



# Direct Gamete Sequencing Reveals No Evidence for Segregation Distortion in House Mouse Hybrids

## Citation

Corbett-Detig, Russell, Emily Jacobs-Palmer, Daniel Hartl, and Hopi Hoekstra. 2015. "Direct Gamete Sequencing Reveals No Evidence for Segregation Distortion in House Mouse Hybrids." PLoS ONE 10 (6): e0131933. doi:10.1371/journal.pone.0131933. <http://dx.doi.org/10.1371/journal.pone.0131933>.

## Published Version

[doi:10.1371/journal.pone.0131933](https://doi.org/10.1371/journal.pone.0131933)

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:17820662>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

RESEARCH ARTICLE

# Direct Gamete Sequencing Reveals No Evidence for Segregation Distortion in House Mouse Hybrids

Russell Corbett-Detig<sup>1\*</sup>, Emily Jacobs-Palmer<sup>1,2</sup>, Daniel Hartl<sup>1</sup>, Hopi Hoekstra<sup>1,2\*</sup>

**1** Department of Organismic & Evolutionary Biology, Harvard University, Cambridge, Massachusetts, United States of America, **2** Howard Hughes Medical Institute, Department of Molecular & Cellular Biology, Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, United States of America

✉ Current Address: Department of Integrative Biology, University of California, Berkeley, California, United States of America

\* [russcd@gmail.com](mailto:russcd@gmail.com) (RCD); [hoekstra@oeb.harvard.edu](mailto:hoekstra@oeb.harvard.edu) (HH)



**OPEN ACCESS**

**Citation:** Corbett-Detig R, Jacobs-Palmer E, Hartl D, Hoekstra H (2015) Direct Gamete Sequencing Reveals No Evidence for Segregation Distortion in House Mouse Hybrids. PLoS ONE 10(6): e0131933. doi:10.1371/journal.pone.0131933

**Editor:** William J. Murphy, Texas A&M University, UNITED STATES

**Received:** January 12, 2015

**Accepted:** June 8, 2015

**Published:** June 29, 2015

**Copyright:** © 2015 Corbett-Detig et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data have been deposited in the NCBI short read trace archive and can be accessed using the following accession: PRJNA280976.

**Funding:** This work was funded by the Howard Hughes Medical Institute. RCD was supported by a Harvard Prize Graduate Fellowship and currently a UCB Chancellor's Postdoctoral Fellowship; EJP was supported by an NSF Predoctoral Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

Understanding the molecular basis of species formation is an important goal in evolutionary genetics, and Dobzhansky-Muller incompatibilities are thought to be a common source of postzygotic reproductive isolation between closely related lineages. However, the evolutionary forces that lead to the accumulation of such incompatibilities between diverging taxa are poorly understood. Segregation distorters are believed to be an important source of Dobzhansky-Muller incompatibilities between hybridizing species of *Drosophila* as well as hybridizing crop plants, but it remains unclear if these selfish genetic elements contribute to reproductive isolation in other taxa. Here, we collected viable sperm from first-generation hybrid male progeny of *Mus musculus castaneus* and *M. m. domesticus*, two subspecies of rodent in the earliest stages of speciation. We then genotyped millions of single nucleotide polymorphisms in these gamete pools and tested for a skew in the frequency of parental alleles across the genome. We show that segregation distorters are not measurable contributors to observed infertility in these hybrid males, despite sufficient statistical power to detect even weak segregation distortion with our novel method. Thus, reduced hybrid male fertility in crosses between these nascent species is attributable to other evolutionary forces.

## Introduction

The Dobzhansky-Muller model [1,2] is widely accepted among evolutionary biologists as a primary explanation for the accumulation of intrinsic reproductive incompatibilities between diverging lineages [3,4]. Briefly, this model posits that genes operating normally in their native genetic background can be dysfunctional in a hybrid background due to epistatic interactions with alleles from a divergent lineage. Although elucidating the molecular basis of speciation has been a central focus for decades, loci contributing to Dobzhansky-Muller incompatibilities

**Competing Interests:** The authors have declared that no competing interests exist.

(DMIs) have proved challenging to identify primarily because they are, by definition, incompatible in combination (reviewed by [3–6]). As a result, the specific genetic changes responsible for the onset of reproductive isolation between lineages remain largely obscure.

