Conspecific sperm precedence is reinforced, but postcopulatory sexual selection weakened, in sympatric populations of *Drosophila*

Supplemental Methods

Conspecific sperm competition assay

Virgin individuals were collected and aged 7 days prior to the initiation of an experimental block. One day before mating, *D. persimilis* tester males were isolated individually (Dixon et al. 2003). The following day, females were individually added (without anesthesia) to a vial containing a tester male and were co-housed for 24 hours, after which the tester male was removed. Females were housed individually in these vials for 7 days before second mating (similar to Davis et al. 2017). After 7 days we inspected all vials for the presence of larvae to determine if females had mated with the first *D. persimilis* tester males. This was used to evaluate evidence for differences in successful first matings (pre-mating isolation) among allopatric and sympatric populations, rather than observing matings directly, as there is high variance in time to copulation in this heterospecific pairing (Davis et al. 2017). Only females that had mated (i.e. had produced larvae within 7 days) were retained for the remainder of the CSP experiment.

For the second mating, each individual female was paired with one of the four *D. pseudoobscura* male genotypes from her own population to determine the strength of CSP. These second males were also isolated one day before the introduction of the female. Seven days after mating with the first male, females were transferred, without anesthesia, to the vial containing the second male. Individual pairs were co-housed for 24 hours and the male was removed on the second morning. The female was kept for five days (transferring after 2 days to avoid overcrowding of larvae). All progeny produced in the five-day window after the second mating were collected; from these progeny, a maximum of 10 males and 10 females, randomly chosen from the total group of progeny, were used to score CSP (P2) as described below.

Generating visibly-marked tester males for quantifying CSP and ISC

To allow efficient progeny scoring, paternity was scored with the aid of visible markers in both CSP and ISC experiments. This required us to generate marked male tester lines with wild-caught *D. persimilis* (for CSP tester males) and *D. pseudoobscura* (for ISC tester males) lines from each study population.

For CSP, we introgressed an X-linked marker ("short" or *sh*) from a *D. persimilis* line (UCSD stock center 14011-0111.57) into four of our collected *D. persimilis* genotypes to act as a visible marker (Supplemental Fig. 2). *D. persimilis* has very few genetic markers, none of which are dominant. The X linked marker allowed us to infer the genotype of female offspring in a single cross. These four *D. persimilis* tester males originated from isofemale lines collected at the Sierra and Mt St. Helena locations and were used to evaluate the mean strength and variation in CSP for all four *D. pseudoobscura* populations in the CSP experiment. We first crossed these *D. persimilis sh* mutant males to females from each of the four wild-type *D. persimilis* isofemale lines (keeping each tester genotype separate throughout this process). This produced F1 daughters heterozygous for the *sh* allele, that were backcrossed to wild type males from the same wildtype isofemale line. From the BC1 progeny we retained *sh* males, and these were backcrossed to the original *D. persimilis* isofemale line to generate BC2s (Supplemental Fig 2). This process of alternating males and females for each backcross generation within each *D. persimilis* isofemale line was completed until the BC12. The alternation of male/female during backcrossing was necessary because recombination only occurs in females, but to retain the marker we had to select for *sh* males every second generation. After the BC12, the progeny within each BC isofemale line were interbred to create males and females homozygous for the *sh* allele. We did not directly evaluate how much of the *sh* line genome was introgressed in each case, however, *D. pseudoobscura* and relatives have a much higher recombination rate than *D.*

melanogaster (McGaugh et al. 2012), and previous introgression lines between these species have eliminated unwanted regions after 4 generations of backcrossing (Ortiz-Barrientos and Noor 2005).

For ISC experiments, the marked tester males were created by introgressing a green fluorescent protein marker (GFP) into 2 wild type *D. pseudoobscura* strains per location (therefore 8 strains in total) using wild-collected isofemale lines that were not used as female or second male genotypes for the ISC experiments. The original GFP strain was obtained from the UCSD stock center (14011-0121.166) the creation of which is described in Holtzman et al. (2010). We chose this marker because it is dominant (Castillo and Moyle 2014). We chromosomally mapped the GFP insertion of the original GFP strain to the second chromosome using a multiply marked (MM) strain, which contains visible recessive markers on all of the major chromosomes (*y;gl;or;inc* kindly provided by N. Phadnis, University of Utah). This mapping was completed in order to ensure the GFP insertion was not on the 3rd chromosome which, in *D. pseudoobscura*, contains large inversions that would have inhibited recombination of the marker into the wild-type backgrounds of our *D. pseudoobscura* isofemale lines.

The original GFP line was created in a stock that carried the X-linked white mutation. To eliminate the white allele from the population, in the parental cross we crossed the WT line with the GFP carrying male, and then used only F1 males with wild-type X chromosomes (no white mutation) to backcross in this initial generation. For the remaining eight backcross generations, we used females to allow recombination. We then chose 10 sibling pairs for each genotype to ensure the GFP marker was homozygous. These sub-lines were inbred for two generations. In the second generation we testcrossed the founder pair of individuals of each sub-line to ensure they were homozygous for the GFP marker. We recovered 2-4 lines that were homozygous for the GFP marker for each genotype. We then combined inbred lines that had originated from the

same isofemale genotype to reduce any potential effects of inbreeding depression that might have arisen during marker introgression.

Scoring conspecific sperm precedence

CSP was scored in F1 male progeny by assessing their fertility (hybrid male offspring of *D. persimilis* sires are sterile), and in F1 female progeny by scoring the presence of a visible marker in their progeny. Our rationale for analyzing both female genotype and male genotype was to provide two complementary estimates of hybrid identity and therefore relative sperm competitive success in our assay. Since these two estimates were highly correlated (as expected) we combined them into a single estimate, increasing sample size per cross and providing more precise estimates of the proportion of hybrid vs non-hybrid progeny. Note that all measurements were made blindly with respect to the provenance of scored individuals, including the selection of twenty progeny to be scored (see below). No information about cross identity was known to the person choosing the individuals to be sampled, while dissecting male testes, or scoring the presence/frequency of the *sh* allele.

To score F1 male fertility, each male was dissected individually in PBS buffer, and its testes moved to a slide that had 1ul of PBS buffer. A cover slip was placed over the slide and the testes were squashed, releasing sperm into the buffer. The slides were examined under an EVOS FL microscope for the presence of motile sperm. If no motile sperm were present, the male was scored as hybrid. 10 F1 males were scored from every cross replicate. To score paternity of F1 females, since the *sh* allele is recessive we could not score F1 females directly, but instead scored their offspring for the presence of the *sh* allele. If an F1 female was hybrid (and carrying the *sh* allele from the *D. persimilis* male) we would expect half of her sons and half of her daughters to have the *sh* phenotype. We previously confirmed that the half segregation held for known hybrid

progeny. For each cross, ten F1 females (that could be hybrid or purebred) were housed individually with a *D. pseudoobscura* male that also carried the *sh* allele (UCSD stock center *Dpse co;sh* 14011-0121.13). We chose a *D. pseudoobscura* male for these crosses to increase the number of progeny to score since *D. pseudoobscura* females (and therefore any purebred female progeny in our experiment) exhibit premating isolation with *D. persimilis* males; hybrid females do not demonstrate a mating preference. After a week the parental individuals were cleared from the vials and the vials were retained to score progeny. As progeny eclosed they were scored for the presence of *sh* allele.

