bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Title:** Direct gamete sequencing reveals no evidence for segregation distortion in house
2 mouse hybrids
3
4 **Authors:** Russell Corbett-Detig[1,2*], Emily Jacobs-Palmer[1,3], Daniel Hartl[1], and Hopi
5 Hoekstra[1,3*]
6
7 **Author Affiliations:**
8 1. Department of Organismic & Evolutionary Biology, Harvard University, Cambridge, MA
9 USA.
10 2. Present Address: Department of Integrative Biology, University of California, Berkeley,
11 CA USA.
12 3. Howard Hughes Medical Institute, Department of Molecular & Cellular Biology, Museum
13 of Comparative Zoology, Harvard University, Cambridge, MA USA.
14 * To whom correspondence should be addressed: russcd@gmail.com,
15 hoekstra@oeb.harvard.edu

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

16 **Abstract:**

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

33

34 **Introduction**:

35 The Dobzhansky-Muller model [1,2] is widely accepted among evolutionary
36 biologists as a primary explanation for the accumulation of intrinsic reproductive
37 incompatibilities between diverging lineages [3,4]. Briefly, this model posits that genes
38 operating normally in their native genetic background can be dysfunctional in a hybrid

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

39 background due to epistatic interactions with alleles from a divergent lineage. Although
40 elucidating the molecular basis of speciation has been a central focus for decades, loci
41 contributing to Dobzhansky-Muller incompatibilities (DMIs) have proved challenging to
42 identify primarily because they are, by definition, incompatible in combination (review by
43 [3-6]). As a result, the specific genetic changes responsible for the onset of reproductive
44 isolation between lineages remain largely obscure.

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

60 Analyses aimed at identifying the genetic targets of positive selection suggest that
61 segregation distorters may be an important source of DMIs in mammalian lineages. One

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

62 particularly intriguing finding shows a substantial overrepresentation of loci associated
63 with spermatogenesis and apoptosis within the set of genes with the strongest evidence for
64 recurrent positive selection in mammals (*e.g.* [21,22]). These functions in turn are
65 potentially driven, at least in part, by segregation distorters, which are expected to leave
66 just such a mark of selection as they sweep through a population. Therefore, mammals are
67 an appealing group in which to test for segregation distortion and its role in speciation.
68 In particular, *Mus musculus domesticus* and *M. m. castaneus* are two subspecies of
69 house mice in the earliest stages of evolving reproductive isolation [23,24]. Indeed, these
70 subspecies are estimated to be approximately 500,000 years diverged from one another
71 [24]. Hybrid males suffer from reproductive deficiencies [25]; specifically, the vas deferens
72 of first-generation hybrid (F_1) males contain more apoptotic sperm cells than either pure
73 strain, and numerous loci affecting fertility in hybrid males have been reported,
74 particularly in F_2 individuals [26]. Finally, Wager [27] identified eight genomic regions that
75 exhibited significant deviations from Mendelian segregation in an F_2 mapping population
76 derived from these two subspecies, which may be consistent with the action of segregation
77 distorters in their hybrids (but see below). In combination with the comparative genomic
78 evidence and phenotypic observations described above, these data suggest that
79 coevolution of SDs and their suppressors may contribute to DMIs in *M. musculus*.
80 The conventional approach to identifying SD relies on detecting a skew in the allele
81 frequencies of second-generation hybrids in a large genetic cross. However, methods that
82 rely on genotyping progeny unavoidably conflate segregation distortion, female effects on
83 sperm function, and differential viability. Additionally, practical issues limit the power of
84 these experiments—specifically, the ability to produce and genotype hundreds to

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

85 thousands of individuals in order to detect distorters of small effect—particularly in
86 vertebrates. Therefore, as a result of modest sample sizes, many experiments designed to
87 detect SD using genetic crosses are underpowered and unable to detect even moderate
88 distortion.

89 Here, we explore a novel approach to surveying the genome for SD by directly
90 sequencing viable gametes from F_1 hybrid *M. m. domesticus*/*M. m. castaneus* males. Briefly,
91 we enriched for viable sperm in hybrids and then sequenced these sperm in bulk, along
92 with a control tissue, to identify any skew in the representation of either parental
93 chromosome in the viable sperm relative to the control (Figure 1). While we demonstrate
94 via simulation that our experimental design has excellent power, we find no evidence of SD
95 in this cross, suggesting that segregation distorters are not a primary contributor to male
96 infertility in *M. m. castaneus* and *M. m. domesticus* hybrids. Nonetheless, this approach can
97 be applied to a wide range of species, and we therefore expect that it will be a useful means
98 to study the frequency and impact of segregation distortion more generally.
99

100 **Materials and Methods**:

101 <u>Reference Genome Assembly:</u> To generate robust genome assemblies for each of the two
102 strains of interest, we aligned all short read data for *M. m. castaneus* strain (CAST/EiJ) and
103 *M. m. domesticus* strain (WSB/EiJ) from a recent large-scale resequencing project [28] to
104 the MM9 genome assembly using BWA v0.7.1 [29] for initial mapping. For reads that failed
105 to map with high confidence, we remapped using stampy v1.0.17 [30]. We realigned reads
106 that overlap indels, and called SNPs and indels for each strain using the Genome Analysis
107 Tool Kit (GATK, [31]). For each program, we used default parameters, except that during

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

108 variant calling we used the option '--sample_ploidy 1,' because the strains are extremely
109 inbred.
110 We generated a consensus sequence for each strain at sites where both assemblies
111 have high quality data. That is, if both CAST and WSB assemblies had a q30 minimum
112 quality genotype (either indels or SNPs) that site was added to both consensus sequences.
113 Otherwise, if either or both assemblies were below this quality threshold at a given site, we
114 used the MM9 reference allele for both.
115
116 Alignment Simulation: Our goal was to align short read data to a single diploid reference
117 genome, comprised of assemblies from the two parental strains. The mapping quality,
118 which indicates the probability that a read is incorrectly mapped in the position indicated
119 by the aligner, should then provide a reliable means of distinguishing whether a read can
120 be confidently assigned to one of the parental genomes. To confirm the accuracy of this
121 approach and to identify suitable quality thresholds, we performed simulations using
122 SimSeq (https://github.com/jstjohn/SimSeq). We used the sequencing error profiles
123 derived from our mapped data (below) and found qualitatively similar error rates using the
124 default error profile included with the SimSeq software package (data not shown). For both
125 the CAST and WSB genomes, we simulated 10,000,000 pairs of 94-bp paired-end reads,
126 whose size distribution was set to match that of our libraries (below). We then mapped
127 these reads back to the single reference genome containing both CAST and WSB consensus
128 sequences. We scored reads as 'mapping correctly' if they mapped to within 10 bp of their
129 expected location measured by their left-most coordinate and on the correct subspecies'
130 chromosome. If the pair mapped, we required that the insert length be less than 500 bp,

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

131 which is well within three standard deviations of the mean insert size of our data and
132 should therefore encompass the vast majority of read pairs. If both reads in a pair mapped
133 and met our criteria above, we used the higher mapping quality of the two, and discarded
134 the other read. This filter is important, here and below, as it avoids counting pairs as
135 though their provenance is independent of their pair.
136
137 Experimental Crosses and Swim-Up Assay: To create first-generation (F_1) hybrids of *Mus*
138 subspecies, we crossed 2 *M. m. castaneus* males to 3 *M. m. domesticus* females and 2 *M. m.*
139 *domesticus* males to 5 *M. m. castaneus* females in a harem-mating scheme. In total, we
140 produced 8 male F_1s in each direction of the cross. F_1 males whose sire was *M. m. castaneus*
141 (CAST genome) are referred to as CW, and those whose sire was *M. m. domesticus* (WSB
142 genome) as WC. All males were housed individually for a minimum of two weeks prior to
143 sacrifice between 90 and 120 days of age.
144 To enrich for viable sperm from each F_1 male, we performed a standard swim up
145 assay [32]. First, immediately following sacrifice, we collected and flash-froze liver and tail
146 control tissues (liver samples, N = 16; tail samples N = 8). Then, we removed and lacerated
147 the epididymides of each male, placed this tissue in 1.5 ml of human tubal fluid
148 (Embryomax® HTF, Millipore), and maintained the sample at a constant 37 °C for 10
149 minutes. Next, we isolated the supernatant, containing sperm that swam out of the
150 epididymides, and spun this sample for 10 minutes at 250 g. We then discarded the
151 supernatant, repeated the wash, and this time allowed sperm to swim up into the solution
152 for an hour to select the most robust cells. Finally, we removed the solution, transferred
153 them to new vial, pelleted these sperm by centrifugation, and froze them at -80 °C.

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

154

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

170 Alignment and Read Counting: We aligned read data to the combined reference genome
171 using 'BWA mem' as described above in the alignment simulation. We removed potential
172 PCR duplicates using Picard v1.73. We then filtered reads based on the alignment filtering
173 criteria described above for the simulated data. Because copy number variations may pose
174 problems for our analysis, we attempted to identify and exclude these regions. Specifically,
175 we broke the genome into non-overlapping 10 kb windows. Then, within each library, we
176 searched for 10 kb regions that had a sequencing depth greater than two standard

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

177 deviations above the mean for that library. All aberrantly high-depth windows identified
178 were excluded in downstream analyses in all libraries. These regions, representing
179 approximately 7% of the windows in the genome, are reported in Supplemental Table S1.
180 Although deletions in one parental strain relative to the MM9 genome could also skew the
181 parental allele frequencies for sequenced tissues, these copy number variable regions
182 would affect both somatic tissues and gametes equivalently, and we therefore do not expect
183 copy number variable regions to yield false positive results.