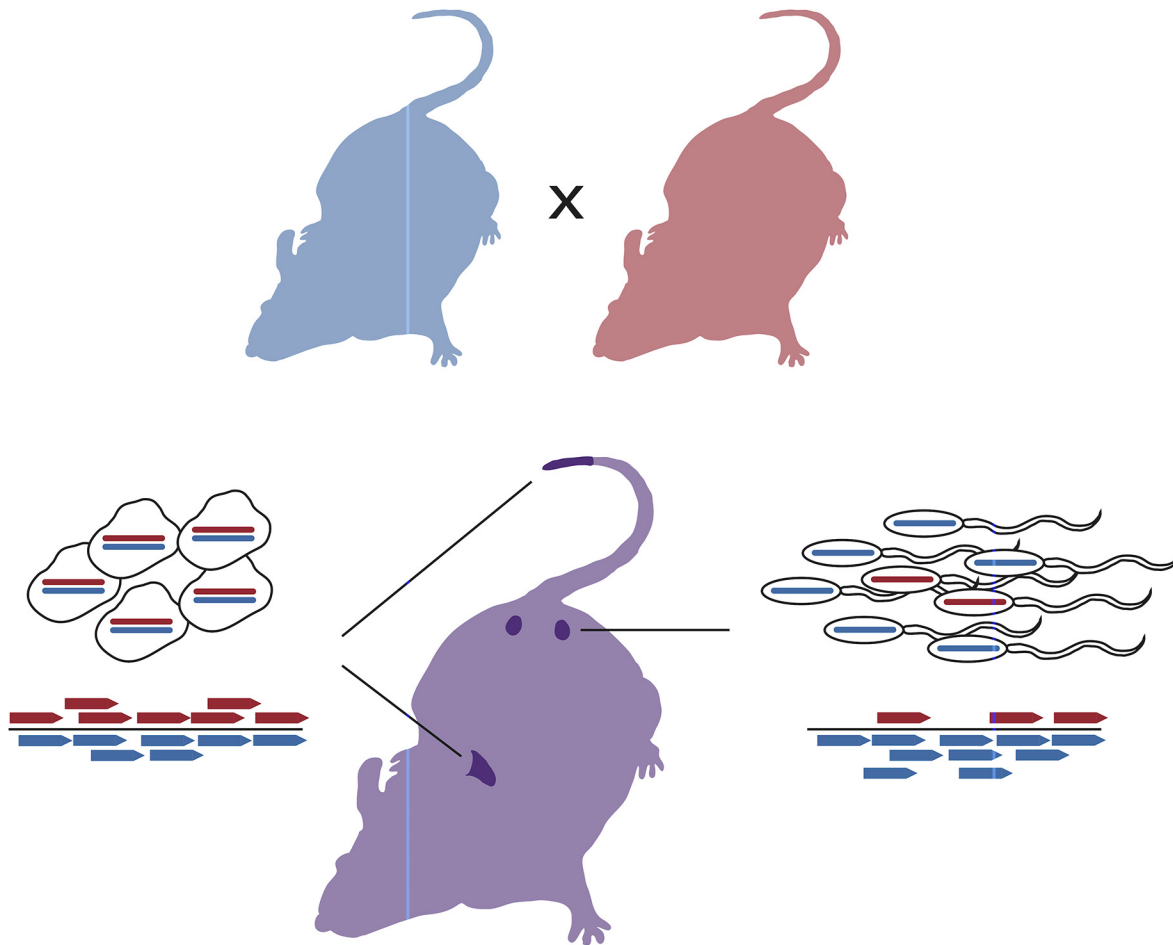
The rapid evolution of selfish genetic elements within lineages is thought to be a potent source of DMIs between diverging taxa. Segregation distorters (SDs) are one such selfish element, increasing their transmission through heterozygotes by either disabling or destroying gametes that failed to inherit the distorting allele [7,8]. Because males heterozygous for a distorter produce fewer viable sperm, SDs can decrease the fitness of carriers. In this case, other loci in the genome are expected to evolve to suppress distortion [9]. This coevolution of drivers and suppressors has been suggested to be a widespread source of DMIs between diverging lineages: hybrids of isolated populations in which such coevolutionary cycles have occurred may suffer lower fertility as drivers become uncoupled from their suppressors in a mixed genome [10–12]. Indeed, there is evidence that SDs contribute to hybrid male sterility in several *Drosophila* species pairs (e.g. [13–15], reviewed in [4,12]) as well as in many crop species (e.g. [16–20]). However, comparatively little is known about genetics of speciation outside of these groups, and a multitude of processes besides SD can contribute to the evolution of DMIs. Hence, it remains unclear if SDs contribute to hybrid sterility in other taxa more generally.

Analyses aimed at identifying the genetic targets of positive selection suggest that SDs may be an important source of DMIs in mammalian lineages. One particularly intriguing finding shows a substantial overrepresentation of loci associated with spermatogenesis and apoptosis within the set of genes with the strongest evidence for recurrent positive selection in mammals (e.g. [21,22]). These functions in turn are potentially driven, at least in part, by SDs, which are expected to leave just such a mark of selection as they sweep through a population. Therefore, mammals are an appealing group in which to test for SD and its role in speciation.

In particular, *Mus musculus domesticus* and *M. m. castaneus* are two subspecies of house mice in the earliest stages of evolving reproductive isolation [23,24]. Indeed, these subspecies are estimated to be approximately 500,000 years diverged from one another [24]. Hybrid males suffer from reproductive deficiencies [25]; specifically, the vas deferens of first-generation hybrid ( $F_1$ ) males contain more apoptotic sperm cells than either pure strain, and numerous loci affecting fertility in hybrid males have been reported, particularly in  $F_2$  individuals [26]. Finally, Wager [27] identified eight genomic regions that exhibited significant deviations from Mendelian segregation in an  $F_2$  mapping population derived from these two subspecies, which may be consistent with the action of SDs in their hybrids (but see below). In combination with the comparative genomic evidence and phenotypic observations described above, these data suggest that coevolution of SDs and their suppressors may contribute to DMIs in *M. musculus*.

The conventional approach to identifying SD relies on detecting a skew in the allele frequencies of second-generation hybrids in a large genetic cross. However, methods that rely on genotyping progeny unavoidably conflate SD, female effects on sperm function, and differential viability of offspring. Additionally, practical issues limit the power of these experiments—specifically, the ability to produce and genotype hundreds to thousands of offspring in order to detect distorters of small effect in an unbiased, genome-wide assay—particularly in vertebrates. Therefore, as a result of modest sample sizes, many experiments designed to detect SD using genetic crosses are underpowered and unable to detect even moderate distortion (e.g. a typical mammalian cross with a few hundred offspring will likely fail to identify 5–10% distortion). These practical issues impose challenges for speciation research generally, and for studies of SD specifically, by limiting our ability to confidently identify SDs within and between natural populations.

Here we explore a novel approach in which we survey the genome for SD by directly sequencing viable gametes from  $F_1$  hybrid *M. m. domesticus*/*M. m. castaneus* males, allowing



**Fig 1. Schematic of experimental cross scheme.** Inbred parental strains were crossed, and individual F1 males (purple) sacrificed at between 3 and 6 months, when their sperm were subjected to a swim-up assay. Libraries were prepared from liver or tail (control; left) and sperm (experiment; right) samples, sequenced, and then aligned to a diploid reference genome; subspecies of origin was determined for as many sequences as possible.

doi:10.1371/journal.pone.0131933.g001

us to circumvent the problems outlined above. Briefly, we enrich for viable sperm in hybrids and then sequence these sperm in bulk preparations, along with control somatic tissues, to identify any skew in the representation of either parental chromosome in the viable sperm relative to the control (Fig 1). While we demonstrate via simulation that our experimental design has excellent power, we find no evidence of SD in this cross, suggesting that SDs are not a primary contributor to male infertility in *M. m. castaneus* and *M. m. domesticus* hybrids. Nonetheless, this approach has a number of advantages relative to conventional methods. Specifically, our method is more cost effective, more specific to the identification of SD, and more generally applicable to a wide variety of organisms than conventional pedigree-based approaches. We therefore expect that it will be a useful means of studying the frequency and impact of SD in other systems.

## Methods

This work was approved by the Institutional Animal Care and Use Committee of Harvard University (protocol 27–23), and all mice were maintained following institutional protocols and

the Guide for the Care and Use of Laboratory Animals. Mice were euthanized by overdose of carbon dioxide.

