres, leading to the observed ring conformation rather than the side-by-side

149    pairing seen for chromosomes 2, 3 and 4. Furthermore, Hi-C analysis showed that

150    chromosome 1 has three putative inversions, each >20 Mb (**Fig. 2D**); such inversions are

151    an established cause of crossover suppression (*25-27*). The rest of the genome has only

152    one large inversion of ~10 Mb on chromosome 2 (**fig. S2**).

153

154    **Island-Specific Evolution of Chromosome 1**

155    To investigate the genetic diversity of *S. mediterranea* in its natural environment and to

156    determine whether chromosome 1 J/V heterozygosity happens in the entire species, we

7

157     used RADseq (*24*) to sample the genomes of 70 sexual individuals from Sardinia, Corsica,

158     Sicily and Tunisia and 2 asexual individuals from Menorca (*13*) (**Fig. 3A**). To look for

159     genetic relationships between the individuals, we determined clustering of relatedness

160     measured by identity-by-state pairwise distances (*28*) and identified two superclusters:

161     animals from Sicily were closely related to those from Tunisia, and animals from Sardinia

162     were closely related to those from Corsica (**Fig. 3B**). Phylogenetic clustering (**fig. S3**) and

163     a STRUCTURE analysis (**fig. S4**) further supported this population structure.

164

165     The relatedness of the populations from Sardinia and Corsica suggested that they may

166     share genome characteristics that differ from those of the populations from Sicily and

167     Tunisia. Indeed, animals in the populations from Sardinia and Corsica had greater

168     heterozygosity on chromosome 1 than expected under Hardy-Weinberg equilibrium (**Fig.**

169     **3C**), whereas heterozygosity on the other three chromosomes in these populations and

170     on all four chromosomes in the other populations closely followed expectation. By

171     analyzing J/V haplotype markers in the wild populations, we found that the animals from

172     Sardinia/Corsica were heterozygous J/V whereas those from Sicily/Tunisia were

173     homozygous J/J. Moreover, the animals from Sardinia (n=28) contained many sites that

174     were heterozygous in more than 80% of the individuals and that were distributed along

175     the length of chromosome 1 except near the ends, whereas the animals from Sicily (n=27)

176     had very few such heterozygous sites (**Fig. 3, D and E, left**) (**table S8**). Few such

177     heterozygous sites were observed on chromosome 3 in either population (**Fig.3, D and**

178     **E, right**). These analyses suggest that chromosome 1 specifically evolved and diverged

179     on the islands of Sardinia and Corsica.

180

**A Sex-Primed Chromosome**

182     To gain insight into the island-specific suppression of recombination on chromosome 1,

183     we examined the genes located on chromosome 1. We found that all currently

184     characterized genes that are essential for development of both male and female

185     reproductive systems were found on chromosome 1, even though differential gene

186     expression analysis between juveniles and adults showed that adult-enriched genes were

187     distributed equally among all 4 chromosomes (**fig. S5**).  The critical regulatory genes—

188     *nanos* (*29*), *nhr-1* (*30, 31*), *npy-8* (*32*), *npyr-1* (*33*), and *CPEB-2* (*34*)—were not clustered

189     but distributed along the length of chromosome 1 (**Fig. 4A**). *Nanos*, *nhr-1* and *CPEB-2*

190     were transcribed from both J and V haplotypes (**fig. S6**). No genetic variants that would

191     allow determination of J- or V-haplotype-specific expression were found in *npy-8* and

192     *npyr-1* transcripts. The presence  of these genes on chromosome 1 and their crucial roles

193     in sexual development indicate the importance of chromosome 1 integrity in the

194     maintenance of sexual reproduction. The asexual lineage of *S. mediterranea,* which has

195     a translocation from chromosome 1 to chromosome 3 and is devoid of any reproductive

196     organs(*12*), likely evolved through loss of function of one or several of these genes.

197

198     We found that chromosome 1 has acquired many genes with male or female-specific

199     functions. The *CPEB-1* gene on chromosome 1 (**Fig. 4B**) is specifically required for

200     development of the female reproductive system (*34*). The presence of a female-

201     determining gene on a chromosome that does not recombine provides the ideal

202     foundation for the evolution of a sex chromosome: a loss-of-function mutation in the gene

203    on one of the two homologous chromosomes would leave the other chromosome with a

204    unique role in determining the female sex. An extended search of the literature identified

205    13 additional genes with well characterized male or female-specific functions that are

206    located on chromosome 1 (**table S9**) (*29, 34*). Two of these genes, Bic-C and sped-1, are

207    in synteny with the ZW sex chromosomes of the parasitic flatworm, *Schistosome mansoni*

208    *(35, 36).* Loss of Bic-C expression leads to a "no testes" phenotype, without affecting the

209    ovary (*29*).

210

211    To examine whether chromosome 1 has evolved a signature of a sex chromosome—

212    haplotype-specific expression of male- or female-specific genes—we examined the

213    expression of these 14 genes in the transcriptomes of sexually mature adult worms and

214    of their male copulatory organ, the penis papilla (pp) (*37*), in a J/V line. Nine of the 14

215    genes had at least one heterozygous variant in the coding sequence, which allowed us

216    to determine if these genes showed biased expression from the J or V chromosome. We

217    found that a gelsolin-like gene (KY847538.1) was expressed predominantly from the J

218    chromosome in both whole-worm and pp transcriptomes, with a larger effect in the male

219    copulatory organ. Three genes (KY847529.1, KY847536.1, and Smed-MSY4) had biased

220    expression in the whole worm transcriptome, and two genes (KY847532.1 and

221    KY847539.1) had biased expression in the male copulatory organ (**Fig. 4B, table S9**).

222    Haplotype-specific expression in 6 of 9 female- or male-related genes suggests that sex-

223    specific *cis*-regulatory variation is maintained on chromosome 1.

224

225 We have identified a chromosome in the hermaphrodite planarian *S. mediterranea* that

226 does not recombine and has acquired many genes with  sex-specific functions, making it

227 primed for evolution into a sex chromosome. The suppression of recombination on this

228 chromosome has evolved specifically on the islands of Sardinia and Corsica; such a

229 phenomenon is absent from the sampled populations of Sicily and Tunisia. Although we

230 cannot know how chromosome 1 will evolve, our findings provide a snapshot that

231 supports the hypothesis that sex chromosomes evolved from homologous autosomes

232 that acquired sex-specific roles and ceased to recombine (*1-8*). The locked J/V

233 heterozygous system may play a facilitating role in this process by maintaining sex-

234 specific alleles in the planarian population prior to the evolution of dioecy (*38*). We

235 propose that the planarian chromosome 1 haplotypes provide a unique opportunity to

236 directly examine the molecular characteristics of a sex-primed autosome.