Our measure of CSP was then the number of purebred progeny out of the total number of F1 individuals scored for a particular cross. Any replicate in which all progeny were scored as hybrid was excluded in our analyses because we could not ensure that a second mating had taken place. Note that the estimated frequency of this failure to remate does not differ between populations (Davis et al 2017). Every CSP estimate was based on at least 10 scored progeny and, for the majority of the crosses, we scored close to 20 individuals. In addition, to ensure that CSP estimated here does not simply reflect stronger fecundity stimulation by conspecific males, in an additional experiment we determined that there was no difference in non-competitive progeny production in heterospecific vs. conspecific matings for any of the allopatric or sympatric populations, consistent with previous work (Lorch and Servedio 2005; Davis et al. 2017). There was also no correlation between the total number of progeny scored for CSP and the magnitude of CSP, and the number of progeny scored did not differ between populations. These observations suggest that there are no postzygotic survivorship barriers in interspecific hybrids that systematically differ between sympatric and allopatric populations, that might otherwise confound our estimate of CSP.

Statistical Analysis

Differences in mean and variance of CSP and ISC between populations

For analyses of mean differences, we pooled the two allopatric populations because they did not differ in mean CSP (Allopatry t = -0.45064, df = 123.62, P = 0.653) and pooled the two sympatric populations for the same reason (Sympatry t = -0.86678, df = 125.87, P = 0.3877). We again pooled the allopatric and sympatric populations because the variance was equivalent between allopatric populations ($\chi 2$ = 0.031899, P = 0.8585), and between sympatric populations ($\chi 2$= 0.80562, P = 0.3711. For ISC we pooled the individual allopatric and sympatric populations as they were not significantly different from one another for either mean (Allopatric mean t=-1.136, df=118,66, P=0.2593; Sympatric mean t=0.191, df=125.72, P=0.8488) or variance (Allopatric variance $\chi 2$=0.949, P=0.3316; Sympatric variance $\chi 2$=0.0796, P=0.7782). Note that, although all results reported in the main text are from tests with these pooled data, we also observed significant differences in pairwise tests between individual allopatric and sympatric populations, for both average and variance measures of CSP and ISC (Supplemental Tables 5, 6, 13 and 14.

We compared the total phenotypic variation between geographical classes of population with a Levene-type test implemented in the lawstat package in R is a Kruskal-Wallis modified Brown-Forsythe Levene-type test. The Brown-Forsythe test is based on the absolute deviations from the median, which retains statistical power for many types of non-normal data (Brown and Forsythe 1974). Kruskal-Wallis tests are rank-based tests. We used the Kruskal-Wallis modification because the variance in proportion data derived from binomial data does not accurately reflect variance in the original data (Warton and Hui 2011).

Genetic variation and genotype effects on CSP and ISC

In the bootstrap procedure to estimate genotype effects, data are simulated from the null model which lacks the random effect of interest. Then the full and reduced models are fit to the simulated data to determine the bootstrap distribution of the Likelihood Ratio test statistic. To the model above we also included a random effect of tester male (*D. persimilis* for CSP and GFP *D. pseudoobscura* strain for ISC). To assess the relative importance of each variable we calculated the intraclass correlation for each coefficient (see supplement for details); a high correlation indicates that the variable explains much of the variance in the data.

To assess the relative importance of each variable we calculated the intraclass correlation for each coefficient; a high correlation indicates that the variable explains much of the variance in the data. The ICC for the female effect, for example, would be:

$$ICC_F = \frac{\sigma_F^2}{\sigma_F^2 + \sigma_M^2 + \sigma_{MF}^2 + \sigma_T^2 + \frac{\pi^2}{3}}$$

Where *F* represents female variance, *M* represents male variance, *MF* represents the interaction, and *T* represents the identity of the tester male. The $\frac{\pi^2}{3}$ replaces the residual variance for the binomial model with logit link function. In the case of binomial regression the ICC values are on the log scale, and there is no convenient transformation to proportion scale (Eldridge et al. 2009), so they are presented here as a relative measure of variance explained.

Genetic association between CSP and ISC

We tested 14 candidate genes –each of which has an ortholog in *D. melanogaster* with a known role in sperm competition– for associations with CSP and ISC. In parallel, we also analyzed 13 control genes, 11 of which have been previously used as controls in studies of sperm competition genes (Begun and Whitley 2000; Begun et al. 2000; Swanson et al. 2001; Supplemental Table 3).

The additional two genes (Myosin 10A and Smrter) were added as genes with a large number of SNPs, to better match the large number of variable sites we observed in several female-acting sperm competition candidate genes. For each candidate gene we determined variable sites (single nucleotide polymorphisms; SNPs) from whole body transcriptome data. Transcriptomes were prepared by extracting total RNA from 20 whole bodies per library. We made separate libraries for male and female for each strain, including *D. persimilis*, used in the experiment. Briefly, virgin individuals were collected and maintained in same sex vials for seven days. Twenty flies were placed, without anesthesia, into an Eppendorf tube and immediately submerged in liquid nitrogen and stored at -80 degrees C. For extraction, tubes were removed from -80 and placed on ice. Samples were homogenized with Trizol reagent and RNA was extracted using the Invitrogen PureLink RNA mini kit (ThermoFisher Scientific). Libraries were prepared by the IU Center for Genomics and Bioinformatics (CGB). Libraries were sequenced by the CGB using two runs of 75bp long reads on the Illumina NextSeq (Illumina). This generated ~3.5million reads per library/sample. Raw reads were processed by removing adapters using scythe (Bufffalo 2014) and quality trimming using sickle (Joshi and Fass 2011). Reads were mapped to the *D. pseudoobscura* r3.04 genome using bwa mem (Li 2013). Duplicates were determined using picard (Broad Institute 2018) and filtered for quality score above 20 using samtools 1.8 (Li et al. 2009). SNPs were called using a standard GATK pipeline that included the haplotype caller algorithm followed with joint calling protocol (De Pristo et al. 2011). Lastly SNPs were filtered with the following GATK options (QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0). Only bi-allelic SNPs were used for association tests. In addition, since we phenotyped individuals, but sequenced pools for each isofemale line (which were not inbred, and therefore can have residual heterozygosity) we treated heterozygous calls within an isofemale line as missing data. The analyses reported in this study used only sequence data from

each genotype for the 14 candidate genes, extracted from the larger transcriptome dataset. To generate these data, FASTA files were created using the GATK alternate FASTA tool. This tool creates FASTA entries based on the annotated genome (in this case, the *D. pseudoobscura* genome) by replacing reference genome alleles with SNPs called by GATK where applicable. We used the option to include heterozygous sites in these files so that one allele was not chosen randomly, but subsequently masked these sites as missing data . All resulting FASTA files are deposited into the Dryad Digital repository (doi:10.5061/dryad.n88kb1m) along with csv files of the variable sites used in the analysis.

