Published in final edited form as:

Nat Ecol Evol. 2016 ; 1: . doi:10.1038/s41559-016-0019.

Widespread maintenance of genome heterozygosity in *Schmidtea mediterranea*

Longhua Guo^{1,2}, Shasha Zhang¹, Boris Rubinstein¹, Eric Ross^{1,3}, and Alejandro Sánchez Alvarado^{1,2,3,*}

¹Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA

²Molecular Biology Program, Department of Neurobiology and Anatomy, University of Utah, Salt Lake City, Utah 84112, USA

³Howard Hughes Medical Institute, Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA

Abstract

Loss of heterozygosity through inbreeding or mitotic errors leads to reductions in progeny survival and fertility. Loss of heterozygosity is particularly exacerbated in geographically isolated populations, which are prone to inbreeding depression and faster rates of extinction. The regenerative capacities of the hermaphroditic biotype of the planarian *Schmidtea mediterranea* allowed us to perform a systematic genetic test of Mendelian segregation and study the loss of heterozygosity in the Spiralian superclade in general and planarians in particular. We discovered that ~300 Mb (~37.5%) of the genome retains heterozygosity even after ten generations of inbreeding, and show that these chromosomal regions have low diversity and recombination rates in wild populations. Our genetic and genomic analyses establish *S. mediterranea* as a genetically tractable system. The research also opens the door to study the evolutionary basis of non-Mendelian mechanisms, the adaptive advantages of chromosome structural heterozygotes and their potential relationship to the robust regenerative capacities of planarians.

The freshwater, free-living flatworm (platyhelminth) *S. mediterranea* has emerged as a powerful model system for studying adult stem cells and organ regeneration^{1,2}. This species exists naturally as asexual and sexual reproductive diploid strains with four pairs of chromosomes, distinguishable by a chromosomal translocation that is only present in the asexual biotype³. Animals harbouring this translocation reproduce by transverse fission and do not differentiate germ lines or the somatic copulatory apparatus, whereas individuals

Reprints and permissions information is available at www.nature.com/reprints.

*Correspondence and requests for materials should be addressed to A.S.A. asa@stowers.org.

Author contributions

L.G. and A.S.A. conceived and designed the project. L.G. performed the experiments and data analysis. S.Z. helped with worm culture, crosses and genotyping. E.R. helped with identifying SNPs from the (S2F5b)F6b family. L.G. and B.R. developed algorithms to analyse Mendelian inheritance in small families and performed correlation analysis. L.G. developed figures. L.G. and A.S.A. interpreted the data and prepared the manuscript.

Supplementary information is available for this paper.

Competing interests

The authors declare no competing financial interests.

lacking this translocation are hermaphroditic (Supplementary Fig. 1) and do not reproduce asexually. Although both biotypes display remarkable regenerative capacities, the ease of cultivation of the asexual *S. mediterranea* has generally favoured its study over the sexual biotype. Therefore, little is known about the mechanisms of heredity in this organism; genetic studies in free-living flatworms using microsatellite and ploidy measurements have been mostly limited to *S. polychroa*, a primarily parthenogenetic polyploid species⁴⁻⁸. Because the sexual *S. mediterranea* biotype can be found on the Mediterranean island of Sardinia, its natural history provides an additional opportunity to study the genetics of an animal that possesses robust regenerative capacities in the context of geographic isolation (see Supplementary Information). By developing the necessary culture conditions, we have taken advantage of the sexual *S. mediterranea* biotype to examine genome-wide Mendelian inheritance and genetic variability in planarians; this effort is not only the first such exploration in the Platyhelminthes, but also in the Lophotrochozoans/ Spiralian, a molecularly understudied, species-rich animal superclade⁹.

Results

To establish methods for the genetic analysis of *S. mediterranea*, we sought to determine their reproductive behaviour. When pairs of worms were kept together in culture, we observed animals conjoining in a tail-to-tail configuration (Fig. 1a, Supplementary Video 1), a behaviour we term copulatory embrace. Once the embrace concluded, animals were separated and kept in isolation. One week after copulation, each animal produced one to two egg capsules per week, with each capsule yielding up to ten hatchlings. Hatchlings eclose from the capsule as early as two weeks after deposition and are devoid of a reproductive system^{10,11}. With regular feedings, the juveniles grow to develop testes, ovaries and the attendant somatic reproductive organs, reaching sexual maturity in about six weeks. Altogether, the life cycle of the sexual *S. mediterranea* biotype spans ~2.5 months (Fig. 1b).

To determine if individual hermaphroditic planaria could reproduce through self-fertilization, we selected juveniles that had yet to develop a reproductive system (virgins) and reared them to sexual maturity in isolation. In parallel, we set up crossing pairs of sexually mature virgins. The numbers of egg capsules and developing progeny produced under each experimental condition were determined. We found single worms were capable of producing egg capsules; however, in all instances the capsules were devoid of hatchlings, suggesting that the capsules deposited by these solitary individuals were probably the result of normal ovulation in the absence of productive fertilization (Fig. 1c). In contrast, mating pairs not only laid egg capsules, but also consistently produced progeny (Fig. 1c). We conclude that as in other flatworm species¹² the sexual biotype of *S. mediterranea* appears incapable of self-fertilization.

To study inbreeding and to simplify the analyses of marker segregation in *S. mediterranea*, we took advantage of the regenerative capacity of planarians to produce clones of one genotype via amputation that could be used for genetic crosses. Histological and molecular studies^{10,11,13,14} indicate that germ cells and somatic tissues of the reproductive system can regenerate after amputation. However, it was not clear whether fertility and fecundity are fully regenerated as well. We found that neither one, nor two rounds of amputation

diminished the capacity to produce egg capsules or hatchlings (Fig. 1d), indicating that regeneration results in the functional restoration of the entire hermaphroditic reproductive system. The ability to combine regeneration (asexual clonal expansion) with sexual reproduction allowed us to cross an animal to its genetically identical clone. We use ‘cross of clones’ to describe matings between such individuals.