237

238 **Methods**

239

240 **Chromosome sequencing**

241 Chromosomes were collected from multiple animals of one clonal line, S2, maintained in

242 the lab by amputation and regeneration. Chromosome spreads were prepared on

243 nuclease-free Membrane Slides (Zeiss) according to a previously developed protocol,

244 except that at the last step, the tissues were dissociated into single nuclei and dropped

245 onto the slides without squashing with a coverslip (*39*). Single chromosomes were

246 identified under a 40x lens and collected into caps of single PCR tubes by PALM

247 MicroBeam laser microdissection (Zeiss). Collected chromosomes were spun down with

11

248  a tabletop spinner in 4*u*l of PBS, and the DNA in the pellets was amplified with REPLI-g

249  Single Cell kit (QIAGEN) for sequencing on the MiSeq or HiSeq 3000 sequencing system

250  (Illumina).

251

252  **Chromosome-scale genome assembly**

253  A Hi-C sequencing library was prepared from multiple animals of the S2 strain using the

254  enzyme DpnII, following the instructions of Phase Genomics Proximo Animal kit version

255  3.0 with few modifications. Sequenced reads were aligned to dd_Smes_g4.fasta (*16*) with

256  bwa mem (version 0.7.17) (*40*). An assembly file was prepared from the SALSA (*21*)

257  output FINAL.fasta with juicebox_scripts (Phase Genomics). The assembly and .hic files

258  were loaded into Juicebox (*23*) for scaffold manipulation (i.e., split, merge, order, orient)

259  and chromosome assembly. The modified assembly file (chromosome-scale) was

260  converted to fasta (Smed_chr_ref_v1) with juicebox_assembly_converter.py.

261

262  **Oocyte and sperm sequencing**

263  Sperm were released from sexually mature S2 strain animals into CMFB (1% BSA) buffer.

264  The cell dissociation solution was dropped onto a slide and visualized under a phase

265  contrast microscope to identify single sperm. Oocytes were released from egg capsules.

266  Single sperm or oocytes were transferred into single PCR tubes for amplification with the

267  REPLI-g Single Cell kit (QIAGEN). RADseq libraries and whole genome libraries were

268  prepared and sequenced on the HiSeq 3000 or NovaSeq S2 sequencing system

269  (Illumina). RADseq sequencing data were analyzed as described in the linkage map

12

270   section. Whole-genome sequencing data were analyzed as described in the

271   recombination section.

272

273   **Linkage map**

274   A J/J line, D5, was crossed to a J/V line, S2F8b, to build an F2 population of 93 animals.

275   Genomic DNA was extracted from a fragment of each animal with Easy-DNA gDNA

276   purification kit (K180001, ThermoFisher). Sequencing libraries for RADseq were

277   prepared according to the procedures of Adapterama III(*24*) with few modifications.

278   Genetic variants were identified with Stacks (version 2.41)(*41, 42*). All variants were

279   filtered with VCFtools (version 0.1.14)(*43*) to remove insertions and deletions and to

280   select biallelic single nucleotide variants. Clusters of markers that are located within

281   200bp were removed as they most likely correspond to repetitive elements. Markers that

282   disobey Mendelian segregation were also removed. Only markers that are homozygous

283   in both parents were used to build the linkage map with R/QTL(*44*).

284

285   **Quantifying recombination**

286   Sequencing reads from S2, its oocytes, and sperm were aligned to Smed_chr_ref_v1 with

287   bwa mem (version 0.7.17). Genetic variants were jointly called by using the genome

288   analysis toolkit GATK (v.4.1.4.1) with GenomicsDB and GenotypeGVCFs (*45*). Biallelic

289   heterozygous markers in S2 were further filtered by removing abnormal markers: markers

290   with no segregation of two alleles in the gametes, clusters of markers in close proximity

291   (<200bp), and markers heterozygous in sperm. J and V haplotypes were manually

292   phased with oocytes without recombination and phased by MPR.genotyping (*46*) with all

13

293    gametes. The MPR.genotyping package was also used to impute or correct missing or

294    erroneous genotypes. The final genotype bins were used to identify and visualize

295    recombination with customized R code. Quantification of recombination was focused on

296    crossovers between long tracks of haplotypes along a chromosome. Putative gene

297    conversion events (i.e., short tracks of haplotype switches encompassing < 1% of

298    chromosome length) were not included.

299

300    **Synteny analysis**

301    Protein sequences of genes of interest were obtained from *S. mediterranea*. Such genes

302    were aligned to *S. mansoni* genomes (*36, 47*) with tblastn.

303

304    **Gene expression analysis**

305    To examine gene contents on chromosome 1, transcriptome data were downloaded from

306    NCBI Sequence Read Archive (SRA). To examine genes related to sexual development,

307    transcriptomes from sexual adults (*30, 34, 48*), sexual juveniles (*30*), and sexual adults

308    with nhr-1 RNAi (*30*) were used. To examine stem-cell-enriched genes, transcriptomes

309    from sorted X1 cells and CIW4 were used (*49, 50*).

310      All sequencing data were aligned to dd_Smed_v6 (*17*) with bwa mem (version 0.7.17).

311    Differential gene expression was analyzed with DEseq2 (version 1.26.0) (*51*). Expression

312    was quantified at the transcript level with kallisto (version 0.44.0) (*52*), and imported and

313    summarized to gene-level count matrices by tximport (*53*).

314

315    **Haplotype-specific expression**

316　To examine haplotype-specific expression of critical regulators of the reproductive system,

317　mRNA was extracted from six sexually mature animals of a J/V line, S2, and analyzed as

318　three biological replicates, with 2 animals pooled into each replicate. Nine penis papillae

319　(pp) were dissected from nine sexual adult animals of the same line and analyzed as

320　three biological replicates, with pp from 3 animals pooled into each replicate. All mRNA

321　samples were extracted on the same day and processed at the same time for library

322　preparation and sequencing to minimize technical variation. Libraries for RNA-Seq were

323　prepared with Clonetech SMARTer Stranded Total RNA-Seq (Pico) Kit. The workflow

324　consists of converting total RNA to cDNA, and then adding adapters for Illumina

325　sequencing through PCR. The PCR products are purified, and then ribosomal cDNA is

326　depleted. The cDNA fragments are further amplified with primers universal to all libraries.