184 Next, to identify regions showing evidence of segregation distortion, we conducted
185 windowed analyses with 1 Mb between the centers of adjacent windows. We counted reads
186 in each window as a decreasing function of their distance from the center of the window,
187 and included no reads at distances greater than 20 cM, thereby placing the most weight in a
188 window on the center of the window. We then analyzed each window in two mixed-effects
189 generalized linear models. Both models included random effects for the libraries and
190 individuals. The first model includes no additional factors. The second had fixed effects for
191 tissue, direction of cross, and an interaction term based on tissue by direction of cross
192 effects, and thus has five fewer degrees of freedom than the first model. Hence, for each
193 window, we assessed the fit of the second model relative to the first using a likelihood ratio
194 test, wherein the log likelihood ratio should be chi-square distributed with 5 degrees of
195 freedom. Afterwards, we applied a false-discovery rate multiple testing correction to the
196 data [33]. We performed all statistical analyses in R [34].
197

198 Power Simulations: To estimate the power of our method, we simulated distortion data. We
199 began by selecting sites randomly distributed across the genome, and for each site drew a

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

200 distortion coefficient from a uniform distribution between -0.05 and 0.05. Each read on the
201 parental genome that was susceptible to distortion was counted on the distorting genome
202 with probability equal to the distortion coefficient multiplied by the probability that no
203 recombination events occurred between the distorted locus and the read. We also did the
204 alternative (*i.e.* switching reads from the distorted against genome to the distorting
205 genome) by multiplying by the probably that a recombination event was expected to occur
206 in the genomic interval between the distorter and the read. We determined recombination
207 probabilities using the genetic map reported in [35]. We performed the simulation for both
208 parental genomes, and then again for each parental genome but with the distortion limited
209 to one direction of the cross (*e.g.* only sperm from CW males experienced distortion). A
210 direction-specific effect could occur if, for example, suppressing alleles are present on the Y
211 chromosome of one subspecies and therefore are only present in CW or WC males.
212
213 **Results**
214 After addressing the possibility of contamination, labeling, and quality issues (see
215 Supplemental Text S1, Supplemental Table S2), we ran our analysis of the data across all
216 autosomes, excluding regions with evidence for copy-number variations (described in
217 Methods). With the exception of windows on chromosome 16 (see below), we found no
218 windows with a statistically significant signature of segregation distortion. The lowest
219 uncorrected p-value for any window (aside from those on chromosome 16) was 0.0224,
220 which is not significant when we corrected for multiple tests. Thus, we did not find
221 evidence for segregation distortion in any of the genomic regions considered (Figure 2).

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

222 By contrast, on chromosome 16, we identified 15 contiguous windows with
223 significantly skewed allele frequencies following correction for multiple comparisons
224 (minimum p = 5.026E-4; Figure 3). However, upon closer examination, it appears that this
225 signal is driven almost entirely by a single liver sample, that of individual CW10. If this
226 sample is removed from the dataset, this chromosome no longer shows significant
227 deviation from expectations. When comparing the relative read depths across
228 chromosomes 16 and 1, CW10's liver sample also appears to have disproportionately lower
229 depth on this chromosome relative to CW10's sperm sample (p = 3.02E-5; X^2-test). These
230 results suggest that this pattern is likely driven by a somatic aneuploidy event in CW10's
231 liver that occurred relatively early in liver development and is not the result of distortion in
232 the sperm library.
233 Through simulation, we ensured that we have sufficient statistical power, given our
234 experimental design and data quality, to detect segregation distortion if it is indeed
235 occurring in hybrid males. We found that we have 50% power to detect SD to
236 approximately 0.014, or 1.4% (this number reflects the positive or negative deviation from
237 the null expectation, 0.5, at α = 0.001) if distortion affects CW and WC males equally (Figure
238 4). In other words, we have 50% power to detect distortion that is greater than 51.5% or
239 less than 48.5%. If there is directionality to the distortion effect (*i.e.* only CW or only WC
240 males experience SD), we have 50% power to detect distortion of 0.016 for CW males and
241 0.018 for WC males (at α = 0.001). This slight difference in power based on cross direction
242 likely reflects differences in sequencing depth between WC and CW sperm and liver
243 samples. It is also important to note that because read mapping and sequencing, as well as
244 divergence between the CAST and WSB strains and their divergence from the reference

