

Research Article

Associations between Variation in X Chromosome Male Reproductive Genes and Sperm Competitive Ability in *Drosophila melanogaster*

Leah Greenspan^{1,2} and Andrew G. Clark¹

¹Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853-2801, USA

²Department of Developmental Biology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065-6094, USA

Correspondence should be addressed to Andrew G. Clark, ac347@cornell.edu

Received 11 January 2011; Accepted 7 March 2011

Academic Editor: Jeremy Marshall

Copyright © 2011 L. Greenspan and A. G. Clark. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Variation in reproductive success has long been thought to be mediated in part by genes encoding seminal proteins. Here we explore the effect on male reproductive phenotypes of X-linked polymorphisms, a chromosome that is depauperate in genes encoding seminal proteins. Using 57 X chromosome substitution lines, sperm competition was tested both when the males from the wild-extracted line were the first to mate (“defense” crosses), followed by a tester male, and when extracted-line males were the second to mate, after a tester male (“offense” crosses). We scored the proportion of progeny sired by each male, the fecundity, the remating rate and refractoriness to remating, and tested the significance of variation among lines. Eleven candidate genes were chosen based on previous studies, and portions of these genes were sequenced in all 57 lines. A total of 131 polymorphisms were tested for associations with the reproductive phenotypes using linear models. Nine polymorphisms in 4 genes were found to show significant associations (at a 5% FDR). Overall, it appears that the X chromosomes harbor abundant variation in sperm competition, especially considering the paucity of seminal protein genes. This suggests that much of the male reproductive variation lies outside of genes that encode seminal proteins.

1. Introduction

In nature, *Drosophila melanogaster* females often mate with multiple partners [1–3]. Due to the female fly’s ability to store sperm, having multiple partners provides an evolutionary advantage for female flies allowing them to choose the strongest and most fit sperm to fertilize their eggs. Females’ tendency to remate has provided an opportunity for selection to operate on differential sperm success, and this opportunity is thought to have resulted in robust sperm competition among male fruit flies. In order for male flies to sire as many progeny as possible, their ejaculate must either provide a means for their sperm to outcompete other sperm in order to fertilize as many eggs as possible, or it may contain substances that discourage the female fly from remating [4]. Although recent studies have shown that males vary in their ability to gain fertilizations under

competitive conditions, the precise mechanism is unknown. Factors that influence the outcome of sperm competition and postcopulatory sexual selection include differences in sperm delivery and storage, seminal fluid composition, female egg laying rate, and female remating latency [4, 5].

Sexual antagonism occurs when there is an evolutionary advantage of a trait in one sex that reduces the fitness of the opposite sex. Since *D. melanogaster* females have the capability of storing sperm for use in fertilizations over a period of many days, they have the opportunity to produce progeny from multiple fathers. In addition to the fact that there may be differential use of the stored sperm from different males, there is an opportunity for postcopulatory sexual selection. Seminal proteins have developed in male flies that manipulate the female’s reproductive behavior, including her tendency to remate and the rate at which she lays eggs [4]. These proteins increase the fitness of the male

fly by increasing the probability of paternity; however, some of these proteins have been found to be toxic to females, resulting in a conflict, whereby what is good for the male appears to have fitness-lowering consequences to the female [4]. To compensate for this decrease in fitness, female fruit flies have developed a defense mechanism against the toxic seminal fluid. This sets up the opportunity for an arms race between the male and female. One consequence of this is that accessory gland protein-encoding genes (Acps) are among the fastest evolving genes in *Drosophila* [6].

Different strains of *D. melanogaster* vary in the degree of toxicity of the male seminal fluid and also in the degree of effectiveness of the defense system in females. Many experiments have been performed exploring the specific genes that are involved in sperm competition. In Fiumera et al. [7], a survey of 95 second chromosome extraction lines identified wide variability in all the reproductive phenotypes measured, including the extent to which the first-mating and second-mating males successfully sire progeny in double-mating contests with a standard marker male. SNPs in Acp-encoding genes and other candidate genes were genotyped across the lines, and nine significant associations were identified between polymorphisms in the male reproductive genes and sperm competition. Pleiotropy was seen between two genes: a polymorphism in the *Acp33A* gene correlates with both P1' and P2' (arcsine square root transformation of P1 and P2) and a polymorphism in the *CG17331* gene correlates with both elevated P2' and reduced refractoriness. A similar study of variation on the third chromosome of *D. melanogaster* [8] yielded four genes related to sperm competition that presented evidence of pleiotropic effects: *CG6168*, *CG14560*, *Acp62F*, and *Esterase-6*. These studies revealed extraordinary magnitudes of difference in fitness among genotypes, and even the single-gene associations reveal effect sizes on fitness of greater than 10%. This poses a challenge to understand how such a high level of intergenotype fitness variation can be maintained in the population. Fiumera et al. [8] identified some cases where the pleiotropic effects of the single nucleotide variants were antagonistic, that is, the allele that was favored by sperm competition might be disfavored in female-induced fecundity. While this is a possible mechanism for maintaining polymorphism, so far, the evidence does not support the idea that antagonistic pleiotropy is sufficiently prevalent to explain standing levels of polymorphism.

To complete a genome-wide survey of polymorphisms that mediate differences in sperm competitive ability, we sought to determine whether the X chromosome contains any genes that also can affect sperm competition. The X chromosome has additional features distinct from the autosomes that impact the tendency for variation in sperm competitive fitness to impact polymorphism. The different doses of the X chromosome in XY males and XX females lead to a dosage-compensation mechanism in males that is based on elevated transcription of X-chromosome genes [9]. However, in *Drosophila*, dosage compensation does not occur in the germ cells, leading to a lower level of expression of certain X chromosome genes [10]. Another reason to study the X chromosome is its hemizygous state in males.

Due to males' hemizygous sex chromosomes, there is no masking of recessive effects. Slightly deleterious mutations are more efficiently selected against when they are on the X chromosome, and so, their frequency under mutation-selection balance will be lower. But the lack of masking also promotes the accumulation of mutations on the X chromosome that present favorable attributes in males. An additional reason to study the X chromosome is that only one or two of the structural genes for accessory gland proteins (Acps) are found on the X chromosome. Therefore, if there is variation in sperm competition in X substitution lines, either it is not mediated by Acps, or there are additional Acps or Acp-modifying enzyme genes to be discovered on the X. Thus, the goal of this study is to investigate the variation among sperm competition of *D. melanogaster* in genetically identical flies that differ only in their X chromosomes and to determine which polymorphisms are causing this variation through sequencing and association tests.