We found that population structure had quite strong effects on SNP variation in both control genes and candidate genes, so we corrected for this by only using SNPs with low Fst between allopatric and sympatric populations, as well as including a term for population origin in our statistical model. This is a very conservative correction that effectively examines only SNPs that are segregating in both sympatry and allopatry. Given the small number of genotypes per geographic local (n=8) our Fst distribution had discrete values from which we could choose a cutoff. We defined low Fst as any SNP with an Fst<0.06. In the context of our samples this value corresponded to a SNP that occurred in a single individual in a single population. Values greater than 0.06 indicated population specific differentiation, while values less than 0.06 ensured this SNP occurred in both populations. We calculated Fst using the Hudson metric in the R package KRIS (Chaichoompu et al. 2018).We estimated the association between individual SNPs and either the CSP or ISC phenotype while including population origin and genotype/line of the mating partner (see Results). To do this we used models that are structurally analogous to the models used to analyze our phenotype data (described above), but which replaced the male genotype or female genotype term with the SNP genotype for each line, depending on whether the focal gene was male- or female- specific. For example, if the gene was male specific, the

model would predict CSP or ISC as a function of whether the genotype came from allopatry or sympatry, the genotype at the specific SNP, the female genotype that male mated with, and the SNP*female genotype interaction. Models were fit using linear regression. Significant SNP associations were estimated using confidence intervals.

Quantifying sexual selection and variance in male reproductive success

To evaluate whether the intensity/opportunity for sexual selection differs among populations we require an estimate of variance in male reproductive success (Wade 1979). In a natural population most males can gain fitness through offensive (P1) and defensive (P2) sperm competition, so the best estimate for variance in reproductive success would be total progeny produced. In our experiment we did not score lifetime progeny production, and specific male genotypes were either used as offensive or defensive males only. As such we estimated male fitness as the proportion of progeny sired, taking into consideration that we had two distinct classes of males—tester first (defensive) males and second (offensive) males--that may differ in their frequency and variance in fitness in the experiment. Following Shuster et al. (2013) we define total variance in male reproductive success as the sum of within and between male class variance

$$V_{total} = (f_{P1})(V_{P1}) + (f_{P2})(V_{P2}) + (\bar{X}_{P2} - \bar{X}_{P1})^2(f_{P2})(f_{P1})$$

The two terms on the left hand of the equation represent the within class variance (for example, V_{P1} is the variance in sperm competitive success between tester males and f_{P1} is the frequency of tester males used in the experiment). The last term represents the between class variance.

We were interested in reproductive variance at the level of male genotype so we averaged biological replicates to generate mean fitness values for each individual genotype. We used empirical bootstrap confidence intervals to estimate error that may have been a product of

averaging over replicates (Efron and Tibshirani 1993; Davison and Hinkley 1997). For the bootstrap procedure we sampled 16 data points, with replacement, from the 16 original empirical replicates for each genotype (32 for defensive males). We then averaged these data points and calculated V_{total} as described above. We completed 1000 bootstrap replicates for each population. We constructed the 95% confidence interval using the bootstrap difference $\delta^* = V_{total} - V^*_{total}$ where * represents each bootstrap replicate. The interval is then $[V_{total} - \delta^*_{0.05}, V_{total} - \delta^*_{0.95}]$.

The confidence intervals for the Zion population did not overlap with the confidence intervals for either sympatric population and can be considered significantly different at the 0.05 level (Supplemental Table 13). The Lamoille population confidence intervals overlapped with the sympatric populations, but overlap in confidence intervals does not mean parameters are not statistically different (Schenker and Gentleman 2001). This is because confidence intervals calculated for independent parameters cannot replace a comparative test of the differences between two parameters. Therefore, we conducted bootstrap hypothesis testing (Efron and Tibshirani 1993; Davison and Hinkley 1997) to determine whether differences in V_{total} between populations were significant, specifically by calculating bootstrap F statistics. The F statistic is a ratio of any two variance parameters, for example $F = V_{total,pop1}/V_{total,pop2}$. We compared the V_{total} in pairwise comparisons following standard bootstrap methods, where bootstrap samples are generated under the null hypothesis, and then this distribution is compared to the empirically observed statistic. For our scenario, the null hypothesis was that there was no differences in V_{total} between populations. Therefore we sampled, with replacement, offensive and defensive genotypes after pooling data from both populations. We then randomly assigned each value to one of the two populations. This generated a bootstrap replicate with approximately equal variance between the populations. We then could calculate $F = V_{total,pop1}/V_{total,pop2}$ for each

replicate. The bootstrap p-value is then calculated by comparing the bootstrap statistic ($F*$) to the observed statistics (F) using $p^*(F) = \frac{1}{B}\sum_{j=1}^{B} I(F^* > F)$. I() is an indicator function that is equal to 1 when the argument is true (bootstrap statistic > observed statistic), and 0 when false. B is the number of bootstrap replicates (1000 per population comparison).

The potential effects of sampling, geography, and adaptation to laboratory conditions on the experimental observations

Several experimental factors, unrelated to predicted differences in reinforcing selection between allopatric and sympatry sites, should also be considered in terms of their potential influence on our experimental observations and the inferences we can draw from them. First, because our analysis drew on a limited number of genotypes collected from each population (n=4 genotypes that were used as both male and female genotypes, and n=2 for tester male genotypes), it is likely that these do not represent the full range of genetic variation within and between natural populations. The associated sampling error has the potential to influence our inferences from these data if it is greater than the biological differences that exist between our populations. This sampling error is also expected to increase or decrease observed population differences with equal probability, without respect to their predicted biological differences. As a result, population phenotypic differences that are due to sampling error alone are equally likely to be observed as differences between allopatric populations, or between sympatric populations, as between the two population types.

Our observations, however, are not consistent with differences that are due to sampling error alone. Our three principle measures in this study were female mating rate, conspecific sperm precedence (CSP), and intraspecific sperm competition (ISC). For each measure, pairwise tests between all populations showed the same systematic pattern: sympatric populations were

always different than allopatric populations, whereas the two sympatric populations were always indistinguishable, as were the two allopatric populations (Supplemental Tables 2 (female matings), 6-7 (CSP), and 11-12 (ISC)). If sampling error was influencing our results we would expect, with equal probability, to see differences among allopatric populations, or among sympatric populations, and not strictly between these two types. The consistency and repeatability of the sympatric-allopatric differences across all measurements argues that sampling error is not principally responsible for the patterns of observed phenotypic variation between populations.

Similar to sampling error during collections, another factor to consider are the sampling effects that can occur when individual lines are established and cultured in the lab. Initiating *D. pseudoobscura* isofemale lines in a laboratory setting requires time-intensive molecular marker analysis and test crosses to confirm species identity, because species identification cannot be assigned using phenotypic differences. In our experiment, genotypes were maintained in the lab for a maximum of 1.5 years before the completion of data collection, during which time the relatively small effective population sizes within individual cultures could have altered the initial genetic and phenotypic variance present in the initial collections. Importantly, similar to sampling error, while genetic drift has the potential to exaggerate slight starting differences between populations, it is equally likely that drift could have decreased differences in population phenotypes. Given this, drift alone is unlikely to produce the observed differences in population phenotypes, as this would require that lines from our allopatric populations independently drifted in one direction while all lines from our sympatric populations independently drifted in the opposite direction. A more plausible force that can affect newly established lab lines would be lab adaptation, as indirect effects and nutrition can affect sperm competition (see Discussion).