327　Lastly, the PCR products are purified once more to yield the final cDNA library. Different

328　adaptors were used to multiplex samples in one lane. Sequencing was performed on

329　Illumina Novaseq 6000 with a paired-end-read 150 bp run. Data quality checks were done

330　on Illumina SAV. Demultiplexing was performed with Illumina Bcl2fastq2 v 2.17.

331　　All sequencing data were aligned to dd_Smed_v6 (*17*) with bwa mem (version

332　0.7.17). To ensure accuracy, haplotype-specific expression of the 14 genes of interest

333　was manually examined in IGV. J or V allele counts were identified for each biallelic

334　variant in the exons. For a particular gene, the allele counts were aggregated from all

335　variants on all exons for J or V haplotypes. The allele counts were then subjected to

336　binomial test and Bonferroni correction (*54*) to determine if the observed allele bias was

337　statistically significant. P-values > 1 after Bonferroni correction were set to 1.

338

### *De novo* map, parameter optimization and phylogenetic inference

339

340 All datasets were run through the *de novo* pipeline as implemented in STACKS version

341 2.52 (*41, 42*). First, paired-end reads were de-multiplexed and filtered for quality using

342 PROCESS_RADTAGS and truncated to a length of 135 bp. Individual reads with Phred-

343 scores below 30 or uncalled bases were discarded (96.9% of reads passed the quality

344 filters). Optimal parameters were identified following Paris et al. (*55*) guidelines by running

345 multiple iterations of the *de novo* pipeline and varying just one parameter with each new

346 iteration on a subset of 13 samples from the same population, [Sic_mar], following

347 recommendations (*56*). We varied the minimum stack depth (*-m*) between 1 and 6 (m1-

348 m6), the number of mismatches allowed between stacks (*-M*) between 1 and 10 (M1-

349 M10), and the number of mismatches allowed to merge catalog loci (*-n*) while keeping all

350 other parameters constant (m3, M2 and n0). We then compared the number of

351 polymorphic assembled loci across samples using the 80% sample representation cutoff

352 (*r80*) and the gain or loss of polymorphic loci with each new iteration. Once *-m* and *-M*

353 were optimized, we assessed *-n* by evaluating the change in number of polymorphic loci

354 for $n = M - 1$, $n = M$ and $n = M + 1$. RAD loci were then assembled using the wrapper

355 'denovo_map.pl' in Stacks and the following parameters set: $m = 3$, $M = 2$ and $n = 3$.

356 Assembled loci that were present in 75% of all individuals were kept from POPULATIONS,

357 and a minor allele frequency (MAF) filter of 0.04 (*--min-maf*) was used to filter out

358 singleton SNPs that could mask population structure, along with a maximum observed

359 heterozygosity (*--max-obs-het*) filter of 0.99 to remove potentially paralogous loci (*57*).

360 Additionally, to build the phylogenetic tree, we concatenated all RADseq loci after filtering

361 (*--phylip-var-all*). Phylogenetic trees were built by maximum likelihood using RAxML-NG

362    v0.9.0 (*58*), starting from a random seed and applying a GTR+G substitution model and

363    1000 bootstrap replicates. The sample from Menorca (Sme7-5_men) was used as an

364    outgroup to root the tree.

365

**Population structure analysis**

367    For this analysis we excluded the *outgroup* sequence and ran POPULATIONS to retain loci

368    present in all populations (-p 10) and 75% of the individuals present in each population (-

369    r 0.75). Based on the loci that passed our filtering criteria, a random whitelist of 1000 loci

370    was generated and run through POPULATIONS once again, with the same criteria but

371    retaining the first SNP at each locus (--*write-single-snp*). The output was exported in

372    STRUCTURE format, and STRUCTURE 2.3.4 (*59*) was used to infer population structure

373    with 10,000 chains as burn-in and 100,000 MCMC chains with 20 iterations for $K = [1-11]$.

374    The resulting files were run through STRUCTURE HARVESTER (*60*), and the optimal $K$

375    was determined (*61*).

376

**Acknowledgements**

385  Center of Excellence at the California Nano Systems Institute at UCLA with funding from

386  NIH Shared Instrumentation Grant S10OD025017 and NSF Major Research

387  Instrumentation grant CHE-0722519.

388

392

393  **Competing interests**

394  The authors have declared that no competing interests exist.

395

396  **Data availability**

397  The authors confirm that all data underlying the findings are fully available without

398  restriction. All sequencing data will be available from the NCBI SRA database (accession

399  number PRJNA731187). The chromosome scale genome assembly for sexual *Schmidtea*

400  *mediterranea* is openly available at: https://planosphere.stowers.org/;

401  http://planmine.mpi-cbg.de/planmine/begin.do

402  **References**

403  1.   D. Bachtrog, A dynamic view of sex chromosome evolution. *Curr Opin Genet Dev* **16**,
404        578-585 (2006).
405  2.   B. T. Lahn, D. C. Page, Four evolutionary strata on the human X chromosome. *Science*
406        **286**, 964-967 (1999).
407  3.   W. R. Rice, Evolution of the Y Sex Chromosome in Animals. *BioScience* **46**, 331-343
408        (1996).
409  4.   B. Charlesworth, The evolution of chromosomal sex determination and dosage
410        compensation. *Curr Biol* **6**, 149-162 (1996).
411  5.   J. Hodgkin, Genetic sex determination mechanisms and evolution. *Bioessays* **14**, 253-261
412        (1992).

413   6.    B. Charlesworth, The evolution of sex chromosomes. *Science* **251**, 1030-1033 (1991).

414   7.    E. Jablonka, M. J. Lamb, The evolution of heteromorphic sex chromosomes. *Biol Rev*
415         *Camb Philos Soc* **65**, 249-276 (1990).

416   8.    J. J. Bull, *Evolution of sex determining mechanisms*. (The Benjamin/Cummings
417         Publishing Company, Inc., 1983), pp. 316.

418   9.    H. J. Muller, Genetic Variability, Twin Hybrids and Constant Hybrids, in a Case of
419         Balanced Lethal Factors. *Genetics* **3**, 422-499 (1918).

420   10.   D. Charlesworth, Evolution of recombination rates between sex chromosomes. *Philos*
421         *Trans R Soc Lond B Biol Sci* **372**, (2017).

422   11.   P. W. Reddien, A. Sanchez Alvarado, Fundamentals of planarian regeneration. *Annu Rev*
423         *Cell Dev Biol* **20**, 725-757 (2004).

424   12.   P. A. Newmark, A. Sanchez Alvarado, Not your father's planarian: a classic model enters
425         the era of functional genomics. *Nat Rev Genet* **3**, 210-219 (2002).