Genes that might be involved in the process of sperm competition have been found by a number of routes. Parisi et al. [10] used an Incyte microarray to assay the expression of 14,142 transcripts that were predicted to have sex-differential expression profiles. From the total number of transcripts identified on the X chromosome, only those that had a high testis-to-ovary expression ratio were considered as possible candidate genes. Other potential candidate genes were the 27 genes identified by Drnevich et al. [11] that showed different expression levels due to within-population variation of male reproductive success under competitive conditions (MCRS). Significant differences were detected between high MCRS and low MCRS genotypes using a combination of traditional quantitative genetics and oligonucleotide microarrays. Genes on the X chromosome, whose *P* values indicated significant association between expression variation and male fertility were selected to be analyzed for polymorphisms affecting sperm competition. Another source of potential candidate genes was the result of efforts to apply mass spectrometry analysis to the sperm proteome [12–14]. There are currently 1108 proteins in the *D. melanogaster* sperm proteome that were identified using whole-sperm mass spectrometry. Of the proteins identified in these studies, those residing on the X chromosome were considered as potential candidate genes for association testing. Many proteins that were identified by mass spectrometry had unknown function, and thus could possibly play a role in sexual antagonism in flies. The data collected from Haerty et al. [6] provided other potential candidate genes. Haerty et al. applied phylogenetic analysis by maximum likelihood and examined 2505 sex and reproduction-related genes (SRR genes) across the twelve species of *Drosophila*, whose genome sequence was available. For the purposes of this study, genes found on the X chromosome that were undergoing positive selection were considered as possible candidate genes for association testing. To refine the list, we ran all candidate genes through FlyAtlas [15] to determine their expression levels in the testes and the male accessory glands compared to the whole body. Only those genes that showed high expression in either the testes or male accessory glands were examined further.

2. Materials and Methods

2.1. *Drosophila* Stocks. Balancer stocks were used to construct 57 X chromosome substitution lines, whose X chromosomes were derived from a wild population in Ithaca, NY, by Dr. Erin Hill-Burns, and details of the line construction can be found in Hill-Burns and Clark [16]. These lines, which were used in all phases of this study, are coisogenic for the second, third, and fourth chromosomes, but each has a unique homozygous X chromosome. The experimental flies had the red wild-type eye color. In the sperm competition assays, tester flies were a standard *cn bw* lab strain, which is phenotypically white eyed. Lines were raised at 24°C on 12 h light/dark cycle using standard agar-dextrose-yeast medium.

2.2. Scoring Sperm Competition Phenotypes. The methods for scoring sperm competition phenotypes follow closely the methods of Fiumera et al. [8] and Baker et al. [9]. *cn bw* females were mated to two different males in succession in order to determine the extent of sperm competition among the lines of flies. In the “defense” crosses, the first male to mate were males from the X-replacement lines, and the second males were *cn bw*. P1 is the sperm competition index for these crosses, and it scores the proportion of progeny sired by the first (X-replacement line) males. In the “offense” crosses, the first males to mate were *cn bw*, and the second males were from the X-replacement lines. P2 is the sperm competition index for these crosses, and it scores the proportion of progeny sired by the second (X-replacement line) males. We also scored the fecundity of the females in the defense crosses (when the experimental male is first to mate) and offense crosses (when the experimental male is second to mate), where fecundity is defined as the total count of offspring produced by each female across all three vials. The remating rate is the proportion of females that had mated to a *cn bw* male and subsequently remate with an experimental male. The final statistic is refractoriness to remating, which expresses the proportion of females that mate once with an experimental male and then fail to remate.

In the offense experiment, virgin *cn bw* females were first mated to virgin *cn bw* males in a mass mating on the morning of day one. In the afternoon of day one, individual females were transferred, without anesthesia, to vial one, and the males were discarded. In the afternoon of day three, two virgin experimental males were added to vial one. The next day, the females were transferred to vial two. On day seven, the female was transferred to vial three and allowed to lay eggs until day 11 when she was discarded. The progeny were then scored for eye color to determine paternity at 15, 17, and 19 days after the initiation of egg laying for vials one, two, and three, respectively. For the defense experiment, the experimental male was the first male to mate. On the first day, the *cn bw* females were individually mated with the experimental X substitution line males. On day three, the females were transferred to vial two, and the males were discarded. On day four, two *cn bw* males were added to vial two. By day six, the females were transferred to vial three, and the males were discarded. The females were discarded on day nine, and the progeny were scored for eye color at 15, 17, and

19 days after the initiation of egg laying for vials one, two, and three, respectively. The sources of variation were tested for statistical significance across both the offense and defense experiments. General linear models were used to assess the significance of among-line variability in the face of block and replicate variation, using the R statistics package *lme4* (<http://www.r-project.org/>). Permutation tests based on the χ^2 statistics were used to test for significant associations with SNP genotypes.

2.3. Sequencing of Candidate Genes. A total of eleven male reproductive candidate genes on the X chromosome (*CG15208*, *CG15200*, *CG4349*, *CG15035*, *CG12470*, *CG9806*, *CG1559*, *CG9156*, *CG17450*, *CG32819*, and *CG32820*) were chosen for DNA sequencing to test for association between any sequence polymorphisms and the degree of sperm competition. Genes were chosen based on (1) microarray data from Parisi et al. [10] examining global gene expression in *D. melanogaster*, (2) quantitative genetics and oligonucleotide microarrays from Drnevich et al. [11] that investigated male reproductive success under competitive conditions, (3) whole-mass sperm spectrometry from Dorus et al. [12] and Findlay et al. [13] that identified proteins of the *D. melanogaster* sperm proteome, and (4) phylogenetic analysis by maximum likelihood from Haerty et al. [6] that examined sex and reproduction-related genes across all twelve species of *Drosophila*. All candidate genes were run through the database FlyAtlas [15] to determine their expression levels in the testes and male accessory glands (MAGs) compared to the whole fly. Nine of these genes demonstrated high expression in the testes, while two genes showed high expression in the male accessory glands (Table 1).