## Reference Genome Assembly

To generate robust genome assemblies for each of the two strains of interest, we aligned all short read data for *M. m. castaneus* strain (CAST/EiJ) and *M. m. domesticus* strain (WSB/EiJ) from a recent large-scale resequencing project [28] to the MM9 genome assembly using BWA v0.7.1 [29] for initial mapping. For reads that failed to map with high confidence, we remapped using stampy v1.0.17 [30]. We realigned reads that overlap indels, and called SNPs and indels for each strain using the Genome Analysis Tool Kit (GATK, [31]). For each program, we used default parameters, except that during variant calling we used the option ‘--sample\_ploidy 1,’ because the strains are extremely inbred.

We generated a consensus sequence for each strain at sites where both assemblies have high quality data. That is, if both CAST and WSB assemblies had a q30 minimum quality genotype (either indels or SNPs) that site was added to both consensus sequences. Otherwise, if either or both assemblies were below this quality threshold at a given site, we used the MM9 reference allele for both.

## Alignment Simulation

Our goal was to align short read data to a single diploid reference genome, comprised of assemblies from the two parental strains. The mapping quality, which indicates the probability that a read is incorrectly mapped in the position indicated by the aligner, should then provide a reliable means of distinguishing whether a read can be confidently assigned to one of the parental genomes. To confirm the accuracy of this approach and to identify suitable quality thresholds, we performed simulations using SimSeq (<https://github.com/jstjohn/SimSeq>). We used the sequencing error profiles derived from our mapped data (below) and found qualitatively similar error rates using the default error profile included with the SimSeq software package (data not shown). For both the CAST and WSB genomes, we simulated 10,000,000 pairs of 94-bp paired-end reads, whose size distribution was set to match that of our libraries (below). We then mapped these reads back to the single reference genome containing both CAST and WSB consensus sequences. We scored reads as ‘mapping correctly’ if they mapped to within 10 bp of their expected location measured by their left-most coordinate and on the correct subspecies’ chromosome. If the pair mapped, we required that the insert length be less than 500 bp, which is within three standard deviations of the mean insert size of our data and should therefore encompass the vast majority of read pairs. If both reads in a pair mapped and met our criteria above, we used the higher mapping quality of the two, and discarded the other read. This filter is important, here and below, as it avoids counting pairs as though their provenance is independent of their pair.

## Experimental Crosses and Swim-Up Assay

To create first-generation ( $F_1$ ) hybrids of *Mus* subspecies, we crossed 2 *M. m. castaneus* males to 3 *M. m. domesticus* females and 2 *M. m. domesticus* males to 5 *M. m. castaneus* females in a harem-mating scheme. In total, we produced 8 male  $F_1$ s in each direction of the cross.  $F_1$  males whose sire was *M. m. castaneus* (CAST genome) are referred to as CW, and those whose sire was *M. m. domesticus* (WSB genome) as WC. All males were housed individually for a minimum of two weeks prior to sacrifice between 90 and 120 days of age.

To enrich for viable sperm from each  $F_1$  male, we performed a standard swim-up assay [32]. First, immediately following sacrifice, we collected and flash-froze liver and tail control tissues

(liver samples,  $N = 16$ ; tail samples  $N = 8$ ). Then, we removed and lacerated the epididymides of each male, placed this tissue in 1.5 ml of human tubal fluid (Embryomax HTF, Millipore), and maintained the sample at a constant 37°C for 10 minutes. Next, we isolated the supernatant, containing sperm that swam out of the epididymides, and spun this sample for 10 minutes at 250 g. We then discarded the supernatant, repeated the wash, and this time allowed sperm to swim up into the solution for an hour to select the most robust cells. Finally, we removed the solution, transferred them to new vial, pelleted these sperm by centrifugation, and froze them at -80°C.

## Library Preparation and Sequencing

For each  $F_1$  hybrid male, we first extracted DNA from sperm, liver, and tail tissues identically using a protocol designed to overcome the difficulty of releasing the tightly packed DNA within sperm nuclei (Qiagen *Purification of total DNA from animal sperm using the DNeasy Blood & Tissue Kit; protocol 2*). We sheared this DNA by sonication to a target insert size of 300 bp using a Covaris S220, then performed blunt-end repair, adenylation, and adapter ligation following the manufacturer protocol (New England BioLabs). Following ligation, libraries were pooled into two groups of 16 and one group of 8 based on the adapter barcodes. Prior to PCR, each pool was subject to automated size selection for 450–500 bp to account for the addition of 175 bp adapter sequences, using a Pippin Prep (Sage Science) on a 2.0% agarose gel cassette. PCR was performed using six amplification cycles, and then we re-ran the size selection protocol to eliminate adapter dimers prior to sequencing. Finally, we combined the three pools and sequenced them on two lanes of a HiSeq 2500. Each sequencing run consisted of 100 bp paired-end reads, of which the first 6 bp are the adapter barcode sequence, and the remaining 94 bp are derived from randomly-sheared gDNA.

## Alignment and Read Counting

We aligned read data to the combined reference genome using ‘BWA mem’ as described above in the alignment simulation. We removed potential PCR duplicates using Picard v1.73. We then filtered reads based on the alignment filtering criteria described above for the simulated data. Because copy number variations may pose problems for our analysis, we attempted to identify and exclude these regions. Specifically, we broke the genome into non-overlapping 10 kb windows. Then, within each library, we searched for 10 kb regions that had a sequencing depth greater than two standard deviations above the mean for that library. All aberrantly high-depth windows identified were excluded in downstream analyses in all libraries. These regions, representing approximately 7% of the windows in the genome, are reported in [S1 Table](#). We exclude high depth regions because translocated duplications that are linked to distant drive alleles could produce a spurious signature of drive in paralogous regions. Importantly, this is unlikely to impact our ability to detect SD that is affected by copy-number variable regions (e.g. [33, 34]) because linked single-copy regions will still display a signature of drive. Although deletions in one parental strain relative to the MM9 genome could also skew the parental allele frequencies for sequenced tissues, these copy number variable regions would affect both somatic tissues and gametes equivalently, and we therefore do not expect copy number variable regions to yield false positive results.