Next, we exploited the cross of clones approach to carry out multiple rounds of inbreeding. To generate clonal individuals, we amputated a single worm from line S2, allowed the fragments to regenerate, reared the resulting animals to sexual maturity and finally crossed them to each other to produce the first filial generation (F_1 ; Fig. 2). A single F_1 progeny ($S2F_1$) was then selected for a new cycle of clonal crosses. To record potential diversity of recombination events, two siblings ($S2F_{4a}$ and $S2F_{4b}$) from the $S2F_4$ generation were intercrossed (Fig. 2, $S2F_4$) to produce two independent $S2F_5$ individuals, each of which was used to continue the cross of clones inbreeding for five more generations. We expected that at this stage (10 generations) both genomes of ($S2F_{5a}$) F_{10} and ($S2F_{6c}$) F_{10a} should be mostly homozygous. Owing to the prolonged longevity of *S. mediterranea*, all individuals from all ten generations remained alive, fertile and fecund. Therefore, the resulting inbreeding pedigree provided us with a living record of the landscape of changes that a single genome (S2) may have experienced through ten generations (for example, transposon mobilizations, mitotic and meiotic recombinations). More importantly, because one can readily cross an S2 individual with, for example, an $S2F_{10}$ individual, the living members of each generation afforded us the unique opportunity to do extensive multi- and intergenerational crosses to define the mechanisms driving genetic and epigenetic changes.

Having generated an inbreeding pedigree, we proceeded to identify bi-allelic single nucleotide polymorphisms (SNPs) with an in-house assembly of the *S. mediterranea* genome^{15,16} to follow the inheritance patterns of SNP loci across generations (Supplementary Fig. 2). RNA deep sequencing of an inbred individual ($S2F_{5b}$) F_{6b} allowed us to identify 3,234 heterozygous SNPs that could be used to test Mendelian inheritance (Fig. 3a). First, we analysed segregation and inheritance of SNPs in a pilot cross between an individual homozygous for one allele and an individual homozygous for the alternate allele (Fig. 3b). Analysis of genomic DNA by PCR and Sanger sequencing (Supplementary Fig. 2) revealed that the resulting progeny genotypes were 100% heterozygous, as predicted by Mendelian rules in a stable diploid organism with haploid gametes (Supplementary Fig. 3). In additional test crosses that monitored segregation of these four loci, we performed reciprocal crosses between heterozygous and homozygous animals and obtained the expected 1:1 ratio of A/a : A/A genotypes (Fig. 3b). Altogether, these results demonstrate that for some loci, independent assortment is observed and that all progeny arise as a result of cross-fertilization, ruling out parthenogenesis and further confirming cross-fertilization, rather than self-fertilization, as the predominant mode of reproduction in *S. mediterranea*.

The pilot cross also uncovered three heterozygous SNPs that failed to segregate in a Mendelian fashion. DNA genotyping of these SNPs in a cross of clones failed to produce the expected 1:2:1 segregation. Instead, heterozygosity for all three loci was retained in all resulting progeny (Supplementary Fig. 4). This unexpected result prompted us to systematically test inheritance patterns in planarians at a larger scale. We first examined the

zygosity of all 3,234 parental SNPs in each of the ten transcriptomes of the resulting progeny from the clonal cross of (S2F₅b)F₆b (Fig. 3a). We expected that, on average, the SNPs present in the heterozygous state in the (S2F₅b)F₆b parent would appear in the offspring in a 1A/A:2A/a:1a/a manner, as predicted by simple Mendelian inheritance patterns. Curiously, only 9.3% of the SNPs ($n = 300$) were homozygous in at least one progeny, while 99.8% of the SNPs ($n = 3,226$) were heterozygous in at least one progeny. When the distribution of the SNP zygosity was determined among the ten offspring, we observed that 20% of the 300 SNPs were homozygous in five out of ten progeny, while nearly 60% of the 3,226 SNPs ($n = 1,931$) were heterozygous in all ten progeny (Fig. 3c). This unexpected outcome is highly unlikely to be random ($P = 0$) and is a significant deviation from the Hardy–Weinberg principle, which predicts how gene frequencies are inherited from generation to generation. Mathematical simulations to account for random transmission of alleles recovered in ten offspring indicate that the curve that best fits the distribution of the 300 SNPs occurs when the two parental alleles assort with equal probabilities (that is, probability of being heterozygous (α) = 0.5, see Methods), the essence of Mendelian laws (Fig. 3c). In contrast, the curve best fitting the distribution of the 3,226 SNPs suggests strongly skewed segregation probabilities for the two parental alleles ($\alpha = 0.037$). These data indicate that a relatively large portion of the *S. mediterranea* genome was not segregating in a Mendelian fashion. For convenience, we refer to loci that followed Mendelian inheritance as MI and loci that maintained heterozygosity as non-MI.