Primers were designed for each gene using Primer 3 [17]. PCR products ranged from about 700–1000 base pairs long. Coding regions of each gene were flanked by these primers and amplified using PCR. The PCR product was then sequenced using Applied Biosystems Automated 3730 DNA Analyzer with Big Dye Terminator Chemistry and AmpliTaq-FS DNA Polymerase. Reference sequences for each gene were obtained using the FlyBase database [18], and the contigs for each primer set were assembled and aligned according to these references. Sequencher 4.6 (Gene Codes Corporation, 2007) was used to detect polymorphic nucleotides and indels within the aligned contigs.

2.4. Statistical Methods. Statistical tests for the fecundity of the offense and defense experiment, P1', P2', remating rate, and refractoriness were performed using the R statistical package version 2.01. General linear models were used to assess the significance of heterogeneity among the different lines for each statistical test. For the tests of sperm competition, the model was

$$Y_{ijk} = \mu + L_i + B(L)_{ij} + \varepsilon_{ijk}, \quad (1)$$

where Y_{ijk} is the response variable (P1', P2', or fecundity), μ is the grand mean, L_i is the effect of the i th line, $B(L)_{ij}$ is the block effect (the design was a balanced nested design), and ε_{ijk} is the error term, quantified by the among replicate

TABLE 1: Candidate male reproductive genes on the X chromosome. Criteria for selecting genes for DNA sequencing and association testing are presented. Testes/whole body and male accessory gland (MAG)/whole body columns present ratios of expression from FlyAtlas data.

Gene	Testes/whole body	MAG/whole body	Predicted function ^c
<i>CG15208</i> ^{a,c}	14.2	0.1	Unknown
<i>CG15200</i> ^{a,b}	10.4	0	Unknown
<i>CG4349</i> ^{a,d}	9.5	0.3	Iron ion binding (<i>Fer3HCH</i>)
<i>CG15035</i> ^a	15.9	0.1	Protein phosphatase
<i>CG12470</i> ^{a,d}	7.6	0	Unknown
<i>CG9806</i> ^c	0.5	92.5	Membrane alanyl aminopeptidase
<i>CG1559</i> ^c	0.7	1.00	DNA/RNA helicase (<i>Upf1</i>)
<i>CG9156</i> ^{c,d}	10.9	0	Protein phosphatase 1 (<i>Pp1-13c</i>)
<i>CG17450</i> ^d	14.4	0	Microtubule associated
<i>CG32819</i> ^d	14.4	0	Microtubule associated
<i>CG32820</i> ^d	14.4	0	Microtubule associated

Gene candidacy determined from: ^aParisi et al. (2003) [10], ^bDrnevich et al. (2004) [11], ^cHaerty et al. (2007) [6], ^dDorus et al. (2006) [12], ^eFlybase (<http://www.flybase.org>).

variance. The linear model was fitted with the lme4 package of the statistics program R (<http://cran.r-project.org/>), and homogeneity of variances, approximate fit of residuals to normality, and absence of outliers were all assessed with R utilities. Tests of heterogeneity in remating rate and refractoriness were essentially tests of equality of proportions under binomial sampling, and were tested by Chi-square. The association between each SNP and the reproductive phenotypes was tested with linear models, testing the null hypothesis that the phenotypes of the lines bearing the alternative two alleles for each SNP were sampled from the same distribution. Because many tests were performed for each phenotype, we report the results of these tests as the false-discovery rates or q -values [19].

3. Results

3.1. Reproductive Phenotype Variation. All 57 X-chromosome lines were successfully scored in both the offense and the defense experiments. Twenty replicate double matings were attempted for each line; however, not all lines yielded all twenty replicates with doubly mated females. Overall 57,919 progeny were scored from the 57 lines for the defense experiment, and 92,467 progeny were scored from the 57 lines for the offense experiment. For the defense experiment in which the X experimental male was presented first, a total of 1130 females successfully mated, with 531 double mating and 599 mating only once. For the offense experiment in which the X experimental male was presented second, a total of 1172 females successfully mated, with 645 double mating and 527 mating only once.

Six statistics were recorded and analyzed for each line. The first two statistics were P1 and P2, which are the proportion of offspring fathered by the experimental X chromosome males when he is either the first or second male to mate with the female. The terms P1' and P2' refer to the arcsine square root transformed values of P1 and P2. This transformation is done to stabilize variances, which otherwise become inherently small when values approach

zero or one. The next two statistics were the fecundity of the offense and defense experiments, which is the total count of offspring produced by each doubly mated female. The “remating rate” is defined as the proportion of females that remate when the experimental male is second to mate. Finally, “refractoriness” expresses the proportion of females that do not remate after having mated once with an X-line experimental male.

The results of the linear model fits revealed significant genetic variation among the X replacement lines in P1', P2', and fecundity of the offense experiment (Table 2). There was no significant variation in fecundity for the defense experiment across the 57 X substitution lines ($P = .5265$). In this case, it appears that genotype does not affect female fecundity across the lines when the X experimental males are first to mate. Fecundity showed a good fit to the normal distribution, and linear models confirmed that the fecundities of females across the defense crosses were homogeneous.

Results for the fecundity counts of the offense experiment differed slightly from those from the defense experiment. For one, fecundity of the offense experiment showed significant variation among the 57 *D. melanogaster* lines ($P = 1.98 \times 10^{-8}$). Thus, in this case, genotype does seem to affect the difference in fecundity seen across lines.

Considering sperm competitive metrics, the P1' statistic shows highly heterogeneous variation among the lines ($P = 1.38 \times 10^{-6}$) indicating that genotype does have a significant effect on the proportion of offspring fathered by the experimental X chromosome males when he is the first to mate. Figure 1 shows the distribution of the P1 (and P2) line means, showing a mode for the line mean of P1 around 0.4. Figure 2 shows a boxplot of the P1 statistics stratified by line, giving a visual confirmation of the linear model that the lines are highly heterogeneous.

The P2' statistic was found to be highly significant across the 57 X substitution lines ($P = 2.2 \times 10^{-16}$), thus indicating that genotype has an effect on the variation seen for the proportion of progeny sired by the experimental X

TABLE 2: Reproductive phenotype statistics. P1' is the proportion of progeny sired by the first male in defense crosses (arcsine transformed). P2' is the proportion of progeny sired by the second male in offense crosses (arcsine transformed). Summary of the reproductive phenotypes statistics using linear models, a total of 57,919 progeny were counted for the defense experiment, and 92,467 progeny were counted for the offense experiment.