However our expectation under this scenario is that all genotypes would evolve to resemble one another (due to identical lab conditions), and in this case we should not see systematic differences between allopatric and sympatric populations.

A third experimental factor that might potentially influence observations in this study is differences in physical proximity between sympatric and allopatric populations. In our study the two sympatric populations are geographically closer together than the two allopatric populations. This is because the sympatric range of *D. pseudoobscura* is highly restricted to the West Coast of the United States, and within this our sampling is focused on two populations in California. One reason populations might appear similar, regardless of selection, is physical proximity, raising the possibility that population phenotypic similarities and differences might be due to spatial effects, rather than qualitative differences in selection between allopatry and sympatry. Nonetheless, our observed patterns of phenotypic differentiation over geographic distance suggests this difference in physical proximity does not explain our results. The two sympatric populations are ~150 miles apart and their phenotypic distributions are indistinguishable, while the two allopatric populations are separated by a distance three times greater (~450 miles apart) but are also indistinguishable from each other. The phenotypic similarities across such large distances are consistent with high gene flow and low genomic differentiation across the range of *D. pseudoobscura* (Kulathinal et al 2009; McGaugh et al. 2012). In contrast, differences in post-copulatory phenotypes are only observed between populations that are predicted to experience different selective pressures (i.e. the presence/absence of the closely related *D. persimilis*), in a direction that is consistent with our a priori expectation for this selection pressure.

Overall, while sampling error, drift within a lab setting, and differences in geographical proximity, can clearly affect population phenotypic distributions, we infer that the influence of these factors on our observations are small compared to predicted differences in selection

between allopatric and sympatry sites. Repeated systematic differences in postcopulatory traits between allopatric and sympatric populations, in the direction predicted by a scenario of reinforcing selection, are not consistent with population differences that are due to these effects alone.

Literature Cited

Begun, D. J., and P. Whitley. 2000. Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. Proc. Natl. Acad. Sci UAS 97:5960-5965.

Begun, D.J., P. Whitley, B.L. Todd, H.M. Waldrip-Dail, and A.G. Clark. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila.* Genetics 156:1879-88.

Broad Institute. 2018. Picard Toolkit http://broadinstitute.github.io/picard/

Brown, M.B., and A. B. Forsythe. 1974. Robust tests for equality of variances. J. Am. Stat. Ass. 69:364-367.

Buffalo, V. 2014. Scythe. https://github.com/vsbuffalo/scythe

Castillo, D.M., and L.C. Moyle. 2014. Intraspecific sperm competition genes enforce post-mating species barriers in *Drosophila*. Proc. Roy. Soc. B281.

Chaichoompu, K., F. Abegaz, S. Tongsima, P.J. Shaw, A. Sakuntabhai, L. Pereira, and K. Van Steen. 2018. KRIS: Keen and Reliable Interface Subroutines for Bioinformatic Analysis. R package version 1.1.1. URL: https://CRAN.R-project.org/package=KRIS

Davis, J.S., D.M. Castillo, and L.C. Moyle. 2017. Remating responses are shaped by male post- copulatory manipulation but not reinforcement in *D. pseudoobscura*. Ecology and Evolution 7:507-515.

Davison, A.C., and C.V. Hinkley. 1997. Bootstrap methods and their application. Cambridge University Press, Cambridge, UK.

DePristo, M., E. Banks, R. Poplin, K. Garimella, J. Maguire, C. Hartl, A. Philippakis, G. del Angel, M.A. Rivas, M. Hanna, A. McKenna, T. Fennell, A. Kernytsky, A. Sivachenko, K. Cibulskis, S. Gabriel, D. Altshuler, M. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43:491-498.

Dixon, S.M., J.A. Coyne, and M.A.F. Noor. 2003. The evolution of conspecific sperm precedence in *Drosophila*. Molec. Ecol. 12:1179–1184.

Efron, B., and R.J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, Inc. London, UK.

Eldridge, S.M., O.C. Ukoumunne, and J.B. Carlin. 2009. The intra-cluster correlation coefficient in cluster randomized trials: a review of definitions. Int. Stat. Rev. 77:378-394.

Holtzman, S., D. Miller, R. Eisman, H. Kuwayama, T. Niimi, and T. Kaufman. 2010 Transgenic tools for members of the genus *Drosophila* with sequenced genomes. Fly 4:349-362.

Joshi, N.A., and J.N. Fass. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. Version 1.33 https://github.com/najoshi/sickle.

Kulathinal, R. J., L. S. Stevison, and M. A. F. Noor. 2009. The genomics of speciation in *Drosophila*: Diversity, divergence and introgression on a genome-wide scale. PLoS Genetics, 5: e1000550.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1

Lorch, P.D., and M.R. Servedio 2005. Postmating-prezygotic isolation is not an important source of selection for reinforcement within and between species in *Drosophila pseudoobscura* and *D. persimilis*. Evolution 59:1039-1045.

McGaugh, S.E., C.S.S. Heil, B. Manzano-Winkler, L. Loewe, S. Goldstein, T.L. Himmel, M.A.F. Noor. 2012. Recombination modulates how selection affects linked sites in *Drosophila*. PLoS Biol., 10: e1001423

Ortiz-Barrientos, D., and M.A.F. Noor. 2005. Evidence for a one-allele assortative mating locus. Science 310:1467.

Schenker, N., and J.F. Gentleman. 2001. On judging the significance of differences by examining the overlap between confidence intervals. Am. Stat. 55:182-186.

Shuster, S.M., W.R. Briggs, and P.A. Dennis. 2013. How multiple mating by females affects sexual selection. Phil. Trans. Roy. Soc. B :20120046.

Swanson, W.J., A.G. Clark, H.M. Waldrip-Dail, M.F. Wolfner, and C.F. Aquadro. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc Natl Acad Sci U S A. 98:7375-9.

Wade, M.J. 1979. Sexual selection and variance in reproductive success. Am. Nat. 114:742-747.

Warton, D.I., and F.K.C. Hui. 2011. The arcsine is asinine: the analysis of proportions in ecology. Ecology 92:3-10.

Supplemental Tables and Figures

Supplemental Table 1. The pairwise comparison of sympatric and allopatric populations for female premating isolation. The proportion of *D. pseudoobscura* females that mated with *D. persimilis* was compared using a χ^2 test. N is the total number of females tested.

Pop Allo.	Pop Sym.	Prop. Mate Allo (N)	Prop. Mate Sym (N)	χ^2	*P-value*
Lamoille	Mt.St. Helena	0.519 (179)	0.460 (200)	0.143	0.705
Lamoille	Sierra	0.519 (179)	0.495 (222)	1.113	0.291
Zion	Mt.St. Helena	0.524 (145)	0.460 (200)	0.184	0.667
Zion	Sierra	0.524 (145)	0.495 (222)	1.139	0.285

Supplemental Table 2. Candidate genes used in tests for an association with conspecific sperm precedence and intraspecific sperm competition. These candidates have orthologs in D. melanogaster with known roles in sperm competition.

