Reduced X-linked nucleotide polymorphism in *Drosophila simulans*

David J. Begun* and Penn Whitley

Section of Integrative Biology and Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712

Population genetic theory predicts that selectively driven changes of allele frequency for both beneficial and deleterious mutants reduce polymorphism at tightly linked sites. All else being equal, these reductions in polymorphism are expected to be greater when recombination rates are lower. Therefore, the empirical observation of a positive correlation between recombination rates and amounts of DNA polymorphism across the *Drosophila melanogaster* genome can be explained by two very different types of natural selection. Here, we evaluate alternative models of effects of selection on linked sites by comparison of X-linked and autosomal variation. We present polymorphism data from 40 genes distributed across chromosome arms X and 3R of *Drosophila simulans*, a sibling species of *D. melanogaster*. We find significantly less silent polymorphism in *D. simulans* on the X chromosome than on 3R, but no difference between arms for silent divergence between species. This pattern is incompatible with predictions from theoretical studies on the effect of negative selection on linked sites. We propose that some form of positive selection having greater effects on sex chromosomes than on autosomes is the better explanation for the *D. simulans* data.

The amount of nucleotide polymorphism in a given region of the *Drosophila melanogaster* genome is positively correlated with the regional recombination rate (1, 2), which varies several-fold from one part of the genome to another (e.g., ref. 3). The absence of a detectable effect of varying recombination rates on amounts of nucleotide divergence between species (1, 2) rules out variation in neutral mutation rates (4) across the genome as the explanation for the polymorphism data from *D. melanogaster*. Two kinds of population genetic models, both of which invoke effects of natural selection on linked neutral sites, have dominated the debate as to the cause of these patterns. One model (5–7) invokes effects of selection for beneficial mutants on linked neutral sites (i.e., hitch-hiking effects), whereas a second model (8–10) invokes effects of selection against deleterious mutants on linked neutral sites (i.e., background selection). Both models predict that selection reduces polymorphism at linked neutral sites and that the magnitude of this reduction is greater when recombination rates are lower. Finally, for both models there is no expectation that variation in recombination rates affects divergence between species at neutral sites (11). The qualitatively similar predictions of the two models have made it difficult to determine their relative importance. The background selection model can provide a reasonably good fit to the polymorphism data, given certain parameters of deleterious mutation rate and recombination rate in *D. melanogaster* (9, 10). Similarly, a simple hitchhiking model also fits the *D. melanogaster* polymorphism data quite well (7, 12). Attempts to distinguish hitchhiking effects from background selection based on the frequency spectrum of variation in regions of very low crossing-over in *D. melanogaster*, or on the distribution of variation within and between populations of *Drosophila ananassae* either have been ambiguous or have not spoken to the larger issue of the genomewide effect of selection on linked, neutral sites (13–15).

Drosophila males carry only one copy of the X chromosome (i.e., males are heterogametic), whereas females carry two copies of the X chromosome. Heterogamety can have important implications for the way in which natural selection operates (16).

For example, theoretical (e.g., ref. 17) and empirical (18, 19) research shows that selection removes recessive deleterious mutants from populations more effectively on X chromosomes than on autosomes. Differences in the population genetics of X chromosomes and autosomes suggest that another approach for understanding the effect of selection on linked sites across the genome is to compare X-linked and autosomal variation (2, 10).

The background selection model makes a clear prediction about levels of neutral sequence variation on X chromosomes and autosomes (8, 10). Consider a neutral site and a linked site at mutation-selection balance experiencing recurrent mutation to deleterious alleles. Under background selection, the long-term effective population size at the neutral site can be thought of as depending on the proportion of genes in the population carrying a deleterious mutation at the linked site. This is because neutral alleles that are linked to a deleterious mutation are quickly removed from the population, and thus do not "contribute" to the long-term population size at the neutral site. The amount of neutral variation depends on the effective population size, thus explaining the effect of background selection on reducing levels of linked, neutral variation. An important genetic property of the deleterious mutants that cause background selection is that they are partially recessive (8, 20). Because recessive deleterious mutants are maintained at lower frequency and removed from populations more quickly on X chromosomes than on autosomes (e.g., ref. 17), neutral alleles on X chromosomes are less likely to be linked to a deleterious mutant compared with neutral alleles on autosomes. Thus, all else being equal, background selection should leave X chromosomes more polymorphic than autosomes at linked, neutral sites (8, 10) after we correct for expected differences in population size between X chromosomes and autosomes; levels of X-linked polymorphism are corrected by multiplying by 4/3 because with equal numbers of males and females there are three X chromosomes for every four autosomes.

Analysis of published *Drosophila simulans* data provided some evidence of reduced X-linked polymorphism for freely recombining regions (21); however, there was no attempt to account for variation in neutral mutation rates among genes. Perhaps more importantly, this supposed effect of sex linkage on polymorphism was almost entirely attributable to an unusually low level of variation at the X-linked gene, *Yp2*, and/or an unusually high level of variation at the autosomal gene, *Est-6*. Some evidence also suggested that X-linked genes were less variable than autosomal genes in *D. melanogaster* (2). However, there was no quantitative support for a reduction in X-linked polymorphism, and no attempt to account for variation in neutral mutation rates across loci. Here we present sequence data from a large number of *D. simulans* genes distributed across chromosome arms X and 3R in an attempt to test predictions

---

of background selection vis-à-vis levels of autosomal vs. X-linked variation.