As this initial set of SNPs was defined as heterozygous in an individual after six generations of inbreeding, it was possible that they reflected selection via balanced lethality or duplicated genomic regions, for example. To rule out these possibilities and to determine the genome-wide incidence and distribution of MI and non-MI loci across generations, we sequenced at 10× coverage the genomes of individuals from four different generations and counted the total numbers of bi-allelic heterozygous SNPs at each generation (Fig. 3d). Heterozygosity at the eighth generation of inbreeding would be expected to be nearly zero since¹⁷ $H_8 = (0.25) \times (0.5)^6 H_0$. Surprisingly, only around 30% of the heterozygous SNPs lost their heterozygosity before the sixth inbreeding generation. Around 70% of the SNPs remained heterozygous even at the eighth generation. Additional resequencing of the founder S2 lineage (that is, S2, S2F₂, (S2F₅a)F₆ and (S2F₅a)F₉) at 160× depth further confirmed this observation. To summarize, studies of the SNP markers suggested ~150 Mb of the genome followed Mendelian inheritance (~18.7%), ~300 Mb of the genome retained heterozygosity (~37.5%) and ~350 Mb of the genome was homozygous without SNP markers (~43.75%) in the S2 line.

It was possible that the continued maintenance of heterozygosity may be a unique trait of the S2 parental clones used to generate the lineage pedigree; therefore, we genotyped different individuals from our laboratory colony via whole genome sequencing and analysed the segregation of alleles in these animals (Fig. 4). We first tested the zygosity state of the S2 reference SNPs ($n = 335,340$) in four different inbred lines at different generations and three other sexual lines (that is, D2E, D5D and D5I) (Fig. 4a). To allow comparisons of SNPs between all tested lines and to increase the stringency of our analyses, we selected only those SNPs ($n = 31,319$) with sufficient high-quality sequencing reads across all tested lines. We further refined genotype calls by resequencing at 160× coverage and verified 30,992

genotype calls (~99%). We defined MI SNPs ($n = 17,088$) and non-MI SNPs ($n = 13,904$) from this rigorous list of SNPs on the basis of whether they lost or retained heterozygosity during inbreeding of the S2 line (Supplementary Tables 1 and 2). For D2E, ~36% of the S2 non-MI SNPs were homozygous. However, lines D5D and D5I had lost heterozygosity for almost 90% of the non-MI SNPs (12,400 out of 13,904). When fertility of these four lines was examined in clonal crosses, we noted that D2E was as fertile as S2F_{8a} (Fig. 4b, left panel), but that D5D and D5I were completely sterile; although they mated and laid large numbers of egg capsules (Fig. 4b, middle panel, examined eggs > 250); this suggests an association between sterility and loss of heterozygosity of the non-MI SNPs.

To further test the association of the observed loss of heterozygosity of the non-MI SNPs with decreased fertility, we crossed the homozygote non-MI SNP lines D5D and D5I to the heterozygote non-MI SNP line S2F_{8a}. In these crosses, not only was infertility rescued in D5D and D5I animals, but S2F_{8a} also gave birth to progeny (Fig. 4b, right panel), indicating that D5D and D5I have functional male and female gametes. Therefore, although D5D and D5I appeared to be sterile in self-crosses, the male and female germ lines of these animals were independently functional because outcrosses worked in both directions. We conclude from these experiments that the heterozygosity of the non-MI SNPs or associated chromosome region(s) may be important for fertility.

The ability to obtain progeny from both fertile (D2E) and non-fertile (D5D and D5I) lines provided us with independent opportunities to test whether the observed preservation of heterozygosity was unique to the S2 pedigree in clonal crosses. We selected S2 MI and non-MI SNPs to genotype the progeny from D2E clonal crosses, D2E/S2F_{8a} and D5D/S2F_{8a} non-clonal crosses (Fig. 4c–f) using SNPtype dynamic array assays (Methods, Supplementary Tables 3 and 4, Supplementary Fig. 2). In D2E clonal crosses, the MI SNPs A84 and A48 segregated in the expected 1:2:1 Mendelian ratio, but almost all progeny ($n = 30$) were heterozygous for the non-MI SNPs (Fig. 4c). Predominant non-Mendelian segregation of non-MI SNPs was also evident in the progeny of D2E/S2F_{8a} and D5D/S2F_{8a} non-clonal crosses (Fig. 4d–f); this was irrespective of whether the fertilization events were between S2F_{8a} oocytes and D2E or D5D spermatozooids, or between S2F_{8a} spermatozooids and D2E or D5D oocytes. The predominant transmission of the heterozygous genotype was even more pronounced when the homozygous non-MI SNPs in D5D were examined (Fig. 4e). Additionally, whole genome sequencing of three random offspring from D5I/S2F_{8a} crosses indicated that nearly 94% of the non-MI SNPs were heterozygous (Supplementary Fig. 7).

Because these lines have been reared in captivity for over 15 years, it is possible that the observed persistence of heterozygosity may have been artificially introduced during laboratory culture. To rule out this possibility, we obtained and genotyped 12 individuals from three geographically distant locations on the island of Sardinia (Fig. 5a,b) using SNPtype dynamic arrays containing MI and non-MI SNPs identified in S2. Genotypes from 82 SNPs (11 MI SNPs, 71 non-MI SNPs) showed great similarity among the animals of these three different populations (12 Sardinia worms, 39 Site04 F₁ and 5 S2 pedigree members). Diversity in these 12 wild animals came primarily from two MI SNPs and one non-MI SNP identified from the S2 inbreeding reference. In total, 39 out of 71 non-MI SNPs

were heterozygous in all individuals. Moreover, there was no homozygosity at these 39 non-MI SNPs in the F₁ progeny of Site04 worms (39 individuals), although we expected half of the F₁ generation to be homozygous if it were Mendelian inheritance (Fig. 5b). We conclude from these experiments that persistence of heterozygosity is occurring in wild type Sardinian populations of *S. mediterranea*. We postulate that these island populations may be highly inbred and may have evolved genetic mechanisms to maintain genome heterozygosity and purge inbreeding depression (Supplementary Information).