Phenotype	Mean across lines	Range of line means	D.F.	Test statistic	P value
Fecundity-D	80.69	52–103.583	56	$F = 0.98$.5265
P1'	0.8666	0.53088–1.5708	56	$F = 2.25$	1.38×10^{-6}
Fecundity-O	110.79	76.125–153.75	56	$F = 2.53$	1.98×10^{-8}
P2'	1.01	0.27237–1.49932	56	$F = 7.12$	2.20×10^{-16}

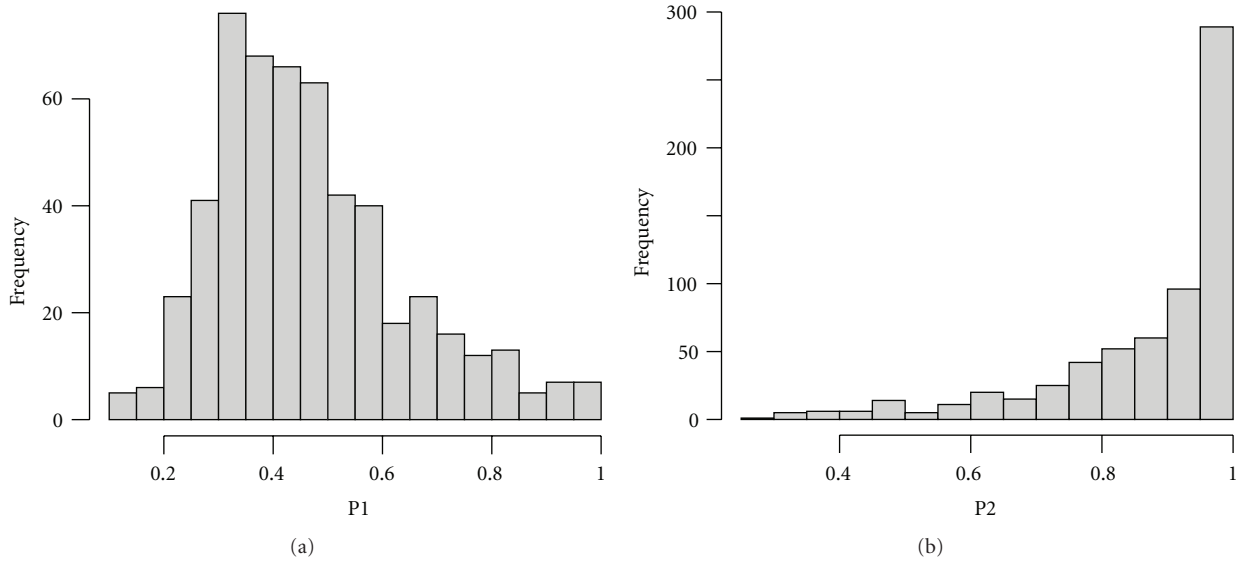


FIGURE 1: The distribution of P1 and P2 across 57 X substitution lines. P1 is the defense test, where the X-line experimental male was the first to mate (followed by a *cn bw* control male), and the P2 assay uses the reverse order. P2 tests the ability of the X experimental males to displace resident sperm. In both cases, arcsine transformation was applied before statistical testing.

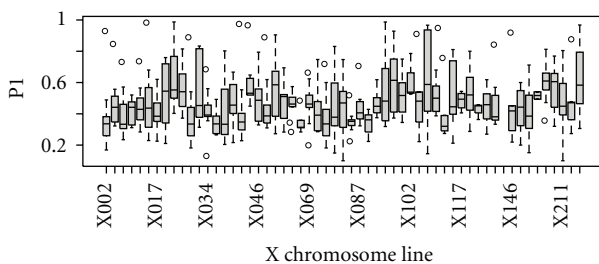


FIGURE 2: Boxplot of the P1 scores of the 57 X replacement lines. The proportion of offspring sired by the X experimental male out of 57 naturally derived X chromosomes of *D. melanogaster* when the respective X-line male is the first to mate (followed by the *cn bw* control male). P1 scores are highly heterogeneous across lines ($P = 1.38 \times 10^{-6}$).

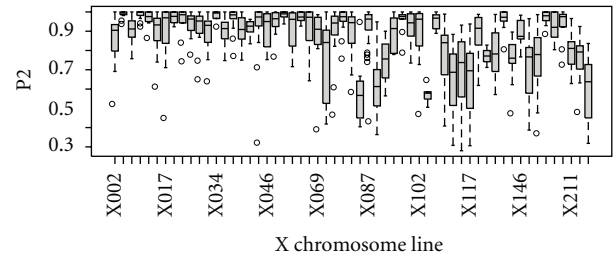


FIGURE 3: Boxplot of the P2 scores of the 57 X replacement lines. The proportion of offspring sired by the X experimental male out of 57 naturally derived X chromosomes of *D. melanogaster* when the respective X-line male is the second to mate (after the *cn bw* control male). P2 scores are highly heterogeneous across lines ($P = 2.2 \times 10^{-16}$).

chromosome male when he is the second to mate. The right panel of Figure 1 displays the distribution of the P2, indicating a highly skewed distribution (statistical analysis was performed on arcsine transformed data). Figure 3 presents a boxplot showing the heterogeneity across lines in the P2 statistic.

Next, we examined the variation among lines in the mating success rate of second males (in the offense crosses), and the ability of the first-mating males to induce refractoriness in their female mates (in the defense crosses). A Chi-square test found that remating rate showed statistically significant heterogeneity across lines ($P < .001$).

TABLE 3: Remating rate and refractoriness. Summary of the remating rate and refractoriness using Chi-square analysis.

Reproductive phenotype	Mean value across lines	Range of line means	D.F.	Chi-square statistic	<i>P</i> value
Remating rate	0.55	0.2105–0.8000	56	94.809	<.001
Refractoriness	0.53	0.0000–0.7500	56	97.072	<.001

This indicates that the different male genotypes among the lines varied in their ability to induce already mated females to remate. Refractoriness, on the other hand, is the proportion of females that do not remate after having mated with an X experimental male. This phenotype also was statistically significantly heterogeneous ($P < .001$) indicating that there is heterogeneity among lines in the ability of males to induce refractoriness among females. Table 3 shows the average remating and refractoriness rates as well as their significance values.