Next, to identify regions showing evidence of SD, we conducted windowed analyses with 1 Mb between the centers of adjacent windows. We counted reads in each window as a decreasing function of their distance from the center of the window, and included no reads at distances greater than 20 cM, thereby placing the most weight in a window on the center of the window. We then analyzed each window in two mixed-effects generalized linear models. Both models

included random effects for the libraries and individuals. The first model includes no additional factors. The second had fixed effects for tissue, direction of cross, and an interaction term based on tissue by direction of cross effects, and thus has five fewer degrees of freedom than the first model. Hence, for each window, we assessed the fit of the second model relative to the first using a likelihood ratio test, wherein the log likelihood ratio should be chi-square distributed with 5 degrees of freedom. Afterwards, we applied a false-discovery rate multiple testing correction to the data [35]. We performed all statistical analyses in R [36].

To investigate the potential contribution of X or Y linked distorters, we performed an analysis identical to that described above for autosomal windows, with one exception: we compared the number of reads that mapped to the X of each species with the proportion of total reads that mapped to the first chromosome of the corresponding individual. Here, the expectation in the absence of SD is that the X chromosome will have about half as many reads as an autosome (after normalizing for length). We performed a similar analysis contrasting reads mapping to the X versus the Y chromosome (and obtained similar results), but because the Y chromosome is largely repetitive, read mapping is unreliable, and we therefore prefer (and recommend) contrasting depth on the X and an autosomal chromosome.

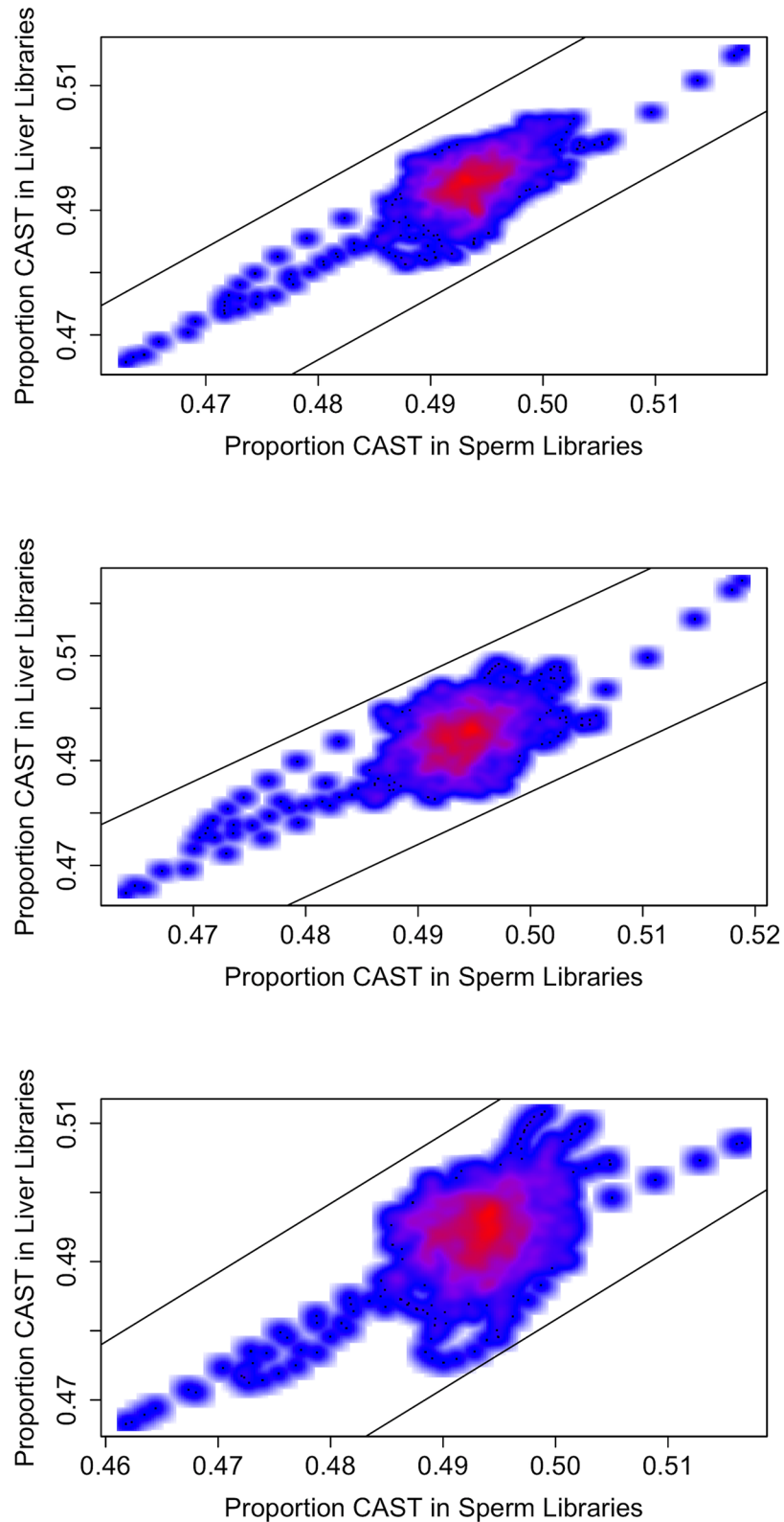
## Power Simulations

To estimate the power of our method, we simulated distortion data. We began by selecting sites randomly distributed across the genome, and for each site drew a distortion coefficient from a uniform distribution between -0.05 and 0.05. Each read on the parental genome that was susceptible to distortion was counted on the distorting genome with probability equal to the distortion coefficient multiplied by the probability that no recombination events occurred between the distorted locus and the read. We also did the alternative (*i.e.* switching reads from the distorted-against genome to the distorting genome) by multiplying by the probability that a recombination event was expected to occur in the genomic interval between the distorter and the read. We determined recombination probabilities using the genetic map reported in [37]. We performed the simulation for both parental genomes, and then again for each parental genome but with the distortion limited to one direction of the cross (*e.g.* only sperm from CW males experienced distortion). A direction-specific effect could occur if, for example, suppressing alleles are present on the Y chromosome of one subspecies and therefore are only present in CW or WC males.

## Results

After addressing the possibility of contamination, labeling, and quality issues (see [S1 Text](#), [S2 Table](#)), we ran our analysis of the data across all autosomes, excluding regions with evidence for copy-number variations (described in Methods). With the exception of windows on chromosome 16 (see below), we found no windows with a statistically significant signature of SD. The lowest uncorrected  $p$ -value for any window (aside from those on chromosome 16) was 0.0224, which is not significant when we corrected for multiple tests. Thus, we did not find evidence for SDs in any of the autosomal genomic regions considered, nor on the X or Y chromosomes ([Fig 2](#)).