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

245 genome, are non-uniform across the genome, different regions of the genome will differ
246 slightly in power to detect distortion.
247
248 **Discussion**
249 Elucidating the genetic mechanisms underlying species formation is a central goal of
250 evolutionary biology. Although there has been progress in identifying the genetic basis of
251 reproductive isolation in a few elegant instances (*e.g.* [36-38]), including several in
252 *Drosophila* (*e.g.* [39,40]), it is unclear how general these results are. For example, SDs
253 contribute to reproductive isolation in some young Drosophila species pairs ([13-15]) but
254 here, to our surprise, we find no evidence for segregation distortion between two nascent
255 species of mouse, *M. m. castaneus*/*M. m. domesticus,* despite strong experimental power.
256 Our conclusion must be qualified to some degree. SDs are generally classified as
257 either gamete disablers or gamete killers depending on their mode of action (reviewed in
258 [7,8]). We expect to detect gamete killers with our approach since their victims may not be
259 present in the epididymides, or, if present, these sperm would not be captured in our
260 stringent swim up assay. Our ability to detect gamete- disablers, however, depends on the
261 specific mechanism by which these genetic elements act. If the motility or longevity of a
262 sperm cell is sufficiently impaired, it is likely that this sperm would fail to swim into
263 solution and remain motile over the course of the assay, but if the distortion effect has a
264 very subtle effect on motility or impairs function later in the sperm life cycle (*e.g.* by
265 causing a premature acrosome reaction), it is unlikely that our method could detect these
266 effects. Thus, although gamete killers are not prevalent sources of DMIs in these subspecies,
267 we cannot completely exclude the possibility that gamete disablers contribute to *M.*

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

268 *musculus* species formation. However, it is worth nothing that disablers cannot explain the
269 reported observation of increased apoptosis of sperm cells in hybrid males [26].
270 Conventional methods of detecting SDs (*i.e.* genotyping progeny) are usually
271 statistically underpowered and thus unable to detect even modest distortion effects.
272 Moreover, requiring the presence of offspring from F_1 hybrids unavoidably conflates
273 viability, gamete competition, and segregation distortion effects. By contrast, our
274 simulations demonstrate that by sequencing high quality gametes from individual hybrid
275 males and comparing allele ratios in these gametes to those of somatic tissues, we have
276 excellent power to detect even relatively weak SDs, of less than two percent. In support of
277 this point, we successfully detected an aneuploidy event that resulted in a four percent
278 difference in allele frequencies relative to expectations within only a single biological
279 replicate. Nonetheless, we found little evidence that SDs are active in F_1 hybrid males,
280 which indicates that segregation distortion (*i.e.* gamete killing) is not a primary contributor
281 to reduced F_1 male fertility in these subspecies.
282 Because our method of determining the allele ratios in bulk preparations of viable
283 gametes relative to somatic tissues is very general, we expect that it will be useful in a wide
284 variety of systems for an array of questions. Provided one can accurately phase the diploid
285 genome of an individual, by *e.g.* using complete parental genotype data when inbred strains
286 are not available, it is straightforward to apply this method to assay segregation distortion
287 in a wide variety of taxa (including humans). Thus, we are now well positioned to survey
288 the prevalence of segregation distortion both within and between a diversity of species.
289 This approach also allows segregation distortion to be weighed against other possible
290 sources of DMIs that may occur during spermatogenesis, oogenesis, fertilization, or

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

291 embryogenesis, but that leaves an identical signature to SD in conventional cross-based
292 experiments. Furthermore, because SDs can increase in frequency in populations despite
293 deleterious consequences for the host, these selfish genetic elements may also be an
294 important source of disease alleles. For example, it has been suggested that SDs contribute
295 to the perpetuation of split-hand/split-foot disease [41], retinal dystrophy [42] and
296 Machado-Joseph disease [43] in humans. Hence, the method introduced here has the
297 potential to improve our understanding of disease evolution in addition to the contribution
298 of SDs to the evolution of reproductive isolation between diverging lineages.

299 While segregation distorters may be an important mechanism of speciation in
300 *Drosophila* and crop plants, efforts to detect SD in other diverging lineages—especially
301 studies with high statistical power—have been limited. We find that at least in *M. m.*
302 *castaneus*/*M. m. domesticus* hybrids, segregation distorters are not measurable
303 contributors to observed infertility in F1 hybrid males, despite strong statistical power to
304 detect them, suggesting that reduced hybrid male fertility in these nascent species is
305 attributable to other underlying genetic causes. Further studies, using the novel approach
306 developed here will provide a powerful way to gain more comprehensive understanding of
307 the role of SDs within and between populations.
308

309 **Acknowledgements**

310 The authors declare no conflicts of interest. We thank Tim Sackton, Julien Ayroles, Brian
311 Arnold, Shelbi Russell, Brant Peterson, Marjorie Lienard, and Heidi Fisher for helpful
312 comments and suggestions, and Catherine Dulac and Stacey Sullivan for assistance with
313 mouse crosses. RCD was supported by a Harvard Prize Graduate Fellowship and currently

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

314 a UCB Chancellor's Postdoctoral Fellowship; EJP was supported by an NSF Predoctoral
315 Fellowship.

316

317 **References**

318 1. Dobzhansky T (1937) Genetics and the Origin of Species. Columbia University Press.
319 Columbia, NY.

320 2. Muller HJ (1942) Isolating mechanisms, evolution and temperature. Biol. Symp. 6:
321 71–125.

322 3. Coyne JA, Orr HA (2004) Speciation. Sinauer Associates Sunderland, MA.

323 4. Presgraves DC (2010) The molecular evolutionary basis of species formation. Nat
324 Rev Genet 11: 175–180.

325 5. Wu C-I, Ting C-T (2004) Genes and speciation. Nat Rev Genet 5: 114–122.
326 doi:10.1038/nrg1269.