Correlation tests were performed to determine if certain reproductive phenotypes were significantly correlated with each other (Table 4). There was a significant correlation seen between fecundity of the defense experiment and fecundity of the offense experiment ($P = .0098$). These two parameters showed a positive linear correlation indicating that females who had a higher number of offspring when they mated with the X experimental males first tended to also have a higher number of offspring when they mated with the X experimental males second. Additionally, the fecundity of the offense experiment demonstrated a statistically significant correlation with the transformed P2 data ($P = 7.58 \times 10^{-7}$). The strong positive correlation implies that females that had a higher number of progeny when they mated with the X experimental second males also had a higher proportion of progeny from that second male. Furthermore, fecundity of the defense experiment also showed a significant correlation P2' ($P = .0134$). This correlation suggests that the higher number of offspring sired in the defense experiment, the larger the proportion of offspring sired from the X experimental males in the offense experiment. A significant correlation was also seen between the P1' and refractoriness ($P < .0001$). This strong negative correlation suggests that males with strong defense induce a higher portion of females to be refractory to remating. Another statistically significant correlation was found between the fecundity of the offense experiment and the remating rate ($P < .001$). This correlation indicates that as the fraction of females that tend to remate increases, the number of progeny sired in the offense experiment increases. A final correlation observed among the reproductive phenotypes was between P2' and remating rate ($P < .0001$ resp.). This strong positive correlation suggests that males that have high P2, and thus strong sperm competition offense, were much more successful at inducing already mated females to remate. None of the other correlations were significant.

3.2. Identifying SNPs and Association Tests. A total of 131 SNPs and indels were found within the 11 candidate genes using Sequencher 4.6. Three of the 11 genes (*CG15200*, *CG4349*, and *CG15035*) were fully sequenced, while the rest

TABLE 4: Correlations Between Reproductive Phenotypes. Correlation coefficients and *P* values for correlations among line means of reproductive phenotypes.

Variables	Correlation coefficient	<i>P</i> value
Fec-O versus Fec-D	0.3395	.0098
P1' versus P2'	0.0411	.7614
Fec-O versus P2'	0.6013	7.58×10^{-7}
Fec-D versus P2'	0.3258	.0134
P1' versus refractoriness	-0.483	<.0001
Fec-O versus remating	0.434	<.001
P2' versus remating	0.695	<.0001

had only portions of the genes sequenced. Association tests were run using linear models in the R statistical software to determine if there were any significant associations between the SNPs and indels, and the reproductive phenotypes.

Nine significant associations were found between the polymorphisms and the reproductive phenotypes ($q < 0.05$). Of the nine associations, two were located in the 3'UTR region, two were located in intronic regions, and five were located in the coding regions. The variants in the coding regions consisted of one nonsynonymous mutation, one synonymous mutation, and two frameshifting deletions. One frameshift deletion was found in gene *CG9156* position 402, associated with both fecundity of the defense experiment and the P1' statistic. Table 5 displays all the significant polymorphism-phenotype combinations.

Two polymorphisms and a deletion were found to associate significantly with fecundity of the offense experiment. Two genes, *CG15200* and *CG15208*, each had one polymorphism, and there was a deletion allele in gene *CG17450*. Alleles of these genes associated with either an increase or decrease in the total amount of progeny sired when the X experimental male was the second to mate. Since high fecundity is related to a higher fitness, these alleles can be considered either advantageous or disadvantageous. The polymorphisms *CG15200* SNP 471 and *CG17450* deletion 487-488 showed a significant decrease in fecundity ($q = 0.008$ and $q = 0.048$ resp.), while *CG15208* SNP 1295 showed a significant increase in fecundity of the offense experiment ($q = 0.012$). These three SNPs explained 11.5%, 19.3%, and 12.8% of the interline variance in fecundity in the offense crosses.

Two polymorphisms were found to be associated with changes in P2'. Because P2 measures the paternity success (it is the proportion of offspring sired by the second-mating male), it is a direct proxy measure for reproductive fitness. The first associated polymorphism was found in gene *CG15208* at position 1019, while the second polymorphism

TABLE 5: Significant Associations between reproductive phenotypes and DNA sequence polymorphisms found in each gene. q -value is derived from the linear model testing homogeneity of the phenotype across lines stratified by the targeted SNP.

Gene name	Association	Index number	q -value	Location in gene	Mutation	Amino acid change
CG15200	Fecundity-O	471	0.0079	3' UTR		
CG15208	P2'	1019	0.0200	Coding	Nonsyn	Serine → Proline
CG15208	Fecundity-O	1295	0.0121	3''UTR		
CG17450	P2'	210	0.0482	Coding	Synon	
CG17450	Fecundity-O	487-488	0.0478	Intron		
CG17450	Fecundity-D	788-795	0.0146	Intron		
CG9156	P1'	159-160	0.0328	Coding	Frameshift	
CG9156	Fecundity-D	402	0.0182	Coding	Frameshift	
CG9156	P1'	402	0.0032	Coding	Frameshift	