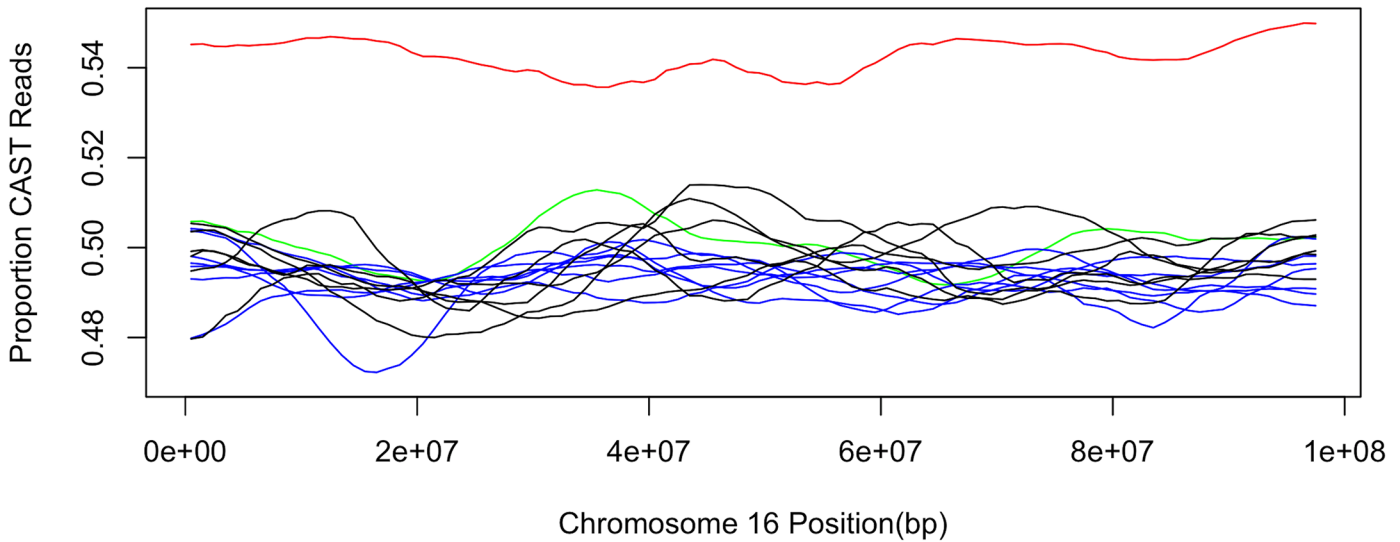
By contrast, on chromosome 16, we identified 15 contiguous windows with significantly skewed allele frequencies following correction for multiple comparisons (minimum  $p = 5.026E-4$ ; [Fig 3](#)). However, upon closer examination, it appears that this signal is driven entirely by a single liver sample, that of individual CW10. If this sample is removed from the dataset, this chromosome no longer shows significant deviation from expectations. When comparing the relative read depths across chromosomes 16 and 1, CW10's liver sample also



**Fig 2. Average proportion CAST reads in sperm libraries versus liver libraries.** Using all males (A), using only CW males (B), and using only WC males (C). Lines indicate the approximate threshold at which we would have 50% power to detect distortion at the  $\alpha = 0.0001$  level (see [Methods](#) for how this threshold was calculated).

doi:10.1371/journal.pone.0131933.g002





**Fig 3. Proportion of informative reads that are derived from the CAST genome across chromosome 16.** CW10's liver sample is shown in red, and CW10's sperm sample is shown in green. All other CW libraries are represented in black for liver and in blue for sperm.