426   13.   E. M. Lazaro *et al.*, Schmidtea mediterranea phylogeography: an old species surviving on
427         a few Mediterranean islands? *BMC Evol Biol* **11**, 274 (2011).

428   14.   L. Guo, S. Zhang, B. Rubinstein, E. Ross, A. S. Alvarado, Widespread maintenance of
429         genome heterozygosity in Schmidtea mediterranea. *Nat Ecol Evol* **1**, 19 (2016).

430   15.   R. M. Zayas *et al.*, The planarian Schmidtea mediterranea as a model for epigenetic
431         germ cell specification: analysis of ESTs from the hermaphroditic strain. *Proc Natl Acad*
432         *Sci U S A* **102**, 18491-18496 (2005).

433   16.   M. A. Grohme *et al.*, The genome of Schmidtea mediterranea and the evolution of core
434         cellular mechanisms. *Nature* **554**, 56-61 (2018).

435   17.   H. Brandl *et al.*, PlanMine--a mineable resource of planarian biology and biodiversity.
436         *Nucleic Acids Res* **44**, D764-773 (2016).

437   18.   S. M. Robb, E. Ross, A. Sanchez Alvarado, SmedGD: the Schmidtea mediterranea
438         genome database. *Nucleic Acids Res* **36**, D599-606 (2008).

439   19.   A. Weise *et al.*, High-throughput sequencing of microdissected chromosomal regions.
440         *Eur J Hum Genet* **18**, 457-462 (2010).

441   20.   L. Ma *et al.*, Direct determination of molecular haplotypes by chromosome
442         microdissection. *Nat Methods* **7**, 299-301 (2010).

443   21.   J. N. Burton *et al.*, Chromosome-scale scaffolding of de novo genome assemblies based
444         on chromatin interactions. *Nat Biotechnol* **31**, 1119-1125 (2013).

445   22.   E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals
446         folding principles of the human genome. *Science* **326**, 289-293 (2009).

447   23.   N. C. Durand *et al.*, Juicebox Provides a Visualization System for Hi-C Contact Maps with
448         Unlimited Zoom. *Cell Syst* **3**, 99-101 (2016).

449   24.   N. J. Bayona-Vasquez *et al.*, Adapterama III: Quadruple-indexed, double/triple-enzyme
450         RADseq libraries (2RAD/3RAD). *PeerJ* **7**, e7724 (2019).

451   25.   T. Dobzhansky, C. Epling, The Suppression of Crossing Over in Inversion Heterozygotes of
452         Drosophila Pseudoobscura. *Proc Natl Acad Sci U S A* **34**, 137-141 (1948).

453   26.   D. E. Miller *et al.*, The Molecular and Genetic Characterization of Second Chromosome
454         Balancers in Drosophila melanogaster. *G3 (Bethesda)* **8**, 1161-1171 (2018).

455  27.  D. E. Miller *et al.*, Rare recombination events generate sequence diversity among
456       balancer chromosomes in Drosophila melanogaster. *Proc Natl Acad Sci U S A* **113**,
457       E1352-1361 (2016).
458  28.  X. Zheng *et al.*, A high-performance computing toolset for relatedness and principal
459       component analysis of SNP data. *Bioinformatics* **28**, 3326-3328 (2012).
460  29.  Y. Wang, R. M. Zayas, T. Guo, P. A. Newmark, nanos function is essential for
461       development and regeneration of planarian germ cells. *Proc Natl Acad Sci U S A* **104**,
462       5901-5906 (2007).
463  30.  S. Zhang *et al.*, A nuclear hormone receptor and lipid metabolism axis are required for
464       the maintenance and regeneration of reproductive organs. *bioRxiv*, 279364 (2018).
465  31.  M. E. Tharp, J. J. Collins, 3rd, P. A. Newmark, A lophotrochozoan-specific nuclear
466       hormone receptor is required for reproductive system development in the planarian.
467       *Dev Biol* **396**, 150-157 (2014).
468  32.  J. J. Collins, 3rd *et al.*, Genome-wide analyses reveal a role for peptide hormones in
469       planarian germline development. *PLoS Biol* **8**, e1000509 (2010).
470  33.  A. Saberi, A. Jamal, I. Beets, L. Schoofs, P. A. Newmark, GPCRs Direct Germline
471       Development and Somatic Gonad Function in Planarians. *PLoS Biol* **14**, e1002457 (2016).
472  34.  L. Rouhana, J. Tasaki, A. Saberi, P. A. Newmark, Genetic dissection of the planarian
473       reproductive system through characterization of Schmidtea mediterranea CPEB
474       homologs. *Dev Biol* **426**, 43-55 (2017).
475  35.  M. Berriman *et al.*, The genome of the blood fluke Schistosoma mansoni. *Nature* **460**,
476       352-358 (2009).
477  36.  A. V. Protasio *et al.*, A systematically improved high quality genome and transcriptome
478       of the human blood fluke Schistosoma mansoni. *PLoS Negl Trop Dis* **6**, e1455 (2012).
479  37.  T. Chong, J. M. Stary, Y. Wang, P. A. Newmark, Molecular markers to characterize the
480       hermaphroditic reproductive system of the planarian Schmidtea mediterranea. *BMC*
481       *Dev Biol* **11**, 69 (2011).
482  38.  D. Charlesworth, S. David, *The evolution of dioecy and plant sex chromosome systems*.
483       Sex Determination in Plants (Garland Science, ed. 1st, 1999).
484  39.  L. Guo *et al.*, An adaptable chromosome preparation methodology for use in
485       invertebrate research organisms. *BMC Biol* **16**, 25 (2018).
486  40.  H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler
487       transform. *Bioinformatics* **25**, 1754-1760 (2009).
488  41.  J. Catchen, P. A. Hohenlohe, S. Bassham, A. Amores, W. A. Cresko, Stacks: an analysis
489       tool set for population genomics. *Mol Ecol* **22**, 3124-3140 (2013).
490  42.  J. M. Catchen, A. Amores, P. Hohenlohe, W. Cresko, J. H. Postlethwait, Stacks: building
491       and genotyping Loci de novo from short-read sequences. *G3 (Bethesda)* **1**, 171-182
492       (2011).
493  43.  P. Danecek *et al.*, The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158
494       (2011).
495  44.  K. W. Broman, H. Wu, S. Sen, G. A. Churchill, R/qtl: QTL mapping in experimental
496       crosses. *Bioinformatics* **19**, 889-890 (2003).
497  45.  A. McKenna *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing
498       next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010).