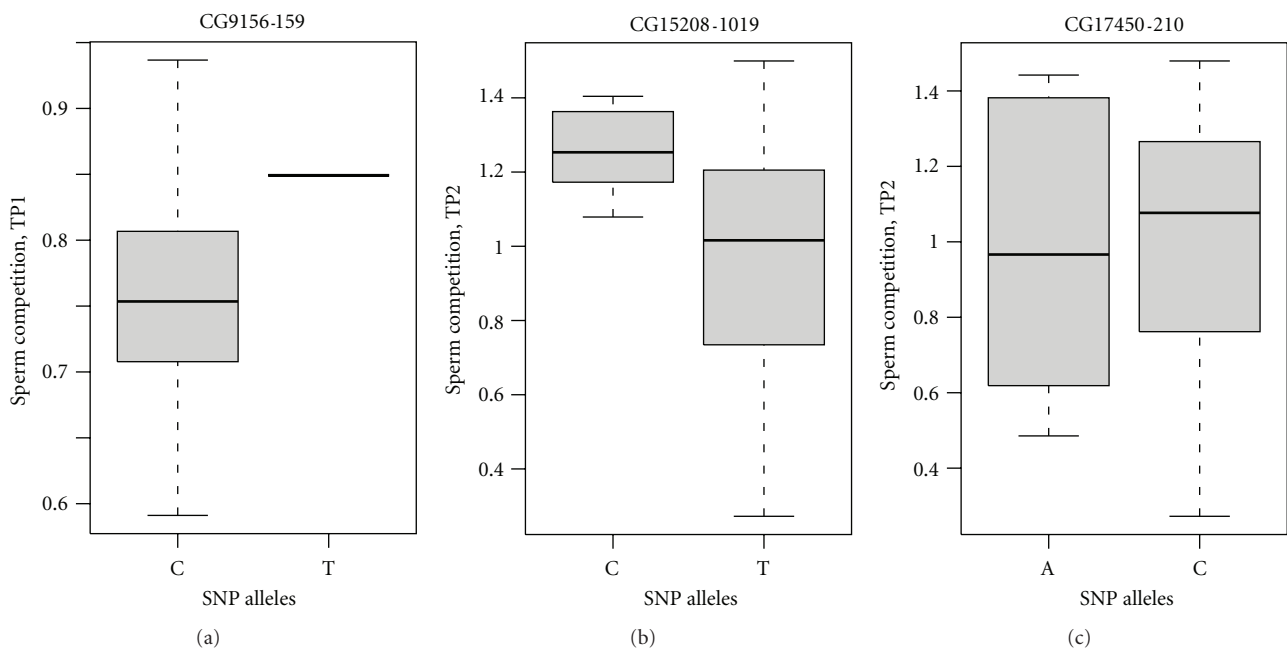


FIGURE 4: SNP allelic associations with P1 and P2. (a): associations between a deletion found on the gene *CG9156* and the arcsine square transformation of the proportion of progeny sired by the X experimental male when he was the first to mate. Gene *CG9156* deletion 159-160 had 21 lines with a deletion in both positions, 4 lines with a deletion in just the second position, and 14 lines with the reference sequence AA. This deletion showed a statistically significant decrease in the P1' statistic ($q = 0.033$). (b): for *CG15208* position 1019, there were 7 lines with the rare allele C and 48 lines with the common allele T. Those lines containing the rare allele showed a significant increase in P2' ($q = 0.020$). (c): for *CG17450* position 210, 4 lines had the rare allele G, while 41 lines contained the common allele C. This rare allele also demonstrated a statistically significant increase in P2' ($q = 0.048$).

was found in gene *CG17450* at position 210. Both polymorphisms were low frequency alleles that showed an increase in P2'. Since both fecundity of the offense and fecundity of the defense experiment were significantly correlated with P2', we can conclude that these low frequency alleles may be advantageous, since in the lab environment, they lead to an increase in fitness. This increase in fitness is a result of the increase in the number of offspring from the X experimental male when he is second to mate. Even though the female already mated, she has a tendency to use more of the second male's sperm if he possessed the novel allele. The right two panels of Figure 4 illustrate the difference in P2' across

genotypes of the associated SNPs. These two SNPs explained 15.4% and 19.8% of the interline variance in P2.

Two mutations were found to associate with fecundity in the defense experiment. One polymorphism was found in gene *CG17450*, and the SNP was immediately flanked by a seven basepair deletion. This mutation demonstrated an increase in the total number of progeny sired in the defense experiment, and it explained 13.4% of the interline variance in fecundity defense. Since an increase in fecundity is related to an increase in fitness, these rare alleles and deletion can be considered advantageous due to the increased reproductive fitness they confer. Gene *CG9156* position 402 contained

a deletion that was significantly associated with the fecundity of the defense experiment. This SNP explained 3.96% of the interline variance in fecundity in the defense crosses.

Two deletions were found to associate with $P1'$, the sperm competition parameter for first-mating males. One deletion was found in gene *CG9156*, at position 159. The leftmost panel of Figure 4 depicts the association between this polymorphism and $P1'$. The deletion allele appears to be associated with a decrease in $P1'$, and it explained 17.8% of the interline variance in $P1'$. No other associations with this deletion allele were detected, so the reduced $P1'$ alone suggests that it has a net negative effect on fitness. Gene *CG9156* at position 402 also contained a deletion that was significantly associated with the $P1'$ statistic, and it explained 3.9% of the interline variance in fecundity.

4. Discussion

The overall magnitude of variation seen among these 57 X substitution lines is roughly consistent with what was found in similar studies of the second and third chromosomes [7, 8]. However, these results are still surprising, given the paucity of seminal protein genes found on the X chromosome [20]. Many of the attributes of male reproductive fitness varied significantly across this set of X-chromosome extracted lines, including fecundity of the offense experiment, $P1'$, $P2'$, remating rate, and refractoriness to remating. Thus male genotype seems to have a significant effect on all these reproductive phenotypes. Although, fecundity of the defense experiment did not show significant variation across lines, two mutations were found to have a significant association with the male-induced differences in fecundity.

The different reproductive phenotypes that were scored in our collection of X-chromosome replacement lines showed a strong pattern of pairwise correlations. The pattern of correlations overall shows exactly the opposite of what we would call antagonistic pleiotropy. The alleles that were found to be best for one reproductive phenotype tended to be best for all. In such a situation, one would expect the alleles with this best-at-everything phenotype to become fixed in the population, begging the question of what maintains this high level of genetic variation in these important fitness traits.

DNA sequence analysis of the 57 X substitution lines using simple statistical tests allowed us to determine if there were polymorphisms among the candidate genes that were associated with variation in aspects of reproductive fitness. A total of 131 mutations including single nucleotide polymorphisms, insertions, and deletions were found among the 11 candidate genes that were analyzed. Association tests were performed between these mutations and the varying reproductive phenotypes seen in the X experimental males. Although association tests with these polymorphisms do not prove causation of the reproductive phenotypes, they do give insight into evolutionary changes that may be occurring, and thus prove important to study.