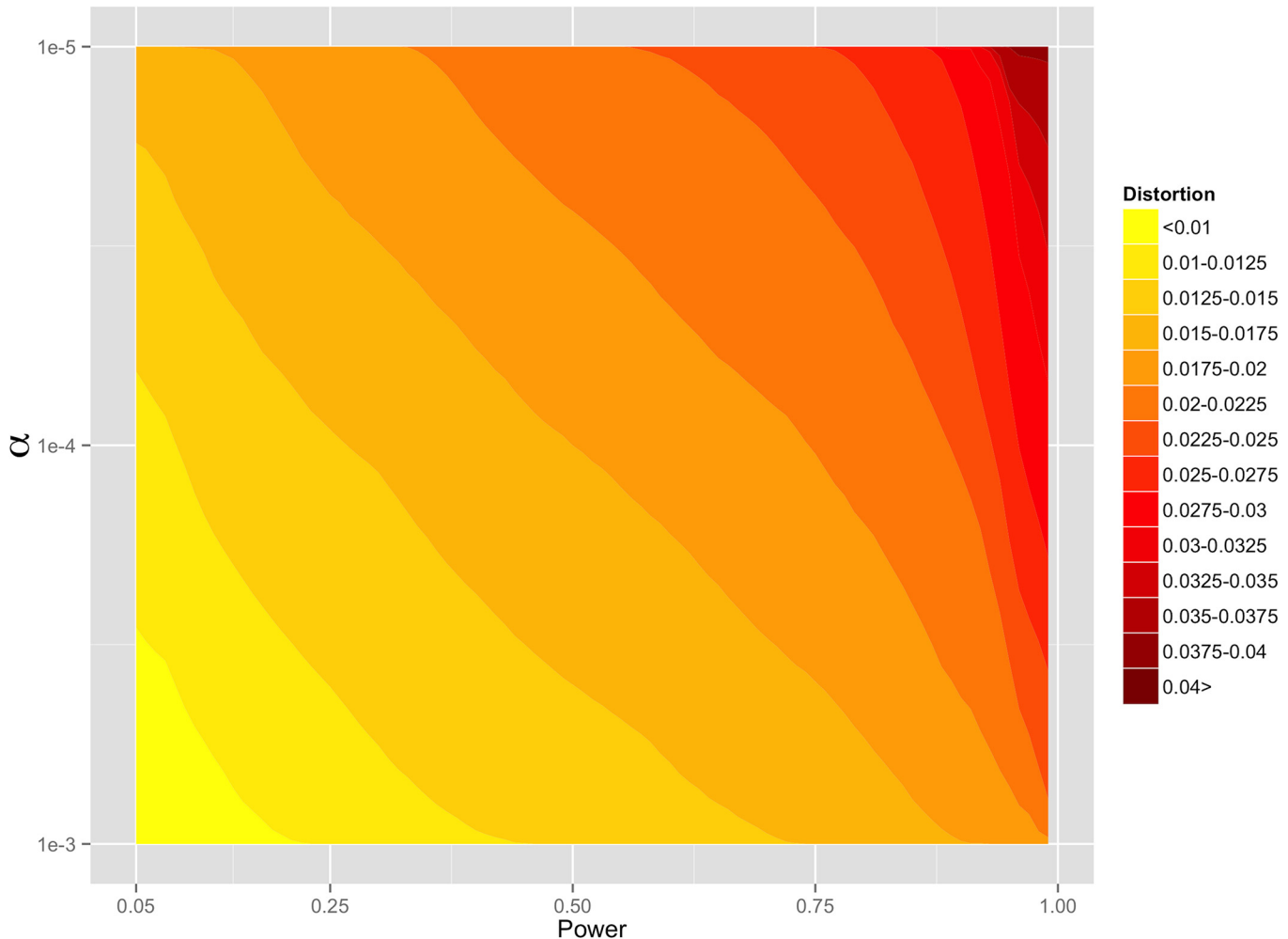
doi:10.1371/journal.pone.0131933.g003

appears to have disproportionately lower depth on this chromosome relative to ratio of depths in CW10's sperm sample ( $p = 3.02E-5$ ;  $X^2$ -test). These results suggest that this pattern is likely driven by a somatic aneuploidy event in CW10's liver that occurred relatively early in liver development and is not the result of distortion in the sperm sample.

Through simulation, we ensured that we have sufficient statistical power, given our experimental design and data quality, to detect SD if it is indeed occurring in hybrid males. We found that we have 50% power to detect SD to approximately 0.014, or 1.4% (this number reflects the positive or negative deviation from the null expectation, 0.5, at  $\alpha = 0.001$ ) if distortion affects CW and WC males equally (Fig 4). In other words, we have 50% power to detect distortion that is greater than 51.5% or less than 48.5%. If there is directionality to the distortion effect (*i.e.* only CW or only WC males experience SD), we have 50% power to detect distortion of 0.016 for CW males and 0.018 for WC males (at  $\alpha = 0.001$ ). This significance level was selected for illustrative purposes because it is approximately what would be required for genome-wide significance given our false discovery rate correction. The slight difference in power based on cross direction reflects differences in sequencing depth between WC and CW sperm and liver samples. It is also important to note that different regions of the genome will differ slightly in power to detect distortion because read mapping and sequencing, as well as divergence between the CAST and WSB strains and their divergence from the reference, are not uniform across the genome.

### Discussion

Elucidating the genetic mechanisms underlying species formation is a central goal of evolutionary biology. Although there has been progress in identifying the genetic basis of reproductive isolation in a few elegant instances (*e.g.* [38–40]), including several in *Drosophila* (*e.g.* [41,42]), it is unclear how general these results are. For example, in the case of SD specifically, we know that SDs contribute to reproductive isolation in several young *Drosophila* species pairs ([13–15]) but here, to our surprise, we find no evidence for SD between two nascent species of mouse, *M. m. castaneus*/*M. m. domesticus*, despite strong experimental power.



**Fig 4. Minimum level of distortion that could be detected given a specified significance threshold ( $\alpha$ , y-axis), and desired power (x-axis).**

doi:10.1371/journal.pone.0131933.g004

Our conclusion must be qualified to some degree. SDs are generally classified as either gamete disablers or gamete killers depending on their mode of action (reviewed in [7,8]). We expect to detect gamete killers with our approach since their “victims” may not be present in the epididymides, or, if present, would not be captured in our stringent swim-up assay. Our ability to detect gamete disablers, however, depends on the specific mechanism by which these genetic elements act. If the motility or longevity of a sperm cell is sufficiently impaired, it is likely that this sperm would fail to swim into solution and remain motile over the course of the assay, but if the distortion effect has a very subtle effect on motility or impairs function later in the sperm life cycle (e.g. by causing a premature acrosome reaction), it is unlikely that our method could detect these effects. Thus, although gamete killers are not prevalent sources of DMIs in these subspecies, we cannot completely exclude the possibility that gamete disablers contribute to *M. musculus* species formation. However, it is worth noting that disablers cannot explain the reported observation of increased apoptosis of sperm cells in hybrid males [26].

Conventional methods of detecting SDs (i.e. genotyping progeny of a cross) will conflate SD, gamete competition, and offspring viability, and due to practical limitations are usually statistically underpowered and thus unable to detect even modest distortion effects. For example,

one would need to genotype 3405 offspring to have 80% power to detect 5% distortion ( $\alpha = 0.001$ ), whereas using our method, we have 100% power to detect the same effect in the bulk-sequenced sperm of a single F<sub>1</sub> male. Moreover, requiring the presence of offspring from F<sub>1</sub> hybrids unavoidably conflates viability, gamete competition, and segregation distortion effects. By contrast, our simulations demonstrate that by sequencing high quality gametes from individual hybrid males and comparing allele ratios in these gametes to those of somatic tissues, we have excellent power to detect even relatively weak SDs, of approximately one percent. In support of this point, we successfully detected an aneuploidy event that resulted in a four percent difference in allele frequencies relative to expectations, within only a single biological replicate. Nonetheless, we found little evidence that SDs are active in F<sub>1</sub> hybrid males, which indicates that segregation distortion (specifically by gamete killers, though not necessarily by gamete disablers) is not a primary contributor to reduced F<sub>1</sub> male fertility in these subspecies.

Because our method of determining the allele ratios in bulk preparations of viable gametes relative to somatic tissues is very general, we expect that it will be useful in a wide variety of systems for an array of questions. Provided one can accurately phase the diploid genome of an individual, *e.g.* by using complete parental genotype data when inbred strains are not available, it is straightforward to apply this method to assay SD in a wide variety of taxa, including humans. Thus, we are now well positioned to survey the prevalence of segregation distortion both within and between a diversity of species. This approach also allows segregation distortion to be weighed against other possible sources of DMIs that may occur during spermatogenesis, oogenesis, fertilization, or embryogenesis, but which leave an identical signature to SD in conventional cross-based experiments. Furthermore, because SDs can increase in frequency in populations despite deleterious consequences for the host, these selfish genetic elements may also be an important source of disease alleles. For example, it has been suggested that SDs contribute to the perpetuation of split-hand/split-foot disease [43], retinal dystrophy [44] and Machado-Joseph disease [45] in humans. Hence, the method introduced here has the potential to improve our understanding of disease evolution in addition to the contribution of SDs to the evolution of reproductive isolation between diverging lineages. When applying this new method to outbred or natural populations, however, one must consider the possibility that SD alleles may be polymorphic (*e.g.* the early DMI model described in [46]). If SD alleles are at low frequency, it will be necessary to survey a large number of hybrid individuals to capture the loci of interest.