The observed persistence of heterozygosity could be a consequence of either pre- or post-zygotic selection, yet two of the most common mechanisms (genome duplications and non-disjunction) are unlikely to be responsible (Supplementary Information). We noted, however, that progeny from intercrosses between individuals of different genetic backgrounds (that is, S2F_{8a} and D2E; S2F_{8a} and D5D) were sometimes homozygous at the non-MI SNPs (Fig. 4d–f). This modification to the robust persistence of heterozygosity in clonal crosses (for example, S2 lineage, Sardinia worms and D2E) indicates potential recombination/mutation events in the life history of these lines. The ability to obtain homozygous progeny at the non-MI SNPs from these crosses provided us with a chance to examine segregation ratios of *A* and *a* alleles. Although infrequent, such homozygous F₁s (*A/A* or *a/a*) seemed to have similar ratios at different non-MI SNPs in multiple crossing experiments (Fig. 4c–f, Supplementary Fig. 9), suggesting that these non-MI SNP alleles may be closely linked or on the same haplotypes. Hence, to examine potential pre-zygotic mechanisms in haploid genomes, we sequenced the genomes of individual male and female gametes.

Whole-genome sequencing of individual spermatozooids ($n=11$) and oocytes ($n=15$) from the S2 line revealed a minute amount of heterozygosity varying from gamete to gamete, suggesting a basal level of gene duplication on different chromosomes, but not whole genome duplication (Fig. 6a,c,e,g). Surprisingly, we found both male and female gametes had lower diversity at the non-MI SNPs (Fig. 6a–d) compared with the MI-SNPs (Fig. 6e–h). Both male and female gametes could be clustered into two groups at the non-MI SNPs (Fig. 6a,b,e,f), representing the segregation of potentially two homologous chromosomes. At the MI-SNPs, gametes were more diverse and did not cluster into two groups (Fig. 6c,d,g,h). We named the two groups of allele combinations in male and female gametes J and V haplotypes. The J and V haplotypes reflect a tight linkage between the non-MI SNPs and low recombination rates.

Haplotypes with low recombination rates are usually maintained by chromosomal rearrangements (for example, inversions¹⁸ or translocations). Our analyses of the karyotypes and oocyte meiotic configurations (Supplementary Figs 1 and 3, and data not shown) rule out the classic multi-chromosome ring mechanism (that is, translocation heterozygosity) discovered in *Oenothera*¹⁹. However, three other potential molecular mechanisms remain, each of which involves two alleles present in the *trans* configuration on two haplotypes. In the first scenario, the two alleles lead to embryonic lethality when homozygous²⁰. In the second case, one such allele produces ‘toxins’ in the male gamete to suppress its own functionality or functionality of the gamete with the alternate allele, which leads to biased inheritance of one haplotype. The alternate allele produces toxins in the female gametes leading to a balanced meiotic drive system^{21,22}. In the third possibility, the two alleles

function during spermatozoid–oocyte recognition and fertilization, as has been shown for the self-incompatibility S-allele system in hermaphroditic plants²³ and the mating type loci in the green alga²⁴. We term this third mechanism haplotype incompatibility.

We found embryonic lethality highly unlikely to be responsible for the maintenance of genome heterozygosity (Supplementary Information). We then tested the balanced meiotic drive system hypothesis by examining haplotype segregation in crosses between S2F₈a and D5D/D5I lines. Analysis of the D5D/D5I diploid genome sequences indicated that they were not only homozygous in the haplotype defining non-MI SNPs, but also the J-haplotype alleles (Fig. 6a,e). Although both male and female J-haplotype gametes could produce progeny, J-haplotype oocytes from D5D/D5I and V-haplotype spermatozooids from S2F₈a produced ~50% less progeny compared with J-haplotype spermatozooids from D5D/D5I and V-haplotype oocytes from S2F₈a (Fig. 4b, right column). These data suggest that oocyte functions were impaired in the J haplotype, a form of meiotic drive. As the J haplotype did not completely abolish the functionality of female gametes, meiotic drive alone cannot explain the complete absence of homozygotes in the clonal crosses of lines heterozygous for J/V haplotypes.

To address the hypothesis of gamete haplotype incompatibility, we examined the mechanisms underlying the sterility of D5D/D5I in clonal crosses. Because D5D/D5I have functional male and female gametes but are sterile when they are crossed to their clones (Fig. 4b), the unhatched egg capsules from such crosses should not have arrested embryos if the hypothesis was valid. Indeed, sectioning then hematoxylin and eosin staining of these egg capsules indicated that all the putative zygotes ($n > 50$) were arrested at the single cell stage (Supplementary Fig. 3). Super-resolution structural illumination imaging of these cells showed that they were, in fact, unfertilized oocytes arrested at pre-metaphase II ($n = 8$, Supplementary Video 2). Hence, we conclude that J-haplotype male and female gametes in D5D/D5I lines rarely fertilized. Our results demonstrate that only gametes of different haplotypes could produce productive zygotes, suggesting the existence of a J/V-haplotype gamete recognition and fertilization programme that drives the heterozygosity maintenance in *S. mediterranea* populations.

In summary, data from the embryos of the S2 lines and infertility in the D5D/D5I lines suggest embryonic lethality is unlikely to be the driving factor or the sole factor in the maintenance of heterozygosity in *S. mediterranea* (Supplementary Information). Instead, the alleles driving the system probably control gamete recognition and fertilization. The fact that ~20% of MI SNP scaffolds co-localized with the non-MI SNP scaffolds (Supplementary Table 1) and that the J haplotype led to meiotic drive in oocytes, indicates that the heterozygosity maintaining complex in *S. mediterranea* may involve multiple mechanisms that are more complex than any one individual model. Mutations that impact multiple biological processes may have accumulated within the chromosome structural heterozygotes. Future identification of the molecular drivers in the maintenance of heterozygosity at the non-MI SNPs will clarify the nature of this balanced system in *S. mediterranea*.