Of the 131 mutations identified, nine were found to significantly associate with specific phenotypes. These mutations were found among four genes which contain

naturally occurring polymorphisms that are associated with differences in mating. *CG15208* showed two significant associations with reproductive phenotypes, but there is as yet no identified known or predicted function. The first was at position 1019 in the coding region. This polymorphism caused a nonsynonymous mutation changing a serine into a proline, and was found to have an association with the proportion of progeny sired by the X experimental male when he was the second to mate. The other polymorphism found in this gene was at position 1295 and located in the 3'UTR region. There was found to be a statistically significant association between this polymorphism and the fecundity of the offense experiment. *CG15208* was originally chosen to be analyzed due to its high testis-to-ovary expression ratio, as determined by Parisi et al. [10], and its high level of expression in the testes compared to the whole body (14.4), determined by FlyAtlas [15]. Furthermore, this gene is one of the sex and reproduction-related genes that was found to be undergoing positive selection [6]. Thus, this gene has been shown to be important in the variation of reproductive phenotypes. However, more investigation is necessary to understand the mechanisms behind these polymorphisms that lead to phenotypic variation.

CG15200 does not have a known function but was chosen as a candidate gene because of its high testis-to-ovary expression ratio determined by Parisi et al. [10] and the significant difference seen in male reproductive success under competitive conditions [11]. This gene was also shown to have a higher expression level in the testes than in the whole body (10.4). One significant polymorphism found in this gene was at position 471 in the 3'UTR region, where many variants influence the transcript abundance of a gene. This polymorphism was shown to be significantly associated with fecundity in the offense experiment.

In *CG17450*, several mutations were found that associated with three different reproductive phenotypes, thus making this gene the most pleiotropic of all the genes analyzed. *CG17450* is known to be associated with microtubule binding and microtubule cytoskeleton organization and biogenesis, but at present, we do not have any understanding of the mechanism whereby polymorphisms in this gene should mediate fitness differences. However, this gene was chosen as a good candidate gene, because it was shown to be part of the *Drosophila* sperm proteome [12–14] as well as to have a high expression level in the testes compared to the whole body (14.4). The first polymorphism identified in this gene was at position 210 in the coding region. This SNP caused a synonymous mutation but was shown to affect the proportion of progeny sired by the X experimental male when he was the second to mate. The second mutation seen in this gene was a deletion in the intronic region of the gene at position 487–488. This deletion seemed to relate to a decrease in the fecundity of the offense experiment. The final mutation identified in this gene was an SNP followed by a deletion at positions 788–795. This mutation was also found in the intronic region of the gene but was associated with an increase in the fecundity of the defense experiment.

The fourth gene that showed segregating variation that was significantly associated with reproductive phenotypes

TABLE 6: Primers used to amplify the desired region of each of the 11 candidate genes.

Gene name	Forward primer (F)	Reverse primer (R)	PCR product
<i>CG15208</i>	CGCGATATGATCGAACAGTG	TTGCAAAATGCGAATAGCTG	787 bp
<i>CG15200</i>	TCCTTCAAAGTCAATGCTGAGA	TTGCAGTGAAGCACAAAACA	785 bp
<i>CG4349</i>	CGACCCAGTTGCCTTTTAGA	AAACAATTGATCGGCCAAAA	852 bp
<i>CG15035</i>	GATGTGACACCCCTTCGATT	CACTTTACCAGATCGCACCA	839 bp
<i>CG15035</i>	GCAGCTGTACGGCGTGTAT	TGACCACTTTGTTTGGAGAGA	850 bp
<i>CG12470</i>	GATCACGTTGCTTTGCTGAA	ACGAAGAAGAGCCGGGTAAT	749 bp
<i>CG9806</i>	TATTTCCGGGAGAGCAGTCGT	TGATGGAGAAGGCCAATTC	821 bp
<i>CG1559</i>	GCAGATGGACGGTTTGCTAT	TGTTTTCGCTGATTTTCCTTG	972 bp
<i>CG9156</i>	TCCATGAAACACGAACTGGA	GTCTGGAGAATGGGACTGGA	709 bp
<i>CG17450</i>	AGAAGAAGCCACCGAGGACT	AAAGTGGCCTACTCCGATTG	823 bp
<i>CG17450</i>	ATATCGCATCAGCCATGTCA	CCGAGAGGTTGCTGAAAGAC	945 bp
<i>CG32819</i>	GCCAGGCTACCGTTAATCA	GAATTTGGACTGGGTGTTGG	987 bp
<i>CG32820</i>	GGCCCACATGCTAATCAACT	ATGACATGGCTGATGCGATA	913 bp

was *CG9156*. This gene is known to be a protein phosphatase and is important for protein amino acid dephosphorylation. It was chosen as a good candidate gene to analyze because it was shown to be a sex and reproduction-related gene that was undergoing positive selection [6], and it is clearly expressed as part of the sperm proteome [12]. This gene also demonstrates an 11-fold increase in expression in the testes compared to the whole body. *CG9156* was found to have two deletions that caused frameshift mutations that affected different reproductive phenotypes. The first deletion is at positions 159-160 and is located in the coding region. This deletion is associated with a decrease in the proportion of progeny sired by the X experimental male when he is the first to mate. The second deletion in this gene is also located in the coding region and is found at position 402. This mutation was found to associate with the fecundity of the defense experiment and also with the proportion of offspring sired by the X experimental male when he is the first to mate.

The other seven genes analyzed did not show significant associations between their mutations and variation in the reproductive phenotypes. However, all of these genes contain important functions that if disabled could potentially alter reproductive phenotypes. *CG4349* is the ferritin 3 heavy chain (gene symbol *Fer3HCH*) and is known to be involved in ferrous and ferric iron binding, while *CG15035* is inferred to be a protein phosphatase. *CG12470* has no known function but is found in the sperm proteome [12] and is shown to have a high testis to ovary expression ratio [10]. Gene *CG9806* is involved in alanyl aminopeptidase activity and gene *CG1559* is needed for DNA/RNA helicase activity as well as nucleic acid binding. The last two genes *CG32819* and *CG32820* are both associated with microtubule function. Although these genes did not contain polymorphisms that significantly associated with reproductive phenotypes, they may still have an important effect on the phenotypes. Most of the genes were not sequenced in their entirety, and it is possible that significant effects of sequence variants within large introns were missed.