While segregation distorters may be an important mechanism of speciation in *Drosophila* and crop plants, efforts to detect SD in other diverging lineages—especially studies with high statistical power—have been limited. We find that at least in *M. m. castaneus*/*M. m. domesticus* hybrids, SDs—specifically gamete killers—are not measurable contributors to observed infertility in F<sub>1</sub> hybrid males, despite strong statistical power to detect them, suggesting that reduced hybrid male fertility in these nascent species is attributable to other underlying genetic causes. Further studies using the novel and powerful approach developed here will improve understanding of the role of SDs within and between populations.

## Supporting Information

**S1 Table. List of genomic windows excluded from all downstream analyses due to detection of individual libraries with unusually high depth.**

(TXT)

**S2 Table. Quality control results for the quantity of reads in each library derived from the Y chromosome, X chromosome, and mtDNA.**

(TXT)

**S3 Table. Alignment simulation results showing the relationship between the reported mapping quality for a read and its probability of correct assignment to the genomic location from which it was derived.**

(TXT)

**S1 Text. Supplemental methods describing quality control steps to ensure samples are not contaminated or mislabeled.**

(TXT)

## Acknowledgments

The authors declare no conflicts of interest. We thank Tim Sackton, Julien Ayroles, Brian Arnold, Shelbi Russell, Brant Peterson, Marjorie Lienard, and Heidi Fisher for helpful comments and suggestions, and Catherine Dulac and Stacey Sullivan for assistance with mouse crosses. RCD was supported by a Harvard Prize Graduate Fellowship and currently a UCB Chancellor's Postdoctoral Fellowship; EJP was supported by an NSF Predoctoral Fellowship.

## Author Contributions

Conceived and designed the experiments: RCD EJP DH HH. Performed the experiments: RCD EJP. Analyzed the data: RCD. Contributed reagents/materials/analysis tools: HH. Wrote the paper: RCD EJP HH.

## References

1. Dobzhansky T (1937) *Genetics and the Origin of Species*. Columbia University Press. Columbia, NY.
2. Muller HJ (1942) Isolating mechanisms, evolution and temperature. *Biol. Symp.* 6: 71–125.
3. Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates. Sunderland, MA.
4. Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nat Rev Genet* 11: 175–180. doi: [10.1038/nrg2718](https://doi.org/10.1038/nrg2718) PMID: [20051985](https://pubmed.ncbi.nlm.nih.gov/20051985/)
5. Wu C-I, Ting C-T (2004) Genes and speciation. *Nat Rev Genet* 5: 114–122. doi: [10.1038/nrg1269](https://doi.org/10.1038/nrg1269) PMID: [14735122](https://pubmed.ncbi.nlm.nih.gov/14735122/)
6. Noor MAF, Feder JL (2006) Speciation genetics: evolving approaches. *Nat Rev Genet* 7: 851–861. doi: [10.1038/nrg1968](https://doi.org/10.1038/nrg1968) PMID: [17033626](https://pubmed.ncbi.nlm.nih.gov/17033626/)
7. Lyttle TW (1991) Segregation distorters. *Annual review of genetics* 25: 511–581. doi: [10.1146/annurev.ge.25.120191.002455](https://doi.org/10.1146/annurev.ge.25.120191.002455) PMID: [1812815](https://pubmed.ncbi.nlm.nih.gov/1812815/)
8. Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. *Genetica* 117: 27–35. doi: [10.1023/A:1022308414864](https://doi.org/10.1023/A:1022308414864) PMID: [12656570](https://pubmed.ncbi.nlm.nih.gov/12656570/)
9. Hartl DL (1975) Modifier theory and meiotic drive. *Theoretical population biology* 7: 168–174. doi: [10.1016/0040-5809\(75\)90012-X](https://doi.org/10.1016/0040-5809(75)90012-X) PMID: [1145501](https://pubmed.ncbi.nlm.nih.gov/1145501/)
10. Frank SA (1991) Divergence of meiotic drive-suppression systems as an explanation for sex-biased hybrid sterility and inviability. *Evolution*: 262–267.
11. Hurst LD, Pomiankowski A (1991) Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* 128: 841–858. PMID: [1916248](https://pubmed.ncbi.nlm.nih.gov/1916248/)
12. McDermott SR, Noor MAF (2010) The role of meiotic drive in hybrid male sterility. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 1265–1272. doi: [10.1098/rstb.2009.0264](https://doi.org/10.1098/rstb.2009.0264)
13. Tao Y, Masly JP, Araripe L, Ke Y, Hartl DL (2007) A sex-ratio Meiotic Drive System in *Drosophila simulans*. I: An Autosomal Suppressor. *PLoS Biol* 5: e292. doi: [10.1371/journal.pbio.0050292](https://doi.org/10.1371/journal.pbio.0050292) PMID: [17988172](https://pubmed.ncbi.nlm.nih.gov/17988172/)
14. Tao Y, Araripe L, Kingan SB, Ke Y, Xiao H, Hartl DL (2007) A sex-ratio meiotic drive system in *Drosophila simulans*. II: an X-linked distorter. *PLoS Biol*. doi: [10.1371/journal.pbio.0050293.g001](https://doi.org/10.1371/journal.pbio.0050293.g001)
15. Phadnis N, Orr HA (2009) A Single Gene Causes Both Male Sterility and Segregation Distortion in *Drosophila Hybrids*. *Science* 323: 376–379. doi: [10.1126/science.1163934](https://doi.org/10.1126/science.1163934) PMID: [19074311](https://pubmed.ncbi.nlm.nih.gov/19074311/)