20

499  46.  W. Xie *et al.*, Parent-independent genotyping for constructing an ultrahigh-density
500      linkage map based on population sequencing. *Proc Natl Acad Sci U S A* **107**, 10578-
501      10583 (2010).
502  47.  F. Luo *et al.*, An improved genome assembly of the fluke Schistosoma japonicum. *PLoS*
503      *Negl Trop Dis* **13**, e0007612 (2019).
504  48.  E. L. Davies *et al.*, Embryonic origin of adult stem cells required for tissue homeostasis
505      and regeneration. *Elife* **6**,  (2017).
506  49.  A. Zeng *et al.*, Prospectively Isolated Tetraspanin(+) Neoblasts Are Adult Pluripotent
507      Stem Cells Underlying Planaria Regeneration. *Cell* **173**, 1593-1608 e1520 (2018).
508  50.  E. M. Duncan, A. D. Chitsazan, C. W. Seidel, A. Sanchez Alvarado, Set1 and MLL1/2
509      Target Distinct Sets of Functionally Different Genomic Loci In Vivo. *Cell Rep* **13**, 2741-
510      2755 (2015).
511  51.  M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for
512      RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
513  52.  N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq
514      quantification. *Nat Biotechnol* **34**, 525-527 (2016).
515  53.  C. Soneson, M. I. Love, M. D. Robinson, Differential analyses for RNA-seq: transcript-
516      level estimates improve gene-level inferences. *F1000Res* **4**, 1521 (2015).
517  54.  Etymologia: Bonferroni correction. *Emerg Infect Dis* **21**, 289 (2015).
518  55.  J. R. Paris, J. R. Stevens, J. M. Catchen, Lost in parameter space: a road map for stacks.
519      *Methods in Ecology and Evolution* **8**, 1360-1373 (2017).
520  56.  N. C. Rochette, J. M. Catchen, Deriving genotypes from RAD-seq short-read data using
521      Stacks. *Nat Protoc* **12**, 2640-2659 (2017).
522  57.  C. S. Stobie, C. J. Oosthuizen, M. J. Cunningham, P. Bloomer, Exploring the
523      phylogeography of a hexaploid freshwater fish by RAD sequencing. *Ecol Evol* **8**, 2326-
524      2342 (2018).
525  58.  A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, RAxML-NG: a fast, scalable
526      and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*
527      **35**, 4453-4455 (2019).
528  59.  J. K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using
529      multilocus genotype data. *Genetics* **155**, 945-959 (2000).
530  60.  D. A. Earl, B. M. vonHoldt, STRUCTURE HARVESTER: a website and program for
531      visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
532      *Genetics Resources* **4**, 359-361 (2012).
533  61.  G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using
534      the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611-2620 (2005).
535  62.  L. A. Ouellette, R. W. Reid, S. G. Blanchard, C. R. Brouwer, LinkageMapView-rendering
536      high-resolution linkage and QTL maps. *Bioinformatics* **34**, 306-307 (2018).
537

538

539 **Figure Legends**

540  **Figure 1 Chromosome-scale genome assembly** (A) Schematic of using Hi-C, ChrSeq,

541  and a linkage map to transform 481 scaffolds (*16*) into a chromosome-scale genome

542  assembly. (B) Chromosomes from mitotic cells were spread onto membrane slides for

543  laser capture and sequencing. Numbers denote the identity of chromosomes present in

544  representative samples. (C) Contact heatmap of chromatin interactions visualized in

545  Juicebox (*23*); black boxes denote the four chromosomes in the final assembly. (D)

546  Linkage map visualized by LinkageMapView (*62*) shows lack of recombination on

547  chromosome 1. Tick marks and labels denote genetic markers.

548

549  **Figure 2 Recombination suppression and inversions on chromosome 1** (A)

550  Distribution of heterozygous variants maintained in the S2 inbreeding pedigree along the

551  four chromosomes; y-axis shows variant counts per 10 Mb. (B) Number of meiotic

552  recombination events on chromosomes 1 and 3 in oocytes (red) and sperm (cyan). Dots

553  represent individual gametes, summarized with violin plots. (C) Chromosome 1 in late

554  prophase I in the oocytes from J/V worms (left and right panels) have fewer crossovers

555  than those in oocytes from J/J worms (middle panel). Chromosomes in the right panel

556  show FISH results with probes for telomeric repeats (TTAGGG, magenta) and

557  centromere-like repeats (yellow). (D) Chromatin contact heatmap for chromosome 1.

558  Dashed-line rectangles indicate locations of potential inversions. Arrows show chromatin

559  contact signals with inversion regions from unassigned scaffolds.

560

561  **Figure 3 Island-specific evolution of chromosome 1 heterozygosity** (A) Collection

562  sites of 72 wild isolates from Mediterranean islands and Tunisia. (B) Maximum likelihood

22

563    tree of 2 asexual isolates from Menorca (MEN_GOR) and 70 sexual isolates from

564    Sardinia (SAR), Corsica (COR), Sicily (SIC) and Tunisia (TUN). Dots denote individual

565    worms, with colors corresponding to their collection sites as noted in the legend. (C)

566    Observed vs. expected heterozygosity for each chromosome (denoted by color) in each

567    population (denoted by shape). Dashed line corresponds to equality between observed

568    values and those expected under Hardy-Weinberg equilibrium and is shown as a guide

569    to the eye; the departure of points corresponding to chromosome 1 in COR and SAR from

570    this expectation is notable. (D-E) Number of sites heterozygous in >80% of the population

571    per 10 kb window along chromosome 1 (left) and chromosome 3 (right) in Sardinia (D)

572    and in Sicily (E).

573

574    **Figure 4 Acquisition and haplotype-specific expression of sex-related genes**

575    (A) A putative model of the structures of the J and V haplotypes of chromosome 1, with

576    locations of critical genes for sexual development (cpeb-1 is female-specific). Arrows

577    denote putative inversions. (B) Comparison of read counts for the J and V alleles of key

578    genes in the transcriptomes of whole sexually mature worms and penis papillae (pp).

579    Bonferroni-corrected P-values from a binomial test of equal expression are shown.

580

581

582    **Supplemental Figure 1 Chromosome assignment by ChrSeq** (A-C) Chromosome

583    sequencing depth of different scaffolds are consistent with linkage group assignment.

584    Linkage groups: lg1, lg11, lg17, lg4, lg7. (D-F) *de novo* assignment of 481 scaffolds to

23

585   chromosomes by sequencing depth.  Chromosomes: chromosome 1 (A, D); chromosome

586   2 (B, E), chromosome 3 (C, F).