327 6. Noor MAF, Feder JL (2006) Speciation genetics: evolving approaches. Nat Rev Genet
328 7: 851–861. doi:10.1038/nrg1968.

329 7. Lyttle TW (1991) Segregation distorters. Annual review of genetics 25: 511–581.
330 doi:10.1146/annurev.ge.25.120191.002455.

331 8. Taylor DR, Ingvarsson PK (2003) Common Features of Segregation Distortion in
332 Plants and Animals. Genetica 117: 27–35. doi:10.1023/A:1022308414864.

333 9. Hartl DL (1975) Modifier theory and meiotic drive. Theoretical population biology 7:
334 168–174. doi:10.1016/0040-5809(75)90012-X.

335 10. Frank SA (1991) Divergence of meiotic drive-suppression systems as an explanation
336 for sex-biased hybrid sterility and inviability. Evolution: 262–267.

337 11. Hurst LD, Pomiankowski A (1991) Causes of sex ratio bias may account for unisexual
338 sterility in hybrids: a new explanation of Haldane's rule and related phenomena.
339 Genetics 128: 841–858.

340 12. McDermott SR, Noor MAF (2010) The role of meiotic drive in hybrid male sterility.
341 Philosophical Transactions of the Royal Society B: Biological Sciences 365: 1265–
342 1272. doi:10.1098/rstb.2009.0264.

343 13. Tao Y, Masly JP, Araripe L, Ke Y, Hartl DL (2007) A sex-ratio Meiotic Drive System in
344 Drosophila simulans. I: An Autosomal Suppressor. PLoS Biol 5: e292.

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

345 doi:10.1371/journal.pbio.0050292.

346 14. Tao Y, Araripe L, Kingan SB, Ke Y, Xiao H, et al. (2007) A sex-ratio meiotic drive
347 system in Drosophila simulans. II: an X-linked distorter. PLoS Biol.
348 doi:10.1371/journal.pbio.0050293.g001.

349 15. Phadnis N, Orr HA (2009) A Single Gene Causes Both Male Sterility and Segregation
350 Distortion in Drosophila Hybrids. Science 323: 376–379.
351 doi:10.1126/science.1163934.

352 16. Bohn GW, Tucker CM (1940) Studies on Fusarium wilt of the Tomato. I. Immunity in
353 Lycopersicon pimpinellifolium Mill, and its inheritance in hybrids. Research Bulletin
354 Missouri Agricultural Experiment Station 311.

355 17. D R Cameron RMM (1957) Inheritance in Nicotiana Tabacum Xxvii. Pollen Killer, an
356 Alien Genetic Locus Inducing Abortion of Microspores Not Carrying It. Genetics 42:
357 326.

358 18. Loegering WQ, Sears ER (1963) Distorted inheritance of stem-rust resistance of
359 Timstein wheat caused by a pollen-killing gene. Canadian Journal of Genetics and
360 Cytology 5: 65–72.

361 19. Sano Y, Chu Y-E, Oka H-I (1979) Genetic studies of speciation in cultivated rice. 1.
362 Genic analysis for the F1 sterility between Oryza sativa L. and Oryza glaberrima
363 Steud. Japanese journal of genetics 54: 121–132.

364 20. Yang J, Zhao X, Cheng K, Du H, Ouyang Y, et al. (2012) A killer-protector system
365 regulates both hybrid sterility and segregation distortion in rice. Science.
366 doi:10.1126/science.1222597.

367 21. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, et al. (2005) A Scan for
368 Positively Selected Genes in the Genomes of Humans and Chimpanzees. PLoS Biol 3:
369 e170. doi:10.1371/journal.pbio.0030170.

370 22. Kosiol C, Vinař T, da Fonseca RR, Hubisz MJ, Bustamante CD, et al. (2008) Patterns of
371 Positive Selection in Six Mammalian Genomes. PLoS Genetics 4: e1000144.
372 doi:10.1371/journal.pgen.1000144.

373 23. Boursot P, Din W, Anand R, Darviche D, Dod B, et al. (1996) Origin and radiation of
374 the house mouse: mitochondrial DNA phylogeny. Journal of Evolutionary Biology 9:
375 391–415. doi:10.1046/j.1420-9101.1996.9040391.x.

376 24. Geraldes A, Basset P, Gibson B, Smith KL, Harr B, et al. (2008) Inferring the history of
377 speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes.
378 Molecular Ecology 17: 5349–5363. doi:10.1111/j.1365-294X.2008.04005.x.

379 25. Davis RC, Jin A, Rosales M, Yu S, Xia X, et al. (2007) A genome-wide set of congenic

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

380 mouse strains derived from CAST/Ei on a C57BL/6 background. Genomics 90: 306–
381 313. doi:10.1016/j.ygeno.2007.05.009.

382 26. White MA, Stubbings M, Dumont BL, Payseur BA (2012) Genetics and Evolution of
383 Hybrid Male Sterility in House Mice. Genetics 191: 917–934.
384 doi:10.1534/genetics.112.140251.