Discussion

Why would such robust mechanisms to ensure heterozygosity persistence exist in planarians? We propose that for long-lived organisms with long-term inbreeding in closed environments, it is likely that mechanisms capable of mitigating loss of heterozygosity and maintaining high rates of genome heterozygosity exist; in planarians, we have uncovered evidence for such processes. Naturally occurring inversion heterozygotes, which have been shown to display superior fitness over homozygotic individuals in several Diptera species^{25,26}, may be such a mechanism. Additionally, the observed suppressed recombination²⁴, haplotype incompatibility and oocyte meiotic drive ultimately select gametes that are capable of yielding genome heterozygosity in planarian progeny and are reminiscent of the evolution of mating types or sex chromosomes^{24,27-29}.

Our findings demonstrate that balanced heterozygosity is essential for zygotic development (Fig. 4b), but not for somatic viability in *S. mediterranea*. Two general implications of our findings are that known adaptive advantages conferred by chromosome structural heterozygotes in other species (Supplementary Information) are probably evolutionarily conserved in the Lophotrochozoa and that such structural heterozygotes may be mechanistically associated with the robust regenerative capacity of planarians. Our study, therefore, firmly establishes *S. mediterranea* as a novel genetic system in which to molecularly dissect complex inheritance mechanisms.

Methods

Worm care, line maintenance and crosses

Worms were maintained in 1× Milli-Q standard planarian water at 20 °C, with constant twice or once a week feeding of organic liver paste. Different food colours were mixed with liver paste in cases when the two parental worms were of different genotypes in a cross and needed to be separated after mating. The most frequently used food colours were red, green or blue. Worms of different genotypes can be easily tracked with different colours three to five days post-feeding. To produce clones of one genotype, one worm was amputated into at least three fragments so that after regeneration each new individual became a virgin. To maintain a clonal line, eggs were taken out of the container if worms reached sexual maturity. If the worms had not reached sexual maturity, low-frequency feeding and water change kept them in constant numbers as pure lines. To amplify a clonal line, amputations were required. To perform experiments (for example, genotyping or fertility assessment), progeny were collected from at least five mating pairs per cross.

Next-generation sequencing and SNP calling

Genomic DNA was extracted from worms with Easy-DNA kit (Life technologies). RNA extraction was performed with TRIzol RNA Isolation Reagents (Life technologies). RNA-Seq libraries were constructed using approximately 1 µg of total RNA per sample and using the Illumina TruSeq RNA Sample Preparation Kits v2 (Cat. No. RS- RS-122-2001 and RS-122-2002). For the DNA-Seq libraries, 500 ng–1 Bioo Scientific µg of genomic DNA per sample was sonicated using a Covaris S220 instrument and libraries were constructed

using adapters from Bioo Scientific (Cat. No. 514104) and the High Throughput DNA kit from KAPA Biosystem (Cat. No. KK8234). All libraries were quantified using a combination of Bioanalyzer (Agilent Technologies), LabChip GX (Perkin Elmer) and a Qubit Fluorometer (Life Technologies). All libraries were pooled, re-quantified and run as either 50 bp single-end or 100 bp paired-end lanes on an Illumina HiSeq 2000/2500 instrument. HiSeq Control Software v.1.5.15.1–2.0.10.0 (Illumina, 2013), Real-Time Analysis (RTA) v.1.13.48.0–1.17.21.3 (Illumina, 2012) and CASAVA v.1.8.2 (Illumina) were used to process the runs, demultiplex reads and generate FASTQ files. Sequencing reads with deteriorating quality towards the 3'-end and 5'-end were trimmed with the programme Sickle v.1.33 (Joshi NA, Fass JN, 2011). Bowtie v.2.0.0 (Langmead B, Salzberg S, 2012) was used to align reads to the sexual genome assembly version 4.0 or asexual genome assembly version 1.1; the default mode of alignment was used with the idea that if there were reads from repetitive regions, the best aligned target from the assembly would be reported. SAMtools v.0.1.19 (Li H, 2011) and Picard tools v.1.96 (Broad Institute, 2013) were used to process the bam files. SAMtools mpileup/bcftools call and GATK v.2.7-4 (Broad Institute, 2013) HaplotypeCaller/ UnifiedGenotyper were used to call SNPs.

Gamete sequencing

Oocytes were identified from egg capsules laid by virgins after sectioning on the basis of nuclear morphology and cell size. Spermatozooids were identified from cell extractions of sexually mature worms on the basis of nuclear staining and flagella tails. Single gamete DNA was extracted and amplified as described³⁰. Genomic DNA sequencing and SNP calling were performed using the same methods as for the whole worms, described above.

Sequence coverage analysis

Genomes of four sexual lines in the S2 inbreeding pedigree (S2, S2F₂, S2F₆ and S2F₉) were sequenced to the depth of 160×. DNA sequencing reads were aligned to the reference genome v4.0 with Bowtie2. Duplicates were marked with Picard tools v1.96. InDels were realigned with GATK v3.3. Read coverage for the S2 heterozygous bi-allelic SNPs was then extracted from the alignments for both MI SNPs and non-MI SNPs. Distribution of the read coverage was examined with a box size of four, which provided the best visual effects while maintaining the information and high resolution. A sequencing read coverage between 0 and 500 for most of both the MI and non-MI SNPs in all four lines was achieved. Non-MI SNPs had a collection of discreet sites with sequencing read coverage in the range of 500 to 7,000. As distribution of these SNPs dramatically skews the scale of the rest of the SNPs, they were plotted separately and provided in Supplementary Fig. 5. Consistently, ~0.8% of non-MI SNPs behaved like this in all four lines. These non-MI SNPs should have been in repetitive regions or transposons.