587

588   **Supplemental Figure 2 Putative inversions in the planarian genome detected by**

589   **chromatin contact heatmap**

590    (A) chromosome 1 chromatin contact heatmap showing 3 potential inversions (dashed-

591   line rectangles) visualized in Juicebox. (B) A complex structural variant on chromosome

592   2. Green outline: dd_Smes_g4 scaffolds. Blue outline: chromosome boundaries.

593

594   **Supplemental Figure 3 Phylogenetic tree of wild isolates of *Schmidtea***

595   ***mediterranea*** The tree is rooted with one asexual animal from Menorca (Sme7-5). The

596   tree on the left (A) includes two animals from SAR and SIC with higher genome

597   heterozygosity than the cohorts from the same collection sites. Phylogenetic relations

598   agree with the previous designation of eastern and western populations (*13*). SAR:

599   Sardinia; COR: Corsica; SIC: Sicily; TUN: Tunisia.

600

601   **Supplemental Figure 4 STRUCTURE analysis of wild isolates.** Evanno     analysis

602   supports a K=2 split, dividing geographically the samples in an Eastern and a Western

603   group. With K=3, a Tunisian signal was found to be present in all other 3 groups (light

604   blue).

605

606   **Supplemental Figure 5 Genes enriched in the sexual reproductive system were**

607   **distributed equally on all 4 chromosomes** (A) Number of genes and total length of

608 each chromosome. Genes were based on dd_Smed_v6 annotation. (B) Differentially

609 expressed genes and their localizations on chromosome 1 and the rest of the

610 chromosomes. chr234/chr1: the ratio of genes on chr2, 3, and 4 to chr1. (C) Differentially

611 expressed genes and their localizations on chromosome 1, 2, 3, and 4. Proportion:

612 number of differentially expressed genes to the total number of genes per chromosome.

613 (B-C) juvenile: comparing public transcriptome of juvenile animals to adult animals (*30,*

614 *34, 48*); nhr1 RNAi: comparing public transcriptome of nhr-1 RNAi animals to adult

615 animals (*30*); pp: comparing transcriptome of penis papillae to whole worm, generated in

616 current study; neoblast: comparing public transcriptome of sorted X1 cells to whole worm

617 in CIW4 animals (*49, 50*).

618

619 **Supplemental Figure 6 *Nanos, nhr-1* and *CPEB-2* transcribed from both J- and V-**

620 **haplotypes** The sequencing reads from genomes (J haplotype, V haplotype, S2) and

621 transcriptomes (pp, ww) were aligned to Smed_chr_ref_v1. Shown above are reads for

622 the gene CPEB-2. Three variants with J and V alleles are marked by colored bars. In the

623 genome of the adult worm S2, both J and V alleles are present at equal ratios. In the

624 genome of oocytes (J and V haplotype), the variants have only one allele. In the

625 transcriptomes of penis papillae (pp) and whole worm (ww), both J and V alleles are

626 present at equal ratios.

627

628 **Table S1 Assigning 481 scaffolds to chromosomes with 17 sequenced**

629 **chromosomes**

630 Chromosome sample information (i.e., ploidy, identity and potential contamination),

631 sequencing coverage and the chromosome assignment for each of the 481 scaffolds from

632 the assembly dd_Smes_g4. In chromosome assignment, 0 (scaffolds that could not be

633 assigned to chromosomes by chrseq); 5 (scaffolds that were assigned to two

634 chromosomes by chrseq); 1-4 (chromosomes).

635

636 **Table S2 Linking 481 dd_Smes_g4 scaffolds into chromosomes with Hi-C**

637 **chromatin contact sequencing**

638 dd_Smes_g4 scaffolds that were assigned or unassigned to Smed_chr_ref_v1

639 (chrAssembly), their identity in SALSA assembly of raw Hi-C sequencing data

640 (salsa_scaffolds), and their assignment to chromosomes by chromosome sequencing

641 (chrseq). In chromosome assignment, 0 (scaffolds that could not be assigned to

642 chromosomes by chrseq); 5 (scaffolds that were assigned to two chromosomes by

643 chrseq); 1-4 (chromosomes).

644

645 **Table S3 Summary of the final chromosome scale genome assembly**

646 **(Smed_chr_ref_v1)**

647 Hi-C detected 26 assembly errors in dd_Smes_g4, 5 of which were confirmed by

648 chromosome sequencing to be inter-chromosomal mis-joining. A total of 97 dd_Smes_g4

649 scaffolds were not assigned to Smed_chr_ref_v1.

650

651 **Table S4 Genetic markers and distances in the linkage map established from a F2**

652 **mapping population**

26

653 Linkage groups L.3, L.6 and L.8 only contained one genetic marker, which were not

654 included in the table. Chromosomes 2 and 4 were split into two linkage groups each, likely

655 due to the small size of the F2 mapping population.

656

657 **Table S5 Distribution of Genetic variants that are maintained heterozygous in the**

658 **inbreeding pedigree**

659 Chromosomes were divided into 10Mb windows. Heterozygous variants were identified

660 from the inbreeding pedigree from S2 to S2F9b.

661

662 **Table S6 Chromosome 1 was heterozygous in all samples of a F2 population**

663 The genetic markers were heterozygous in the J/V parent (parent_S2F10B_2A and

664 parent_S2F10B_2B) and homozygous in the J/J parent (parent_D5-1).

665 parent_S2F10B_2A and parent_S2F10B_2B are clones. Genotyping data from RADseq

666 of 291 F2 samples were listed. The 291 F2 samples correspond to 93 unique segregants.

667 0/0: homozygous reference allele; 1/1: homozygous alternative allele; 0/1:

668 heterozygous; ./.: missing data.

669

670 **Table S7 Recombination in the gametes**

671 Distribution of heterozygous variants identified in the J/V line S2 and used in the crossover

672 assessment along chromosome 1 and 3, in 20Mb windows. Number of crossovers

673 identified per gamete on chromosome 1 and 3.

674

675 **Table S8 Chromosome 1 genotypes of wild isolates in Sardinia and Sicily**

27

676    Genotyping data for 6 different collection sites in Sardinia and 2 different collection sites

677    in Sicily were aggregated. Row 1 are names of individual animals. 0/0: homozygous

678    reference allele; 1/1: homozygous alternative allele; 0/1: heterozygous; ./.: missing data.