385 27. Wagner J (2010) Transmission Ratio Distortion in the Hybrid Offspring of Three
386 Mouse Subspecies. BS Thesis, University of Wisconsin, Madison Available
387 athttpbitstreamhandleWagnerpdf Accessed : 1–54.

388 28. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, et al. (2011) Mouse genomic
389 variation and its effect on phenotypes and gene regulation. Nature 477: 289–294.
390 doi:10.1038/nature10413.

391 29. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler
392 transform. Bioinformatics 25: 1754–1760. doi:10.1093/bioinformatics/btp324.

393 30. Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast
394 mapping of Illumina sequence reads. Genome Research 21: 936–939.
395 doi:10.1101/gr.111120.110.

396 31. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. (2011) A framework
397 for variation discovery and genotyping using next-generation DNA sequencing data.
398 Nat Genet 43: 491–498. doi:10.1038/ng.806.

399 32. Holt WV, Hernandez M, Warrell L, Satake N (2010) The long and the short of sperm
400 selection in vitroand in vivo: swim-up techniques select for the longer and faster
401 swimming mammalian sperm. Journal of Evolutionary Biology 23: 598–608.
402 doi:10.1111/j.1420-9101.2010.01935.x.

403 33. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and
404 powerful approach to multiple testing. Journal of the Royal Statistical Society Series
405 B 57: 289–300.

406 34. Team RDC (n.d.) R: A language and environment for statistical computing. R
407 Foundation for Statistical Computing, Vienna, Austria ISBN .

408 35. Cox A, Ackert-Bicknell CL, Dumont BL, Ding Y, Bell JT, et al. (2009) A New Standard
409 Genetic Map for the Laboratory Mouse. Genetics 182: 1335–1344.
410 doi:10.1534/genetics.109.105486.

411 36. Bradshaw HD, Schemske DW (2003) Allele substitution at a flower colour locus
412 produces a pollinator shift in monkeyflowers. Nature. doi:10.1007/s00294-003.

413 37. Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J (2009) A Mouse Speciation
414 Gene Encodes a Meiotic Histone H3 Methyltransferase. Science 323: 373–375.

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

415 doi:10.1126/science.1163601.

416 38. Lassance J-M, Groot AT, Liénard MA, Antony B, Borgwardt C, et al. (2010) Allelic
417 variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones.
418 Nature 466: 486–489. doi:10.1038/nature09058.

419 39. Ting C-T, Tsaur S-C, Wu M-L, Wu C-I (1998) A Rapidly Evolving Homeoboxat the Site
420 of a HybridSterility Gene. Science 282: 1501–1504.

421 40. Masly JP, Jones CD, Noor MAF, Locke J, Orr HA (2006) Gene Transposition as a Cause
422 of Hybrid Sterility in Drosophila. Science 313: 1448–1450.
423 doi:10.1126/science.1128721.

424 41. Jarvik GP, Patton MA, Homfray T, Evans JP (1994) Non-Mendelian transmission in a
425 human developmental disorder: split hand/split foot. American Journal of Human
426 Genetics 55: 710.

427 42. Evans K, Fryer A, Duvall-Young J, Inglehearn C, Whittaker J, et al. (1994) Genetic
428 linkage of cone–rod retinal dystrophy to chromosome 19q and evidence for
429 segregation distortion. Nature 6: 210–213. doi:10.1038/ng0294-210.

430 43. Ikeuchi T, Igarashi S, Takiyama Y, Onodera O, Oyake M, et al. (1996) Non-Mendelian
431 transmission in dentatorubral-pallidoluysian atrophy and Machado-Joseph disease:
432 the mutant allele is preferentially transmitted in male meiosis. American Journal of
433 Human Genetics 58: 730.

434

435 **Author Contributions**

436 RCD, EJP, DH, and HEH conceived and designed experiments. RCD, EJP, and HEH wrote the
437 manuscript. RCD and EJP performed experiments. RCD analyzed the data.

438

439 **Figure and Table Legends**

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

bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

445

446 Figure 2. Average proportion CAST reads in sperm libraries versus liver libraries, using all
447 males (A), using only CW males (B), and using only WC males (C). Lines indicate the
448 approximate threshold at which we would have 50% power to detect distortion at the
449 alpha = 0.0001 level (see Methods for how this threshold was calculated).

450

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

454

455 Figure 4. Minimum level of distortion that could be detected given a specified significance
456 threshold (α, y-axis), and desired power (x-axis).