Mendelian/non-Mendelian inheritance in cross of clones with ten progeny

If genes follow Mendelian inheritance, there are two mutually exclusive states of a gene: being heterozygous or homozygous. We defined '1' as being heterozygous and '0' as being homozygous. For each gene we defined a probability of being in state 1 as $0 < \alpha < 1$. For Mendelian inheritance $\alpha = 0.5$, as being heterozygous and being homozygous have equal chances when all alleles follow random segregation. We generated a random number

between zero and one for a given gene and if this number was larger than α we set the outcome to one, if smaller it was set to zero. We repeated this procedure for each gene several times, equal to the number of replicates (n). In our case, n was the number of progeny ($n = 10$). Then for each gene we computed the total number of '0's or '1's that ranged between zero and n . With $\alpha = 0.5$, we had a theoretical distribution for Mendelian inheritance in a function of the percentage of heterozygous or homozygous genes to the percentage of progeny. This is the Hardy–Weinberg Distribution curve.

In the resulting set we found the probability to observe the number of counts producing a numerically simulated distribution that could be compared with the experimental one. The goal of the numerical simulation was to find the value of α for which the difference between the numerical and experimental distributions was minimal (determined as a least square difference). We found that for MI SNPs the fitted value of probability was $\alpha = 0.5$, while for non-MI SNPs it was $\alpha = 0.037$. Although $\alpha = 0.5$ indicated MI SNPs followed a Mendelian inheritance distribution, $\alpha = 0.037$ did not really have a biological meaning, except that the non-MI SNPs strongly deviated from Mendelian inheritance.

To calculate the probability of n out of a total of m parental heterozygous loci being heterozygous in all ten progeny, the following command was used in R: `pbinom(n-1, m, prob=(0.5)10, lower.tail=F, log.p=T)`.

Definition of genotypes with allele ratios

With transcriptome sequencing data, an SNP was found to be homozygous if total sequencing read coverage was more than 50 and all reads were exactly the same at the SNP position with one allele (100% allele ratio). An SNP was found to be heterozygous if the sequencing read depth of each allele to the total of two alleles was between 30% and 70%, and the total sequencing read depths of two alleles were more than 20.

With genome sequencing data, an SNP was found to be homozygous if total sequencing read coverage was more than six and all reads were exactly the same at the SNP position with one allele (100% allele ratio). An SNP was found to be heterozygous if the sequencing read depth of each allele to the total of two alleles was between 30% and 70%, and total sequencing read depths of two alleles were more than six.

Gamete genotype correlation coefficient analysis

There were three possible genotyping outcomes at an SNP position, 'hom_ref'; 'hom_alt' and 'het', in the sequenced gametes. We used the binary coding settings (0 and 1) to numerically define the genotypes, as all SNPs genotyped were bi-allelic. Hence, the numerical genotype for 'het' was $(0 + 1)/2$, or 0.5. Correspondingly, the numerical genotype for 'hom_ref' was $(1 + 1)/2$, or 1, and that for 'hom_alt' was $(0 + 0)/2$, or 0. There were cases when the genotype could not be determined; these were defined as 'undetermined' or '-0.5'. With these definitions, the genotype of a gamete at all SNPs could be replaced by a string of numbers, which allowed computation of pairwise correlation coefficients. Pairwise correlation coefficients among the gametes were then clustered using the correlation distance as the distance function. The correlation values were organized into a square symmetric correlation matrix where an entry at the position (i,j) gave the value of the correlation

between i th and j th data sets. The results were colour-coded using the hue colour table for better visual presentation.

Genotyping

Besides genome sequencing, genotyping is frequently carried out with restriction fragment length polymorphism (RFLP) or Sanger sequencing of a PCR product. Candidate RFLP loci were selected from SNPs with polymorphic restriction enzyme digestion sites. Genotyping was also carried out with Dynamic Array integrated fluid circuits (Fluidigm) for thorough inspection of a large collection of SNPs in a large collection of animals (96×96).

Embryonic lethality

To examine embryonic lethality, egg capsules that did not hatch 30 d post deposition were fixed with 4% PFA and processed into 8 μ m thick paraffin sections. Cells on the paraffin sections were stained with hematoxylin and eosin stain. The nuclei of the blastomeres were stained dark blue. Arrested embryos had distinctive morphologies and distributions of the blastomeres, which were different from the normal developing embryos. It was also possible to detect arrested embryonic development events in egg capsules that were ~6–10 d post-deposition, due to their abnormal morphologies and distributions of the blastomeres.

Abnormal embryos in developing egg capsules

Developing egg capsules that were 7, 8, 9 or 11 d post-deposition were fixed for sectioning. These pre-hatching egg capsules should have contained both healthy embryos and potentially lethal embryos. The morphologies of the embryos were examined and categorized into three groups; they were considered ‘arrested’ if they were delayed in development (group 2) when compared with developed embryos in the same batch of capsules (group 1) or classified to have abnormal morphology (group 3). Total numbers of group 1 embryos and group 2/3 embryos were quantified for comparison.