679

680    **Table S9 Allele specific expression of "male" and "female" genes**

681    Genes with well characterized sex-related functions were identified from published work.

682    Their locations on Smed_chr_ref_v1 and *Schistosoma mansoni* chromosomes were

683    determined. The expressed J or V alleles for each of the 9 'male' or 'female' genes were

684    quantified in the transcriptomes of whole worms and isolated male copulatory organ,
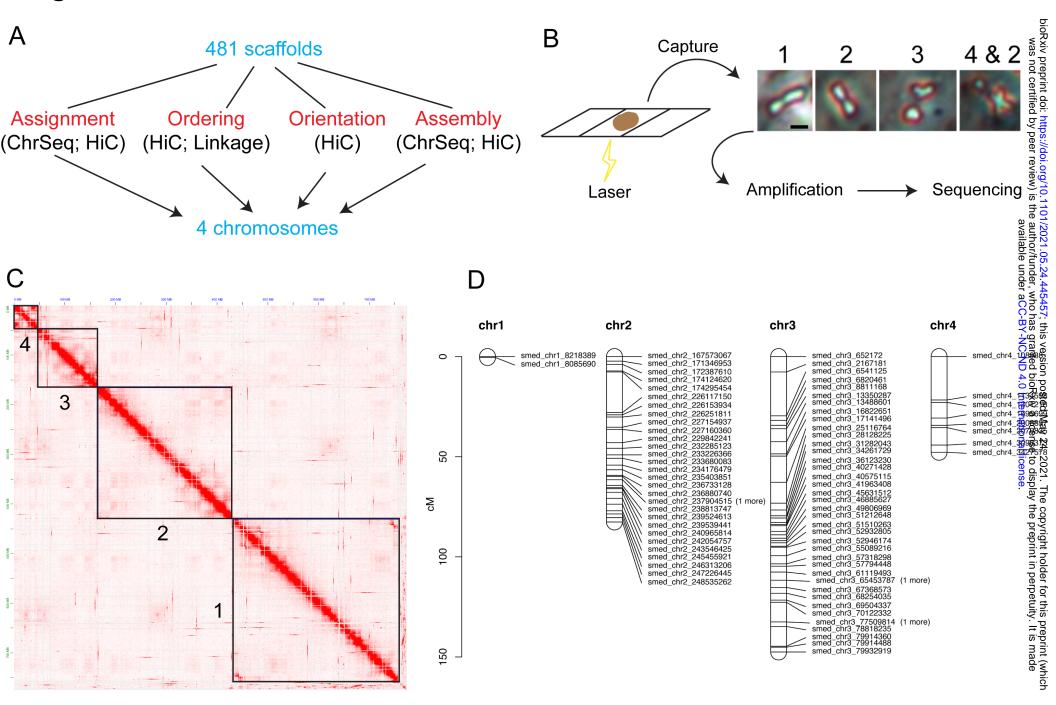
685    penis papillae.
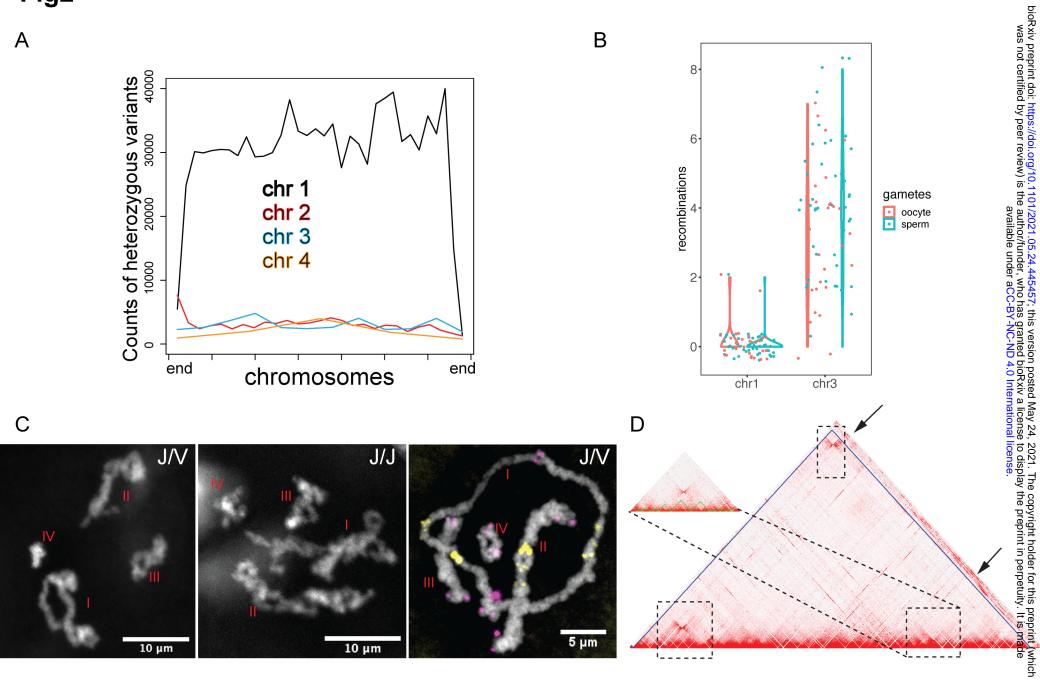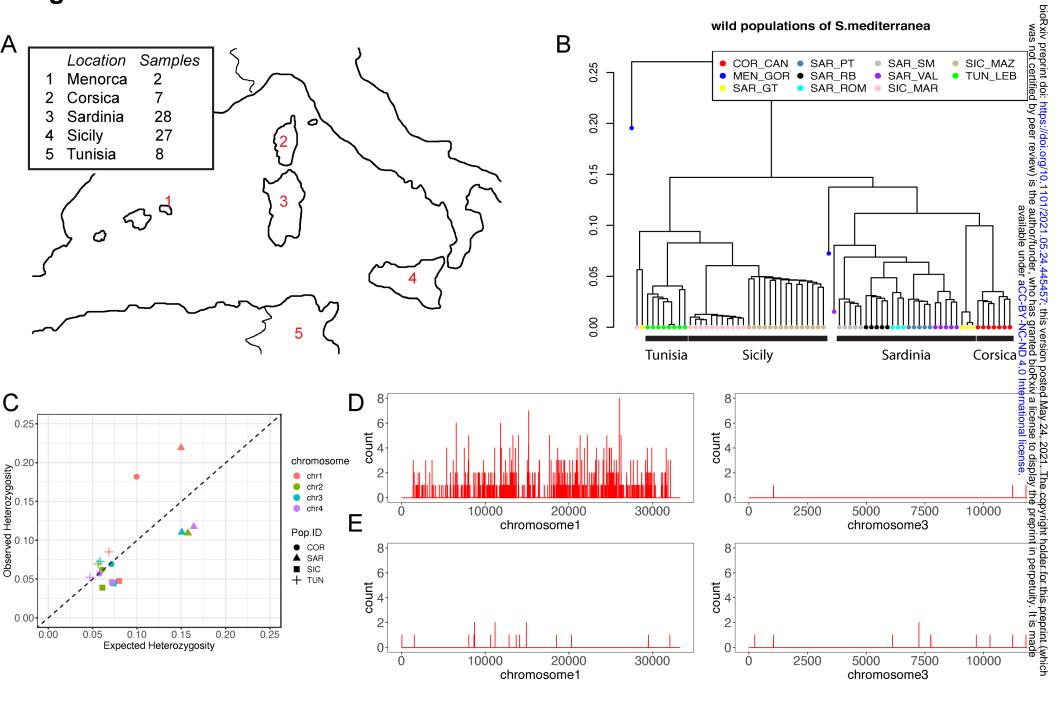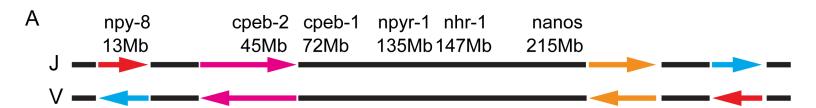
686