457

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

460

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

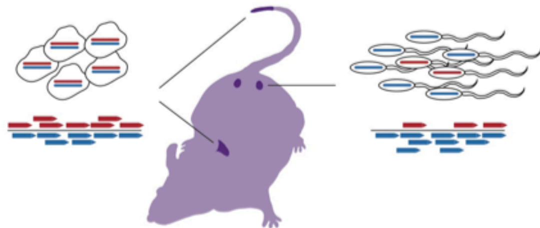
463

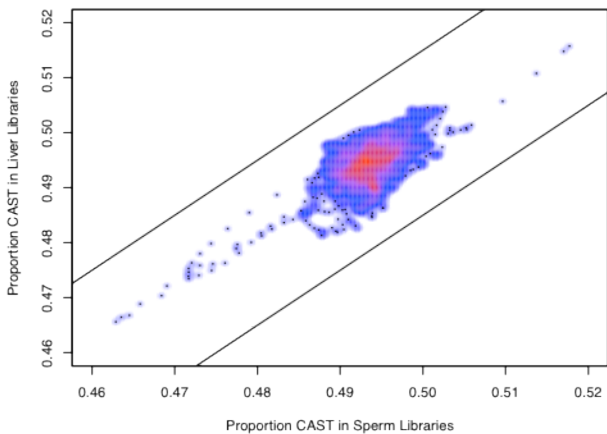
464 Supplemental Table S2. Quality control results for the quantity of reads in each library
465 derived from the *Y* chromosome, *X* chromosome, and mtDNA.

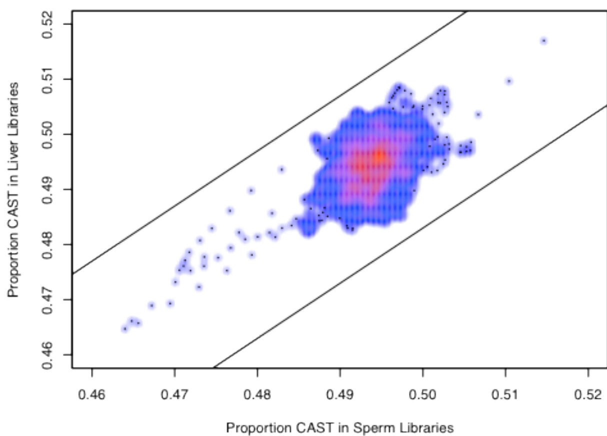
466

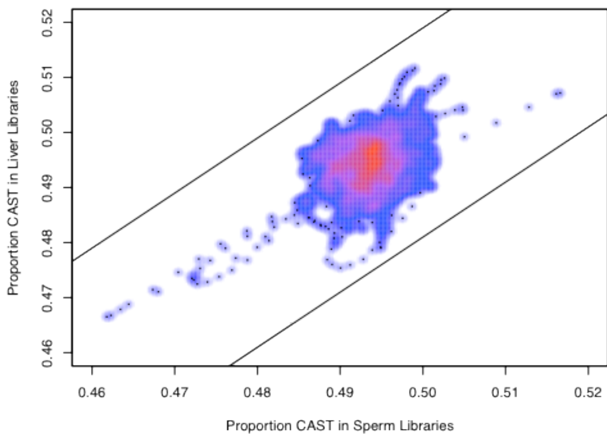
bioRxiv preprint first posted online September 2, 2014; doi: http://dx.doi.org/10.1101/008672; The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

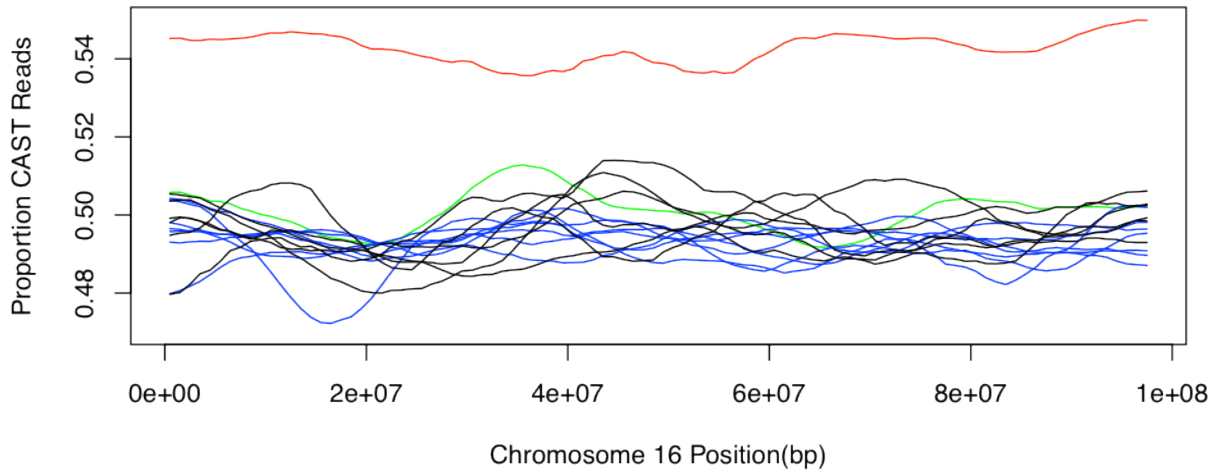
467 Supplemental Table S3. Alignment simulation results showing the relationship between
468 the reported mapping quality for a read and its probability of correct assignment to the
469 genomic location from which it was derived.
470