D. pseudoobscura ID	*D. melanogaster* ortholog	Number of segregating sites (cds length)
FBgn0070474	*Seminase* (sems)	10 (825)
FBgn0248361	*aquarius* (aqrs)	17 (1047)
FBgn0071743	*intrepid* (intr)	36 (948)
FBgn0245599	*antares* (antr)	10 (849)
FBgn0270940	*Sex Peptide* (SP)	2 (174)
FBgn0082155	CG9997	28 (909)
FBgn0074106	lectin-46Cb	18 (1440)
FBgn0245732	lectin-46Ca	20 (1773)
FBgn0074591	CG17575	7 (876)
FBgn0080019	*Epidermal stripes and patches* (Esp)	60 (1971)
FBgn0076878	*fra mauro* (frma)	67 (1851)
FBgn0079015	*hadley* (hdly)	27 (1227)
FBgn0082000	*Neprilysin 2* (Nep2)	34 (2307)
FBgn0074155	*Sex Peptide Receptor* (SPR)	2 (1515)

Supplemental Table 3. Control genes used in tests for an association with conspecific sperm precedence and intraspecific sperm competition. These candidates have orthologs in *D. melanogaster* that are not involved in sperm competition and have been previously identified as a robust control set.

D. pseudoobscura ID	*D. melanogaster* ortholog	Number of segregating sites (cds length)
FBgn0243792	*zeste* (z)	10 (1755)
FBgn0012735	*period* (per)	78 (4101)
FBgn0243709	*Yolk protein 2* (Yp2)	14 (1392)
FBgn0071729	*Zwischenferment* (Zw)	5 (1590)
FBgn0243561	*Alcohol dehydrogenase* (Adh)	7 (873)
FBgn0075375	*FMRFamide* (FMRFa)	38 (1062)
FBgn0246212	*brown* (bw)	11 (2103)
FBgn0070521	*lamina ancestor* (lama)	38 (1989)
FBgn0078384	*knirps* (kni)	1 (1521)
FBgn0071371	*Relish* (Rel)	11 (2907)
FBgn0080509	*glass* (gl)	1 (2040)
FBgn0244150	*Smrter* (Smr)	72 (9880)
FBgn0074248	*Myosin 10A* (Myo10A)	89 (2307)

Supplemental Table 4. The average levels of reproductive isolation for each *D. pseudoobscura* population measured from two barriers to reproduction: female preference (proportion of females that did not mate with heterospecifics) and conspecific sperm precedence (CSP). Higher values indicate stronger reproductive isolation. Interpopulation sperm precedence (ISC) is included for comparison. The mean and variance estimates for CSP and ISC are based on 64 replicates per population. A = allopatric; S = sympatric.

	Female Pref.	CSP		ISC	
Population	Proportion (n)	Mean	Variance	Mean	Variance
Lamoille (A)	0.481 (179)	0.75	0.054	0.76	0.028
Zion (A)	0.476 (145)	0.77	0.041	0.80	0.047
Mt. St. Helena (S)	0.540 (200)	0.90	0.017	0.69	0.057
Sierra (S)	0.505 (222)	0.92	0.018	0.68	0.052

Supplemental Table 5. Significant variation between female genotypes in female mate preference, measured as the probability of mating with heterospecifics across populations is consistent across populations (Left). Only in one population (Lamoille) does the identity of the *D. persimilis* male influence female mate preference (Right). Significance was determined by a Wald's test on the logistic regression. A = allopatric; S = sympatric.

	Female Genotype		*D. persimilis* tester male	
Population	χ^2 test statistic df=3	*P*-value	χ^2 test statistic df=3	*P*-value
Lamoille (A)	14.0	**0.003**	8.1	**0.044**
Zion (A)	7.7	0.053	1.1	0.78
Mt. St. Helena (S)	22.6	**<0.001**	4.2	0.25
Sierra (S)	40.4	**<0.001**	0.15	0.99

Supplemental Table 6. The pairwise comparison of sympatric and allopatric populations for mean conspecific sperm precedence (CSP). Means were compared using *t*-test and Wilcox sum rank test. Since both are significant we present the t-test only.

Pop Allo.	Pop Sym.	Mean CSP Allo.	Mean CSP Sym.	*t* statistic (df)	*P-value*
Lamoille	Mt.St. Helena	0.75	0.90	-4.4 (99.5)	<0.001
Lamoille	Sierra	0.75	0.92	-4.9 (101.4)	<0.001
Zion	Mt.St. Helena	0.77	0.90	-4.3 (108.1)	<0.001
Zion	Sierra	0.77	0.92	-4.9 (110.1)	<0.001

Supplemental Table 7. The pairwise comparison of sympatric and allopatric populations for variance of conspecific sperm precedence (CSP). Variances were compared using a Kruskal-Wallis modified Brown-Forsythe Levene type test (see methods).

Pop Allo.	Pop Sym.	Var. CSP Allo.	Var. CSP sym	Levene type χ^2	*P-value*
Lamoille	Mt.St. Helena	0.054	0.017	13.70	<0.001
Lamoille	Sierra	0.054	0.018	27.78	<0.001
Zion	Mt.St. Helena	0.041	0.017	16.98	<0.001
Zion	Sierra	0.041	0.018	33.25	<0.001

Supplemental Table 8. The genotype effects that predict CSP. The maximum likelihood estimate (ML est.) and intraclass correlation (ICC) are reported as point estimates from the full model. The *P*-value for each term was calculated by comparing the observed Likelihood ratio test statistic (LR) to the distribution generated by parametric bootstrap. Data were bootstrap sampled according to the null hypothesis where the random effect of interest is not included. The full and reduced models are then fit to each bootstrap sample to determine the distribution for the LR test statistic. A = allopatric; S = sympatric. Bold indicates significance at *P*<0.05. Italics indicates marginal significance *P*<0.06.

Lamoille (A)				
Effect	ML est.	LR	*P*-Value	Intraclass Corr.
Female	**0.4024**	**8.10**	**0.0067**	0.096
Male	0.0000	0.00	0.7509	0.00
M x F	**0.1154**	**3.52**	**0.0383**	0.027
D. persimilis	**0.3413**	**37.49**	**0.0013**	0.082
Zion (A)				
Effect	ML est.	LR	*P*-Value	Intraclass Corr.
Female	*0.2683*	*2.72*	*0.05632*	0.067
Male	0.0000	0.00	0.4190	0.00
M x F	**0.3315**	**16.30**	**0.00238**	0.0833
D. persimilis	**0.0865**	**6.44**	**0.0068**	0.0217
Mt St. Helena(S)				
Effect	ML est.	LR	*P*-Value	Intraclass Corr.
Female	**0.8408**	**5.77**	**0.0068**	0.188
Male	0.0000	0.00	0.9891	0.000
M x F	**0.3266**	**8.76**	**0.0026**	0.0737
D. persimilis	0.0000	0.00	0.9851	0.000
Sierra (S)				
Effect	ML est.	LR	*P*-Value	Intraclass Corr.
Female	0.3287	0.72	0.1760	0.071
Male	0.1529	0.27	0.2673	0.033
M x F	**0.5975**	**7.28**	**0.0046**	0.129
D. persimilis	**0.2487**	**8.16**	**0.0012**	0.053

Supplemental Table 9. The genotype effects that predict ISC. The maximum likelihood estimate (ML est.) and intraclass correlation (ICC) are reported as point estimates from the full model. The *P*-value for each term was calculated by comparing the observed Likelihood ratio test statistic (LR) to the distribution generated by parametric bootstrap. Data were bootstrap sampled according to the null hypothesis where the random effect of interest is not included. The full and reduced models are then fit to each bootstrap sample to determine the distribution for the LR test statistic. A = allopatric; S = sympatric. Bold indicates significance at $P<0.05$.