Data availability

The data sets generated and/or analysed during the current study are available in the NCBI BioProject repository at <http://www.ncbi.nlm.nih.gov/bioproject/318690>. More data sets are available from the corresponding author on reasonable request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank D. Chao, R. Krumlauf, K. Golic, S. Hawley, T. Piotrowski, E. Jorgensen and D. Grunwald for comments, discussions and suggestions during the preparation of this manuscript. We thank H. Li, J. Vallandingham and M. Gogol for help with data analysis and visualization. We are grateful to M. Pala for the original gift in 1999 of the sexual specimens of *S. mediterranea* and the Stowers Institute Planarian Core facility for skilful maintenance of our planarian colony. We acknowledge S.M.C. Robb and P. Reddien for the initial establishment of the S2 line, A. Rossi for the discussions and S. Sánchez-Piotrowski for his help in specimen collection. This work was funded in part by the National Institutes of Health (NIH R37GM057260) to A.S.A.

References

1. Sánchez Alvarado A. Planarian regeneration: its end is its beginning. *Cell*. 2006; 124:241–245. [PubMed: 16439195]
2. Reddien PW, Sánchez Alvarado A. Fundamentals of planarian regeneration. *Annu Rev Cell Dev Biol*. 2004; 20:725–757. [PubMed: 15473858]
3. Newmark PA, Sánchez Alvarado A. Not your father's planarian: a classic model enters the era of functional genomics. *Nat Rev Genet*. 2002; 3:210–219. [PubMed: 11972158]
4. Pongratz N, Gerace L, Michiels NK. Genetic differentiation within and between populations of a hermaphroditic freshwater planarian. *Heredity*. 2002; 89:64–69. [PubMed: 12080371]
5. Pongratz N, Storhas M, Carranza S, Michiels NK. Phylogeography of competing sexual and parthenogenetic forms of a freshwater flatworm: patterns and explanations. *BMC Evol Biol*. 2003; 3:23. [PubMed: 14622447]
6. D'Souza TG, Storhas M, Schulenburg H, Beukeboom LW, Michiels NK. Occasional sex in an 'asexual' polyploid hermaphrodite. *Proc Biol Sci*. 2004; 271:1001–1007. [PubMed: 15293852]
7. D'Souza TG, Schulte RD, Schulenburg H, Michiels NK. Paternal inheritance in parthenogenetic forms of the planarian *Schmidtea polychroa*. *Heredity*. 2006; 97:97–101. [PubMed: 16721392]
8. D'Souza TG, Michiels NK. The costs and benefits of occasional sex: theoretical predictions and a case study. *J Hered*. 2010; 101(Suppl. 1):S34–S41. [PubMed: 20212007]
9. Dunn CW, Giribet G, Edgecombe GD, Hejnol A. Animal phylogeny and its evolutionary implications. *Annu Rev Ecol Evol S*. 2014; 45:371–395.
10. Wang Y, Zayas RM, Guo T, Newmark PA. Nanos function is essential for development and regeneration of planarian germ cells. *Proc Natl Acad Sci USA*. 2007; 104:5901–5906. [PubMed: 17376870]
11. Newmark PA, Wang Y, Chong T. Germ cell specification and regeneration in planarians. *Cold Spring Harb Symp Quant Biol*. 2008; 73:573–581. [PubMed: 19022767]
12. Hyman, LH. *The Invertebrates: Platyhelminthes and Rhynchocoela, the Acoelomate Bilateria*. Vol. II. McGraw-Hill; 1951.
13. Collins JJ III, et al. Genome-wide analyses reveal a role for peptide hormones in planarian germline development. *PLoS Biol*. 2010; 8:e1000509. [PubMed: 20967238]
14. Zayas RM, 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*. 2005; 102:18491–18496. [PubMed: 16344473]
15. Robb SM, Ross E, Sánchez Alvarado A. SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Res*. 2008; 36:D599–D606. [PubMed: 17881371]
16. Robb SM, Gotting K, Ross E, Sánchez Alvarado A. SmedGD 2.0: The *Schmidtea mediterranea* genome database. *Genesis*. 2015; 53:535–546. [PubMed: 26138588]
17. Nagylaki, T. *Introduction to Theoretical Population Genetics*. Springer; 1992.
18. Sturtevant AH, Beadle GW. The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. *Genetics*. 1936; 21:554–604. [PubMed: 17246812]
19. Golczyk H, Massouh A, Greiner S. Translocations of chromosome end-segments and facultative heterochromatin promote meiotic ring formation in evening primroses. *Plant Cell*. 2014; 26:1280–1293. [PubMed: 24681616]
20. Wallace H. The balanced lethal system of crested newts. *Heredity*. 1994; 73:41–46.
21. Zanders SE, et al. Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. *eLife*. 2014; 3:e02630. [PubMed: 24963140]
22. Charlesworth B. Selection for gamete lethals and S-alleles in complex heterozygotes. *Heredity*. 1979; 43:159–164.
23. Takayama S, Isogai A. Self-incompatibility in plants. *Ann Rev Plant Biol*. 2005; 56:467–489. [PubMed: 15862104]
24. Charlesworth B. Evolutionary genetics. The nature and origin of mating types. *Curr Biol*. 1994; 4:739–741. [PubMed: 7953566]

25. Levene H, Dobzhansky T. New evidence of heterosis in naturally occurring inversion heterozygotes in *Drosophila pseudoobscura*. *Heredity*. 1958; 12:37–49.
26. Dobzhansky T, Pavlovsky O. An extreme case of heterosis in a Central American population of *Drosophila tropicalis*. *Proc Natl Acad Sci USA*. 1955; 41:289–295. [PubMed: 16589666]
27. Charlesworth B. The evolution of sex chromosomes. *Science*. 1991; 251:1030–1033. [PubMed: 1998119]
28. De Hoff PL, et al. Species and population level molecular profiling reveals cryptic recombination and emergent asymmetry in the dimorphic mating locus of *C. reinhardtii*. *PLoS Genet*. 2013; 9:e1003724. [PubMed: 24009520]
29. Umen JG. Evolution of sex and mating loci: an expanded view from Volvocine algae. *Curr Opin Microbiol*. 2011; 14:634–641. [PubMed: 22035946]
30. Ottolini CS, et al. Genome-wide maps of recombination and chromosome segregation in human oocytes and embryos show selection for maternal recombination rates. *Nat Genet*. 2015; 47:727–735. [PubMed: 25985139]