Due to a lack of accessory gland proteins on the X chromosome, it is tempting to assume that differences in

the X chromosome would not cause differences in sperm-competition phenotypes. However, this study has identified several mutations on the X chromosome that are associated with either fecundity, remating tendencies, or measures of sperm competition. Perhaps it is the hemizygous state of male flies that is allowing recessive polymorphisms that would normally be masked to be seen or the differences in dosage compensation that occurs in the X chromosome that is causing these changes. It is also possible that we have not yet been exhaustive in our discovery of accessory gland proteins and that there are yet to be discovered Acps on the X. But, we can be confident that the candidate genes that had significant SNP associations are not Acps, because of their absence of expression in the male accessory gland. However, these genes could affect Acps as modifying enzyme genes that affect development of function of accessory glands. Whatever the cause, the data indicates that genotypic variation on the X chromosome does affect reproductive phenotypes including sperm competition. Genes *CG15208*, *CG15200*, *CG17450*, and *CG9156* are good candidates for further investigation seen as they all contain mutations that are significantly associated with different reproductive phenotypes. Finally, we note that the significant effects that were found in this study were obtained in a particular context, and while we have confidence that the genes involved have effects in that context, changes in many experimental attributes, including the genotypes of other competing flies in the assays, may alter the sperm competitiveness rankings in unpredictable ways [21].

Appendix

The forward and reverse primers utilized can be found listed in Table 6.

Acknowledgments

The authors thank Xiaoyun Wang, Sean Hackett, Anna Beavis, and Scott Pilla for technical assistance. This work was

supported in part by NIH Grant no. RO1 HD059060 to A. G. Clark and Mariana Wolfner and NSF DEB-0743125 to A. G. Clark and Anthony Fiumera.

References

- [1] R. C. Griffiths, S. W. McKechnie, and J. A. McKenzie, "Multiple mating and sperm displacement in a natural population of *Drosophila melanogaster*," *Theoretical and Applied Genetics*, vol. 62, no. 1, pp. 89–96, 1982.
- [2] L. G. Harshman and A. G. Clark, "Inference of sperm competition from broods of field-caught *Drosophila*," *Evolution*, vol. 52, no. 5, pp. 1334–1341, 1998.
- [3] M. Imhof, B. Harr, G. Brem, and C. Schlötterer, "Multiple mating in wild *Drosophila melanogaster* revisited by microsatellite analysis," *Molecular Ecology*, vol. 7, no. 7, pp. 915–917, 1998.
- [4] M. F. Wolfner, "The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in *Drosophila*," *Heredity*, vol. 88, no. 2, pp. 85–93, 2002.
- [5] F. W. Avila, L. K. Sirot, B. A. Laflamme, C. D. Rubinstein, and M. F. Wolfner, "Insect seminal fluid proteins: identification and function," *Annual Review of Entomology*, vol. 56, pp. 21–40, 2011.
- [6] W. Haerty, S. Jagadeeshan, R. J. Kulathinal et al., "Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*," *Genetics*, vol. 177, no. 3, pp. 1321–1335, 2007.
- [7] A. C. Fiumera, B. L. Dumont, and A. G. Clark, "Sperm competitive ability in *Drosophila melanogaster* associated with variation in male reproductive proteins," *Genetics*, vol. 169, no. 1, pp. 243–257, 2005.
- [8] A. C. Fiumera, B. L. Dumont, and A. G. Clark, "Associations between sperm competition and natural variation in male reproductive genes on the third chromosome of *Drosophila melanogaster*," *Genetics*, vol. 176, no. 2, pp. 1245–1260, 2007.
- [9] B. S. Baker, M. Gorman, and I. Marin, "Dosage compensation in *Drosophila*," *Annual Review of Genetics*, vol. 28, pp. 491–521, 1994.
- [10] M. Parisi, R. Nuttall, D. Naiman et al., "Paucity of genes on the *Drosophila* X chromosome showing male-biased expression," *Science*, vol. 299, no. 5607, pp. 697–700, 2003.
- [11] J. M. Drnevich, M. M. Reedy, E. A. Ruedi, S. Rodriguez-Zas, and K. A. Hughes, "Quantitative evolutionary genomics: differential gene expression and male reproductive success in *Drosophila melanogaster*," *Proceedings of the Royal Society B*, vol. 271, no. 1554, pp. 2267–2273, 2004.
- [12] S. Dorus, S. A. Busby, U. Gerike, J. Shabanowitz, D. F. Hunt, and T. L. Karr, "Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome," *Nature Genetics*, vol. 38, no. 12, pp. 1440–1445, 2006.
- [13] G. D. Findlay, M. J. MacCoss, and W. J. Swanson, "Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*," *Genome Research*, vol. 19, no. 5, pp. 886–895, 2009.
- [14] E. R. Wasbrough, S. Dorus, S. Hester et al., "The *Drosophila melanogaster* sperm proteome-II (DmSP-II)," *Journal of Proteomics*, vol. 73, no. 11, pp. 2171–2185, 2010.
- [15] V. R. Chintapalli, J. Wang, and J. A. T. Dow, "Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease," *Nature Genetics*, vol. 39, no. 6, pp. 715–720, 2007.
- [16] E. M. Hill-Burns and A. G. Clark, "X-linked variation in immune response in *Drosophila melanogaster*," *Genetics*, vol. 183, no. 4, pp. 1477–1491, 2009.
- [17] S. Rozen and H. Skaletsky, "Primer3 on the WWW for general users and for biologist programmers," in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, S. Krawetz and S. Misener, Eds., pp. 365–386, Humana Press, Totowa, NJ, USA, 2000.
- [18] M. A. Crosby, J. L. Goodman, V. B. Strelets, P. Zhang, and W. M. Gelbart, "FlyBase: genomes by the dozen," *Nucleic Acids Research*, vol. 35, supplement 1, pp. D486–D491, 2007.
- [19] J. D. Storey, "A direct approach to false discovery rates," *Journal of the Royal Statistical Society. Series B*, vol. 64, no. 3, pp. 479–498, 2002.
- [20] K. R. Ram and M. F. Wolfner, "Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction," *Integrative and Comparative Biology*, vol. 47, no. 3, pp. 427–445, 2007.
- [21] C. Y. Chow, M. F. Wolfner, and A. G. Clark, "The genetic basis for male x female interactions underlying variation in reproductive phenotypes of *Drosophila*," *Genetics*, vol. 186, no. 4, pp. 1355–1365, 2010.