Lamoille (A)				
Effect	ML est.	LR	*P*-Value	Intraclass Corr.
Female	0.0668	0.825	0.2131	0.018
Male	0.0000	0.000	0.5037	0.000
M x F	**0.2098**	**29.93**	**0.0023**	0.057
GFP male	**0.0879**	**23.88**	**0.0010**	0.024
Zion (A)				
Effect	ML est.	LR	*P*-Value	Intraclass Corr.
Female	0.3003	3.202	0.0647	0.074
Male	0.0405	0.170	0.3721	0.010
M x F	**0.3056**	**22.47**	**0.0022**	0.076
GFP male	**0.0835**	**12.21**	**0.0011**	0.020
Mt. St. Helena (S)				
Effect	ML est.	LR	*P*-Value	Intraclass Corr.
Female	0.0000	0.000	1.0000	0.00
Male	0.0184	0.096	0.4120	0.005
M x F	**0.2195**	**52.44**	**0.0019**	0.060
GFP male	**0.0825**	**35.24**	**0.0010**	0.022
Sierra (S)				
Effect	ML est.	LR	*P*-Value	Intraclass Corr.
Female	0.0000	0.000	0.3744	0.00
Male	0.0000	0.000	1.0000	0.00
M x F	**0.4139**	**70.85**	**0.0021**	0.111
GFP male	0.0077	0.902	0.0886	0.002

Supplemental Table 10. The variance in male reproductive success for allopatric (Lamoille, Zion) and sympatric (Mt. St. Helena, Sierra) populations of *D. pseudoobscura* estimated from ISC data. Empirical bootstrap intervals were calculated for each population.

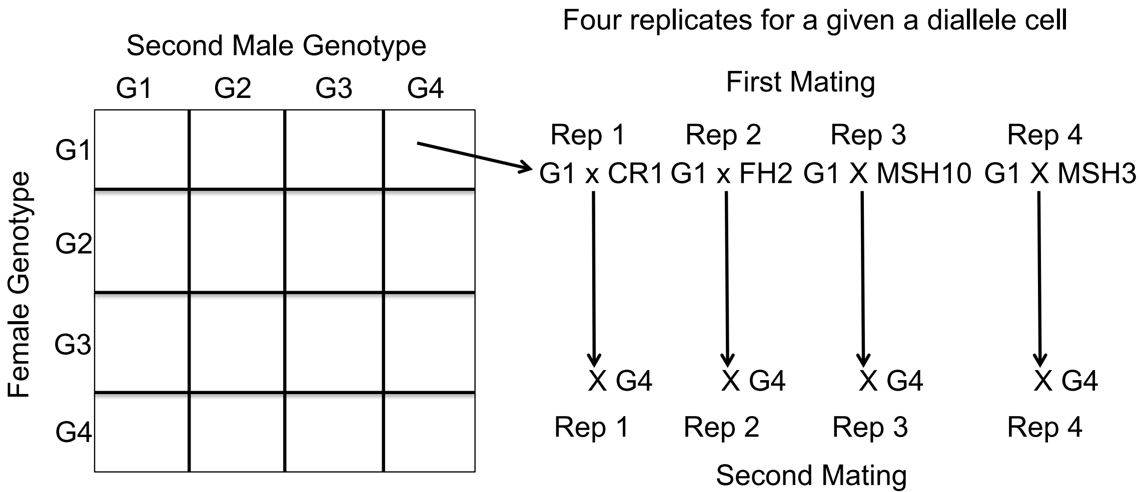
Population	Empirical Mean	Lower Bootstrap CI	Upper Bootstrap CI
Lamoille (A)	0.066	0.052	0.085
Zion (A)	0.090	0.070	0.117
Mt. St. Helena (S)	0.033	0.022	0.056
Sierra (S)	0.031	0.020	0.053

Supplemental Table 11. The pairwise comparison of sympatric and allopatric populations for mean intraspecific sperm competition (ISC). Means were compared using *t*-test and Wilcox sum rank test. Since both are significant we present the *t*-test only.

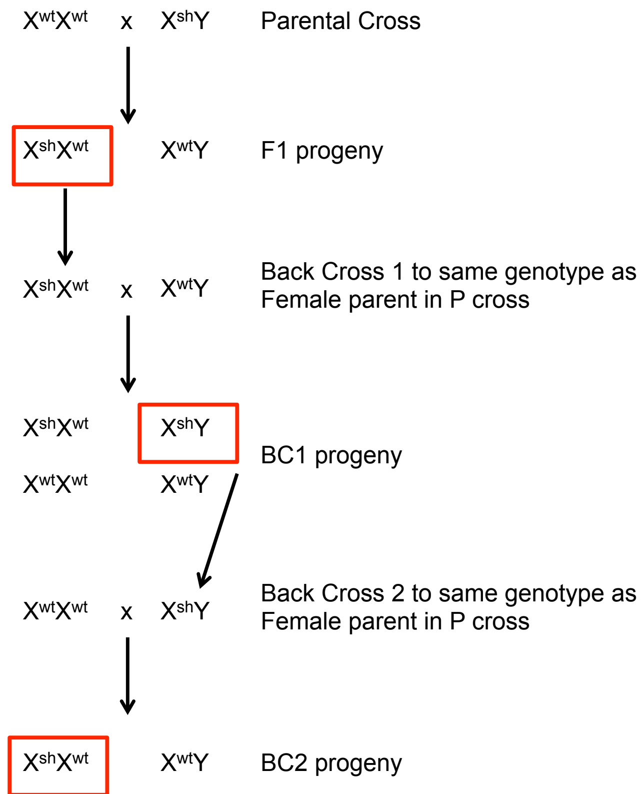
Pop Allo.	Pop Sym.	Mean ISC Allo.	Mean CSP Sym.	*t* statistic (df)	*P-value*
Lamoille	Mt.St. Helena	0.76	0.69	2.1 (99.5)	0.037
Lamoille	Sierra	0.76	0.68	2.3 (101.4)	0.018
Zion	Mt.St. Helena	0.80	0.69	2.8 (108.1)	0.004
Zion	Sierra	0.80	0.68	3.1 (110.1)	0.002

Supplemental Table 12. The pairwise comparison of sympatric and allopatric populations for variance of intraspecific sperm competition (ISC). Variances were compared using a Kruskal-Wallis modified Brown-Forsythe Levene type test (see methods).