687

# Fig1

## A



481 scaffolds

Assignment (ChrSeq; HiC)    Ordering (HiC; Linkage)    Orientation (HiC)    Assembly (ChrSeq; HiC)

4 chromosomes

## B



Capture

1    2    3    4 & 2

Laser    Amplification → Sequencing

## C



## D



**chr1**

smed_chr1_8218389
smed_chr1_8085690

**chr2**

smed_chr2_167573067
smed_chr2_171346953
smed_chr2_172387610
smed_chr2_174124620
smed_chr2_174295454
smed_chr2_226117150
smed_chr2_226153934
smed_chr2_226251811
smed_chr2_227154937
smed_chr2_227160360
smed_chr2_229842241
smed_chr2_232285123
smed_chr2_233226366
smed_chr2_233680083
smed_chr2_234176479
smed_chr2_235403851
smed_chr2_236733128
smed_chr2_236880740
smed_chr2_237904515 (1 more)
smed_chr2_238813747
smed_chr2_239524613
smed_chr2_239539441
smed_chr2_240965814
smed_chr2_242054757
smed_chr2_243546425
smed_chr2_245455921
smed_chr2_246313206
smed_chr2_247226445
smed_chr2_248535262

**chr3**

smed_chr3_652172
smed_chr3_2167181
smed_chr3_6541125
smed_chr3_6820461
smed_chr3_8811168
smed_chr3_13350287
smed_chr3_13488601
smed_chr3_16822651
smed_chr3_17141496
smed_chr3_25116764
smed_chr3_28128225
smed_chr3_31282043
smed_chr3_34261729
smed_chr3_36123230
smed_chr3_40271428
smed_chr3_40575115
smed_chr3_41963408
smed_chr3_45631512
smed_chr3_46885627
smed_chr3_49806969
smed_chr3_51212648
smed_chr3_51510263
smed_chr3_52932805
smed_chr3_52946174
smed_chr3_55089216
smed_chr3_57318298
smed_chr3_57794448
smed_chr3_61119493
smed_chr3_65453787 (1 more)
smed_chr3_67368573
smed_chr3_68254035
smed_chr3_69504337
smed_chr3_70122332
smed_chr3_77509814 (1 more)
smed_chr3_78818235
smed_chr3_79914360
smed_chr3_79914488
smed_chr3_79932919

**chr4**

smed_chr4_...
smed_chr4_...
smed_chr4_...
smed_chr4_...
smed_chr4_...
smed_chr4_...

cM

# Fig2

A



B



C



D

# Fig3

# Fig4

A

npy-8      cpeb-2 cpeb-1  npyr-1  nhr-1         nanos
13Mb       45Mb   72Mb   135Mb 147Mb         215Mb



B

| Accession | J_worm | V_worm | J_proportion | P value (worm) | J_pp | V_pp | J_proportion | P value (pp) |
|---|---|---|---|---|---|---|---|---|
| KY847529.1 | 242 | 146 | 0.62 | 2.26E-05 | 295 | 297 | 0.50 | 1 |
| KY847536.1 | 2451 | 2142 | 0.53 | 9.82E-05 | 42 | 29 | 0.59 | 1 |
| Smed-MSY4 | 5398 | 5802 | 0.48 | 0.00251 | 60 | 81 | 0.43 | 1 |
| KY847538.3 | 448 | 353 | 0.56 | 0.0159 | 6330 | 1926 | 0.77 | 9E-323 |
| KY847532.1 | 135 | 145 | 0.48 | 1 | 2 | 18 | 0.10 | 0.00724 |
| KY847539.1 | 142 | 168 | 0.46 | 1 | 446 | 341 | 0.57 | 0.0037 |
| KY847534.1 | 11 | 9 | 0.55 | 1 | 0 | 0 | NA | NA |
| Smed-elav2 | 182 | 155 | 0.54 | 1 | 0 | 0 | NA | NA |
| Smed-bic-C | 278 | 257 | 0.52 | 1 | 0 | 0 | NA | NA |

# SupFig1



**Chr1_b1_7a**

**Chr2_b1-4e**

**Chr3_b1-4b**

# SupFig2

A



B

# SupFig3

# SupFig4

# SupFig5

**A**

|  | genes | length (bp) |
|---|---|---|
| chr1 | 12018 | 332698990 |
| chr2 | 10296 | 265042666 |
| chr3 | 5470 | 118597112 |
| chr4 | 2932 | 47107876 |

**B**

|  | chr1 | chr234 | chr234/chr1 |
|---|---|---|---|
| juvenile | 1101 | 1763 | 1.601271571 |
| neoblast | 2261 | 3440 | 1.521450686 |
| nhr1RNAi | 844 | 1387 | 1.643364929 |

**C**

|  | chr1 | | chr2 | | chr3 | | chr4 | |
|---|---|---|---|---|---|---|---|---|
|  | number | proportion | number | proportion | number | proportion | number | proportion |
| juvenile | 1101 | 0.092 | 990 | 0.096 | 517 | 0.095 | 256 | 0.087 |
| nhr1RNAi | 844 | 0.070 | 783 | 0.076 | 406 | 0.074 | 198 | 0.068 |
| pp | 240 | 0.020 | 208 | 0.020 | 100 | 0.018 | 66 | 0.023 |
| neoblast | 2261 | 0.188 | 1880 | 0.183 | 1036 | 0.189 | 524 | 0.179 |

# SupFig6