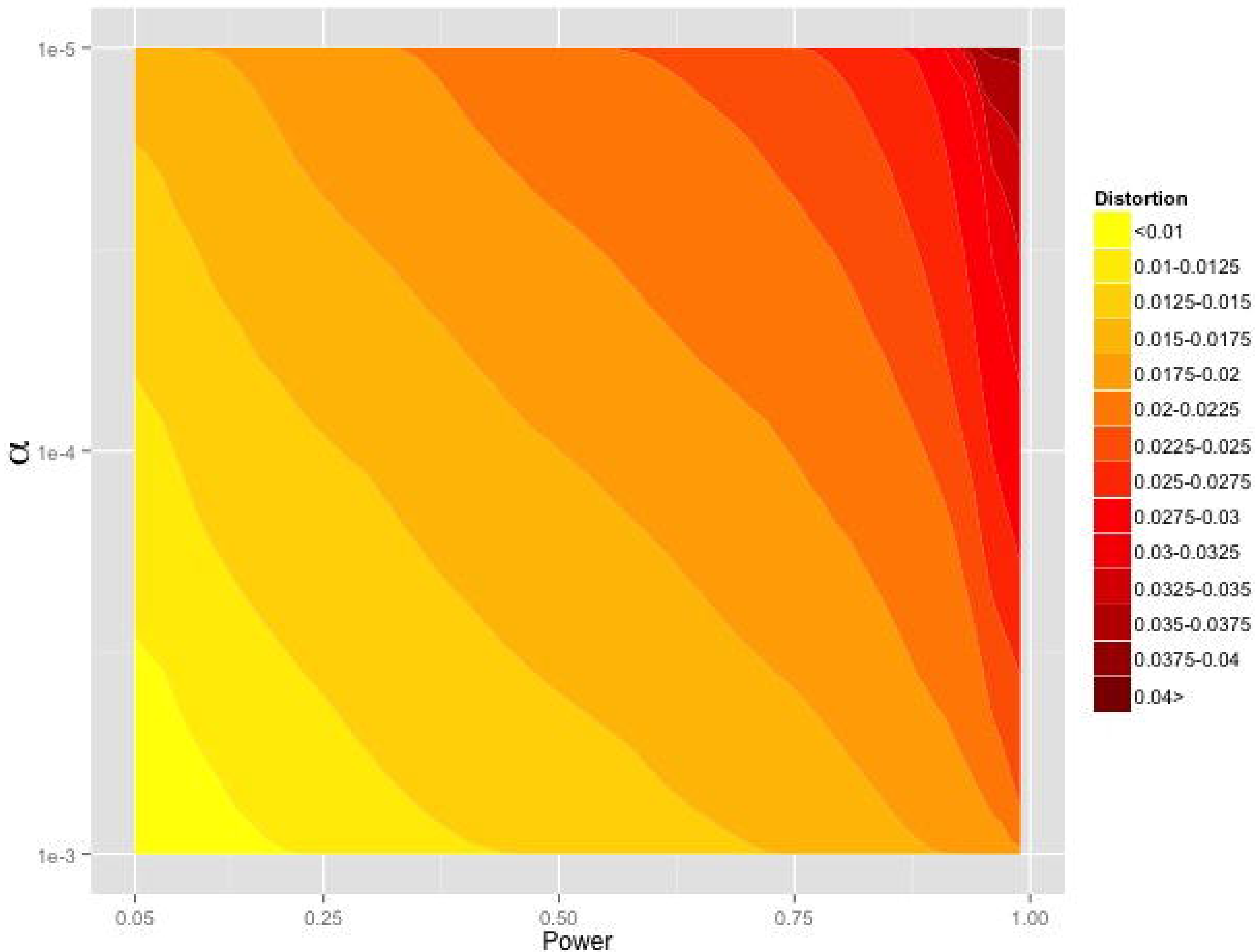



Proportion CAST in Liver Libraries
Proportion CAST in Sperm Libraries
0.46
0.47
0.48
0.49
0.50
0.51
0.52


Proportion CAST in Liver Libraries
Proportion CAST in Sperm Libraries
0.46
0.47
0.48
0.49
0.50
0.51
0.52


Proportion CAST in Liver Libraries
Proportion CAST in Sperm Libraries
0.46
0.47
0.48
0.49
0.50
0.51
0.52


Proportion CAST Reads
0.48
0.50
0.52
0.54
0e+00
2e+07
4e+07
6e+07
8e+07
1e+08
Chromosome 16 Position(bp)


1e-5
1e-4
1e-3
α
0.05
0.25
0.50
0.75
1.00
Power
Distortion
<0.01
0.01-0.0125
0.0125-0.015
0.015-0.0175
0.0175-0.02
0.02-0.0225
0.0225-0.025
0.025-0.0275
0.0275-0.03
0.03-0.0325
0.0325-0.035
0.035-0.0375
0.0375-0.04
0.04>