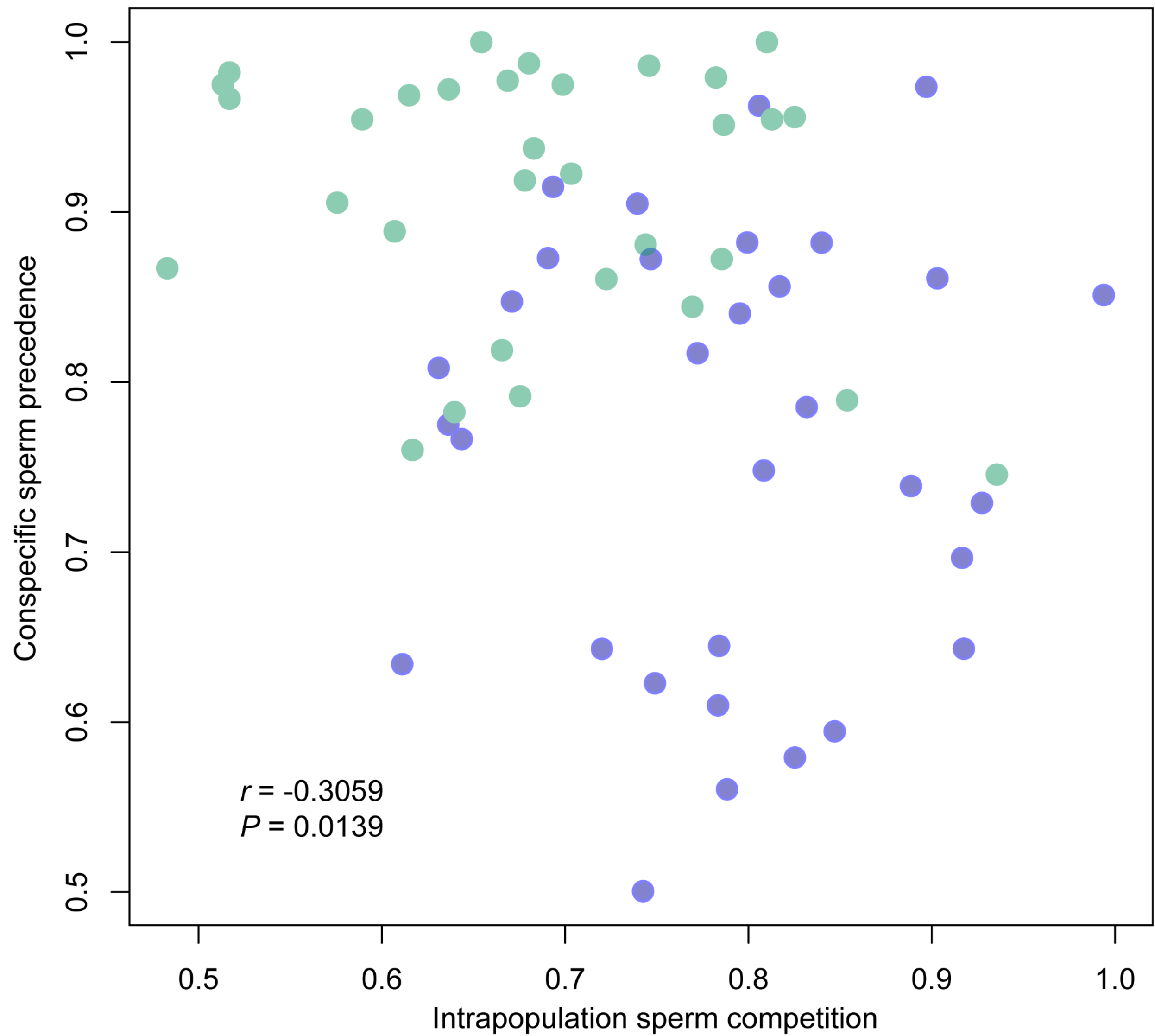
Pop Allo.	Pop Sym.	Var. ISC Allo.	Var. ISC sym	Levene type χ^2	*P-value*
Lamoille	Mt.St. Helena	0.028	0.057	5.12	0.023
Lamoille	Sierra	0.028	0.052	5.63	0.017
Zion	Mt.St. Helena	0.047	0.057	5.68	0.017
Zion	Sierra	0.047	0.052	6.03	0.014



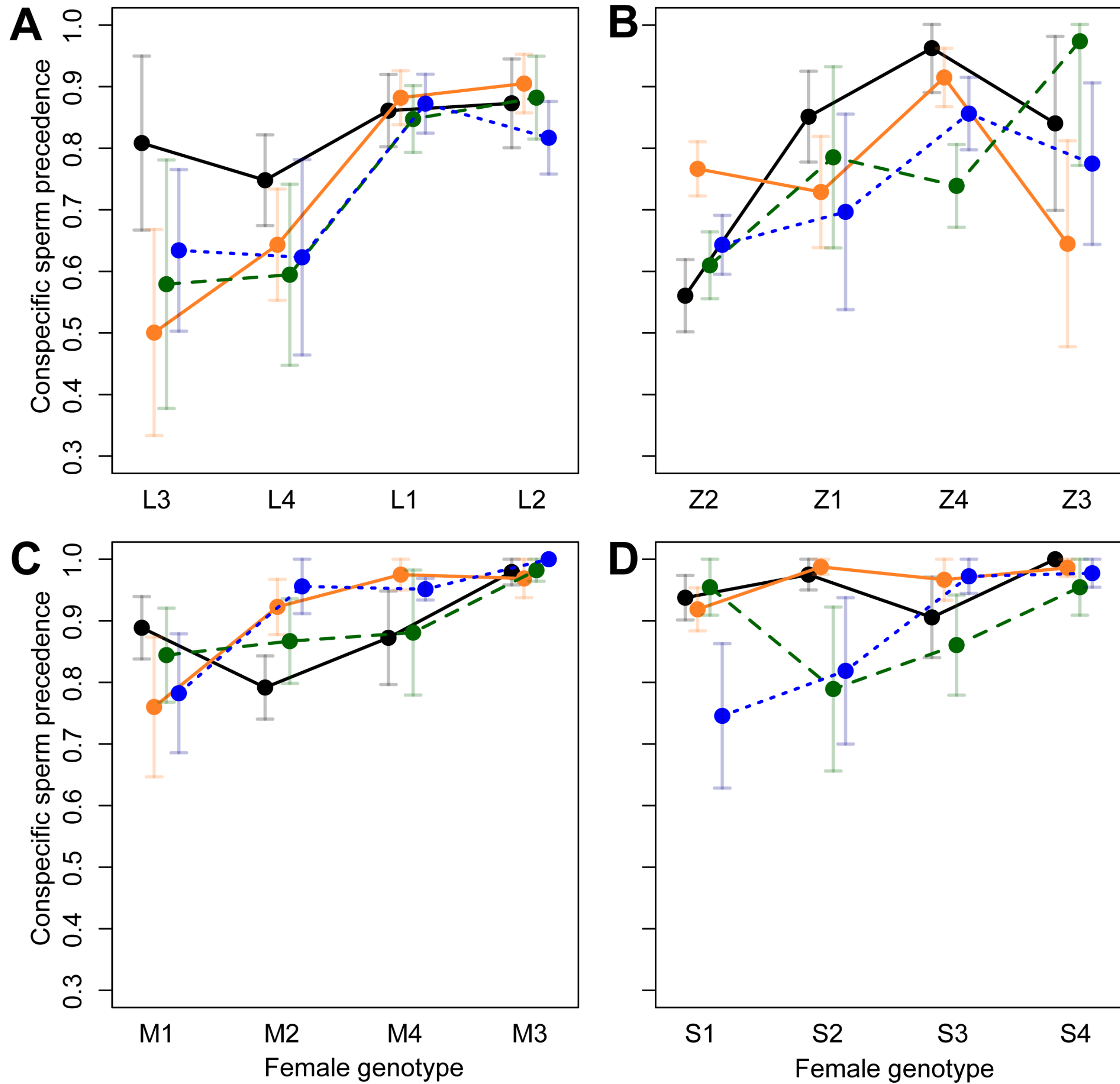
Supplemental Figure 1. The basic experimental design for the conspecific sperm precedence experiment. The grid on the left represents the 4 x 4 male-female genotype combinations that comprised the *D. pseudoobscura* diallel. Each cell represents a female genotype and a *D. pseudoobscura* genotype (the second male to mate). The diagram on the right illustrates how four replicates are completed for each diallel cell using each *D. persimilis* tester male genotype (i.e., CR1, FH2, MSH10, or MSH3) once.