Figure 1. Sexual reproduction in *S. mediterranea*

a, Worms mating. Scale bar = 1 mm. **b**, Their reproduction cycle. **c**, Production of gravid egg capsules; total eggs quantified were around 200 for one worm ($n = 10$) and around 50 for two worms ($n = 3$). Error bars are from three S2 line replicates. **d**, Effects of amputation on reproduction. One worm was amputated into three fragments first. Two were crossed after full regeneration as Cross I. The other fragment was amputated a second time into two fragments and crossed as Cross II. Error bars are from replicates in three different lines,

(S2F₅b)F_{6a}, (S2F₅b)F_{6b} and (S2F₅b)F_{6c}. There were no significant differences between Cross I and Cross II.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. Heterozygosity maintenance in the genomes of S2 and its lineages

a, Segregation analysis of SNPs. Transcriptome sequencing of (S2F₅b)F₆b clonal individuals was used to identify potential SNPs. The zygosity of parental SNPs in each one of the ten progeny of the clonal cross was deduced from the transcriptome of each progeny. **b**, Mendelian inheritance. Sanger sequencing of PCR amplicons of four loci (498, 90, 125 and 1056) that encode SNP candidates was used to define genotypes of three parental worms (for example, (S2F₅a)F₇; genotype: A/A) and the progeny produced by pairwise crosses of the parents. All alleles displayed a simple pattern of Mendelian inheritance and progeny

appeared to be products of cross-fertilization. The numbers of progeny genotyped were between 60 and 150 for each SNP. **c**, Distributions of MI (left) and non-MI (right) SNPs. Left: Distribution of MI SNPs from (S2F₅b)F₆b that are homozygous in at least one out of ten progeny (total SNPs = 300); this is consistent with the distribution of two alleles segregating independently into the offspring. The first data point deviates from the curve due to the high stringency used to identify homozygous SNPs in the progeny, which makes such SNPs more likely to be present in one animal instead of multiple animals. Right: Distribution of non-MI SNPs from (S2F₅b)F₆b that are heterozygous in at least one out of ten progeny (total SNPs = 3,226). The last data point (asterisk) shows SNPs (1,931 or 60%) that are heterozygous in all ten progeny. **d**, Minimal loss of whole genome heterozygosity after eight generations of inbreeding in S2 and its lineages (total SNPs in S2 = 335,340). The expected loss of whole genome heterozygosity is $H_{x+1} = 0.5H_x$ in clonal crosses and $H_{x+1} = 0.25H_x$ in sibling crosses (where x = generation).

Figure 4. Suppression of loss of heterozygosity during sexual reproduction and mitosis

a, Genotyping by whole genome sequencing in S2 inbreeding lineages (for example, S2F₂), other *S. mediterranea* sexual lines (for example, D2E, D5D) and the *S. mediterranea* asexual biotype (CIW4) at 30,992 heterozygous bi-allelic SNPs derived from S2. **b**, In the crosses of clones, S2, S2F_{8a} (S2F_{8a} = (S2F_{6c})F_{8a}) and D2E were fertile, but D5D and D5I were sterile. In non-clonal crosses with S2F_{8a}, D5D and D5I were fertile. **c–f**, Genotyping with an array of non-MI SNPs and MI SNPs. Maintenance of heterozygosity of the non-MI SNPs (red shaded bars) was observed in the progeny of clonal crosses of D2E (**c**), non-clonal

crosses between S2F_{8a} and D2E (**d**), and non-clonal crosses between S2F_{8a} and D5D (**e,f**), with simultaneous regular Mendelian inheritance of the MI SNPs (blue shaded bars). One allele of the non-MI SNPs from S2F_{8a} segregated predominantly into the progeny when crossed with D5D (**e**).

Figure 5. Persistence of genome heterozygosity in wild populations from Sardinia

a, Locations of the three collection sites on the island of Sardinia (>30 km apart). Scale bar: 50 km. **b**, Genotypes from 82 unambiguous SNPs (derived using the SNPtype assays on the 96 × 96 Dynamic Array IFC) are clustered and displayed for each line. Site08#1 is from collection site a, Site07#4 is from site b. Site04 lines are from site c ($n = 10$). Unlabelled lines represent F₁ progeny from the mixed Site04 population ($n = 39$).

Figure 6. J/V haplotypes and low recombination rates at non-MI SnPs in male/female gametes a–d. Genotyping by whole genome sequencing in 11 S2 spermatozooids (S1–S11) at 224 MI SNPs (c,d) and 585 non-MI SNPs (a,b). **e–h.** Genotyping in 15 S2 oocytes (O1–O15) at 2,443 MI SNPs (g,h) and 10,983 non-MI SNPs (e,f). Two diploid somatic samples (that is, D5D and D5I) were also included in the haplotype analysis (a,c,e,g). Eight spermatozooids (a) and six oocytes (e) had similar allele combinations (J haplotype) to D5D and D5I at the non-MI SNPs. Three spermatozooids (a) and nine oocytes (e) had complementary allele combinations (V haplotype). Clustering of correlations at the non-MI SNPs among the 11 spermatozooids (b) and 15 oocytes (f) revealed the highly similar gamete group and the gamete group of complementary haplotypes. Such clustering at the MI SNPs did not result in two distinct groups of allelic combinations of the gametes (d and h). Black bars in b,d,f and h indicate D5D and D5I.