16. Bohn GW, Tucker CM (1940) Studies on Fusarium wilt of the Tomato. I. Immunity in *Lycopersicon pimpinellifolium*, and its inheritance in hybrids. Research Bulletin Missouri Agricultural Experiment Station 311.
17. Cameron DR, Moav RM (1957) Inheritance in *Nicotiana tabacum* Xxvii. pollen killer, an alien genetic locus inducing abortion of microspores not carrying it. Genetics 42: 326. PMID: [17247701](#)
18. Loegering WQ, Sears ER (1963) Distorted inheritance of stem-rust resistance of Timstein wheat caused by a pollen-killing gene. Canadian Journal of Genetics and Cytology 5: 65–72.
19. Sano Y, Chu Y-E, Oka H-I (1979) Genetic studies of speciation in cultivated rice. 1. Genic analysis for the F1 sterility between *Oryza sativa* L. and *Oryza glaberrima* Steud. Japanese journal of genetics 54: 121–132.
20. Yang J, Zhao X, Cheng K, Du H, Ouyang Y, Chen J, et al. (2012) A killer-protector system regulates both hybrid sterility and segregation distortion in rice. Science. doi: [10.1126/science.1222597](#)
21. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz M, et al. (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3: e170. doi: [10.1371/journal.pbio.0030170](#) PMID: [15869325](#)
22. Kosiol C, Vinař T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, et al. (2008) Patterns of positive selection in six mammalian genomes. PLoS Genetics 4: e1000144. doi: [10.1371/journal.pgen.1000144](#) PMID: [18670650](#)
23. Boursot P, Din W, Anand R, Darviche D, Dod B, Von Deimling F, et al. (1996) Origin and radiation of the house mouse: mitochondrial DNA phylogeny. Journal of Evolutionary Biology 9: 391–415. doi: [10.1046/j.1420-9101.1996.9040391.x](#)
24. Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu HT, et al. (2008) Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. Molecular Ecology 17: 5349–5363. doi: [10.1111/j.1365-294X.2008.04005.x](#) PMID: [19121002](#)
25. Davis RC, Jin A, Rosales M, Yu S, Xia X, Ranola K, et al. (2007) A genome-wide set of congenic mouse strains derived from CAST/Ei on a C57BL/6 background. Genomics 90: 306–313. doi: [10.1016/j.ygeno.2007.05.009](#) PMID: [17600671](#)
26. White MA, Stubbings M, Dumont BL, Payseur BA (2012) Genetics and evolution of hybrid male sterility in house mice. Genetics 191: 917–934. doi: [10.1534/genetics.112.140251](#) PMID: [22554891](#)
27. Wagner J (2010) Transmission ratio distortion in the hybrid offspring of three mouse subspecies. BS Thesis, University of Wisconsin, Madison. Available: [https://minds.wisconsin.edu/bitstream/handle/1793/46044/2010\\_Wagner.pdf?sequence=1](https://minds.wisconsin.edu/bitstream/handle/1793/46044/2010_Wagner.pdf?sequence=1).
28. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, et al. (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477: 289–294. doi: [10.1038/nature10413](#) PMID: [21921910](#)
29. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760. doi: [10.1093/bioinformatics/btp324](#) PMID: [19451168](#)
30. Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Research 21: 936–939. doi: [10.1101/gr.111120.110](#) PMID: [20980556](#)
31. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491–498. doi: [10.1038/ng.806](#) PMID: [21478889](#)
32. Holt WV, Hernandez M, Warrell L, Satake N (2010) The long and the short of sperm selection in vitro and in vivo: swim-up techniques select for the longer and faster swimming mammalian sperm. Journal of Evolutionary Biology 23: 598–608. doi: [10.1111/j.1420-9101.2010.01935.x](#) PMID: [20487132](#)
33. Cocquet J, Ellis PJ, Mahadevaiah SK, Affara NA, Vairan D, Burgoyne PS (2012) A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the mouse. PLoS Genet 8: e1002900. doi: [10.1371/journal.pgen.1002900](#) PMID: [23028340](#)
34. Merrill C, Bayraktaroglu L, Kusano A, Ganetzky B (1999) Truncated RanGAP encoded by the *Segregation Distorter* locus of *Drosophila*. Science 283: 1742–1745. PMID: [10073941](#)
35. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B 57: 289–300.
36. Team RDC (n.d.) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria ISBN.
37. Cox A, Ackert-Bicknell CL, Dumont BL, Ding Y, Bell JT, Brockman GA, et al. (2009) A new standard genetic map for the laboratory mouse. Genetics 182: 1335–1344. doi: [10.1534/genetics.109.105486](#) PMID: [19535546](#)

38. Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*. doi: [10.1007/s00294-003](https://doi.org/10.1007/s00294-003)
39. Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323: 373–375. doi: [10.1126/science.1163601](https://doi.org/10.1126/science.1163601) PMID: [19074312](https://pubmed.ncbi.nlm.nih.gov/19074312/)
40. Lassance J-M, Groot AT, Liénard MA, Antony B, Borgwardt C, Andersson F, et al. (2010) Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature* 466: 486–489. doi: [10.1038/nature09058](https://doi.org/10.1038/nature09058) PMID: [20592730](https://pubmed.ncbi.nlm.nih.gov/20592730/)
41. Ting C-T, Tsaur S-C, Wu M-L, Wu C-I (1998) A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* 282: 1501–1504. PMID: [9822383](https://pubmed.ncbi.nlm.nih.gov/9822383/)
42. Masly JP, Jones CD, Noor MAF, Locke J, Orr HA (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science* 313: 1448–1450. doi: [10.1126/science.1128721](https://doi.org/10.1126/science.1128721) PMID: [16960009](https://pubmed.ncbi.nlm.nih.gov/16960009/)
43. Jarvik GP, Patton MA, Homfray T, Evans JP (1994) Non-Mendelian transmission in a human developmental disorder: split hand/split foot. *American Journal of Human Genetics* 55: 710. PMID: [7942849](https://pubmed.ncbi.nlm.nih.gov/7942849/)
44. Evans K, Fryer A, Duvall-Young J, Inglehearn C, Whittaker J, Gregory CY, et al. (1994) Genetic linkage of cone-rod retinal dystrophy to chromosome 19q and evidence for segregation distortion. *Nature* 6: 210–213. doi: [10.1038/ng0294-210](https://doi.org/10.1038/ng0294-210)
45. Ikeuchi T, Igarashi S, Takiyama Y, Onodera O, Oyake M, Takano H, et al. (1996) Non-Mendelian transmission in dentatorubral-pallidoluysian atrophy and Machado-Joseph disease: the mutant allele is preferentially transmitted in male meiosis. *American Journal of Human Genetics* 58: 730. PMID: [8644735](https://pubmed.ncbi.nlm.nih.gov/8644735/)
46. Cutter AD (2012) The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. *Trends in Ecology & Evolution* 27: 209–218. doi: [10.1016/j.tree.2011.11.004](https://doi.org/10.1016/j.tree.2011.11.004)