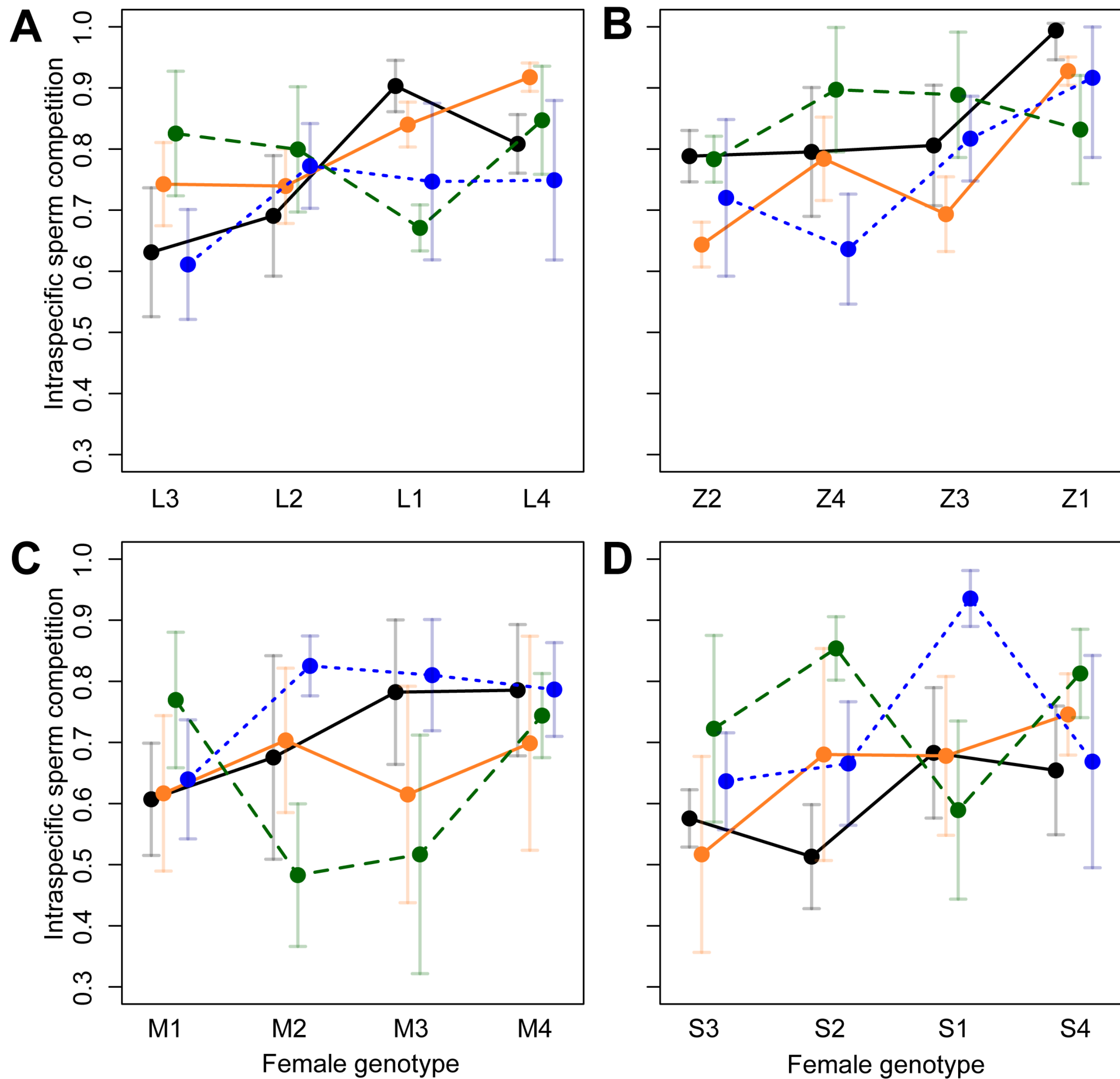
Supplemental Figure 2. A schematic diagram depicting how the *sh* allele was introgressed into the genotypes that would serve as the *D. persimilis* tester males.



Supplemental Figure 3. The negative correlation between intrapopulation sperm competition (ISC) and conspecific sperm precedence (CSP) across all four populations with each point representing a male-female genotype combination. Blue points are from allopatric populations and green points are from sympatric populations.



Supplemental Figure 4. Conspecific sperm precedence (CSP) for all male-female genotype combinations in each population demonstrating a significant effect of female genotype and male-female genotype interaction on the outcome of CSP. A) Lamoille-Allopatry, B) Zion-Allopatry, C) Mt. Dt. Helena-Sympatry, and D) Sierra-Sympatry. Each point represents a specific male-female genotype combination. Error bars are ± one standard error. Female genotypes are ordered by mean CSP. Each color represents a single male genotype for each population. Colors were re-used between each population panel, but actual second male genotypes were unique to each population.



Supplemental Figure 5. Intrapopulation sperm competition (ISC) for all male-female genotype combination in each population demonstrating a significant male-female genotype interaction on the outcome of ISC. A) Lamoille-Allopatry, B) Zion-Allopatry, C) Mt. Dt. Helena-Sympatry, and D) Sierra-Sympatry. Each point represents a specific male-female genotype combination. Error bars are ± one standard error. Female genotypes are ordered by mean ISC. Each color represents a single male genotype for each population. Colors were re-used between each population panel, but actual second male genotypes were unique to each population.