The *D. simulans* genome has three properties that make it well suited for comparisons of X-linked and autosomal polymorphism. First, the high nucleotide variability of its genome (21) facilitates comparisons of levels of variation from one part of the genome to another. Second, although there are fewer genetic data for *D. simulans* than for *D. melanogaster*, the available *D. simulans* data indicate that there is relatively little heterogeneity in recombination rates from one part of the genome to another (22–24). For example, there are 0.06 map units per polytene chromosome band for both the X chromosome and 3R. The centromere-associated reduction of crossing-over is restricted to a much smaller physical region in *D. simulans* than in *D. melanogaster*, and there is no obvious variation in recombination rates over much of X and 3R (22–24). More generally, although data are scarce, comparison of genetic and cytological maps of other *Drosophila* species (25–30) provides no evidence that X-chromosome recombination rates are generally lower than autosomal recombination rates in *Drosophila*. In fact, given that X chromosomes spend a smaller proportion of their time (0.33) in the zero-recombination environment of males than do autosomes (0.5) it is likely that the X chromosome experiences higher recombination rates than 3R in natural *D. simulans* populations. Thus, the expectation under background selection of reduced autosomal polymorphism is probably conservative (although the effect of slightly different recombination rates for X chromosomes vs. autosomes on expected levels of polymorphism for the two types of chromosomes has not been studied theoretically). *Drosophila* inversions can modify long-term recombination rates across the genome (31) in a complex manner that depends on their histories and genetic properties (32). Unlike populations of *D. melanogaster*, *D. simulans* populations harbor virtually no chromosome inversion polymorphism (33). This suggests that genetic data from crosses provide better estimates of long-term average recombination rates in this species than in *D. melanogaster*. Furthermore, the distribution of nucleotide variation within populations is less likely to have been affected by selection on inversion polymorphism in *D. simulans* than in *D. melanogaster*.

Our approach for comparing X-linked and autosomal polymorphism is robust to small errors in recombination rates per physical distance estimated in mapping crosses. For example, the fact that we have sampled many loci scattered across the euchromatic regions of both chromosome arms means that undocumented variability of recombination rates over small physical regions within chromosome arms is not expected to affect our inferences regarding average differences in the population genetics of different chromosome arms. The correction we use for expected differences in population size (i.e., polymorphism) between X chromosomes and autosomes assumes equal effective sizes for males and females. Several types of data suggest that there is sexual selection on males in *D. melanogaster* populations (34–37), with the consequence that the effective population size for males is likely to be smaller than the effective population size for females. The fact that females are refractory to remating to a much greater extent than males (e.g., ref. 38) also may promote a pattern of limited mating opportunities for males in natural populations. If these patterns also hold for the very closely related species, *D. simulans*, then our correction factor results in an overestimate of the expected X chromosome polymorphism relative to the autosomal polymorphism because the difference in effective population size of X chromosomes and autosomes is diminished as male effective population size becomes smaller relative to that of females [with extreme sexual selection on males, X chromosome effective population sizes are actually larger than autosomal population sizes (16)]. Thus, our assumption of equal numbers of males and females is probably conservative. In principle, an extraordinary skew toward a male-biased sex ratio (i.e., much larger effective population sizes for males than for females)

could bias our analysis (16, 39); however, the evidence argues against such a phenomenon (34–37).

Materials and Methods

D. simulans data not from previously published studies are from a set of highly inbred lines made from field-caught inseminated females collected in the Wolfskill Orchard, Winters, CA in summer 1995. PCR products were directly sequenced; sequencing reactions were run on an Applied Biosystems 377 automated sequencer. Data were analyzed primarily with the DNASP program (40). None of the genes in the survey are located in regions of severely reduced crossing-over or were sampled because of *a priori* hypotheses that they were under selection in *D. simulans*. The cytological locations given for *D. simulans* are identical to those of homologous gene in *D. melanogaster*, with the exception of *Rh3, mir, nos, Hsc70, Rel, hyd, eld, CP190*, and *ry*. These genes lie within a fixed inversion difference between *D. melanogaster* and *D. simulans* (85A; 93F6). Therefore, physical locations of these genes in *D. simulans* were estimated by aligning the physical map of *D. melanogaster* with the approximate locations of the genes in *D. simulans*. For example, *Rh3*, located at 92D1 in *D. melanogaster*, would be approximately at a position physically equivalent to 85E in the *D. simulans* genome. Heterozygosities of X-linked genes shown in Table 1 already have been multiplied by 4/3 to allow direct comparison of data from X-linked and 3R genes.

Results

Tables 1–3 show summaries of silent polymorphism and divergence for 21 X-linked genes and 19 genes on 3R in *D. simulans* (mean number of alleles sampled for genes on X and 3R are very similar, 7.1 and 7.4, respectively). Data from approximately 5,782 silent sites on 3R and 5,442 silent sites on the X are presented. These sites are exclusively from protein-coding regions (i.e., "silent" in this paper does not refer to intron or flanking sites). Ratios of fixed to polymorphic silent variants (Table 2) for genes on the X (531 to 228) vs. 3R (412 to 458) are highly significantly heterogeneous ($G = 86.1$; $P = < 10^{-5}$). Small differences in numbers of alleles sampled per locus for the two chromosomes cannot explain this result; comparison of polymorphisms and fixed differences in samples of five *D. simulans* alleles from each locus gives the same result (X-linked fixed and polymorphic silent sites −509 and 205; 3R fixed and polymorphic silent sites −417 and 398; $G = 65.4$, $P < 10^{-5}$; note, however, that the significance of these tests is inflated by the lack of complete independence of polymorphisms within loci). Correcting the observed number of X-linked polymorphisms in Table 2 by multiplying by 4/3 does not change this conclusion ($G = 31.9$, $P < 10^{-5}$).

One interpretation of the data is that there is a deficit of silent polymorphism on the *D. simulans* X chromosome. That this is the case can be seen from comparisons of silent heterozygosity (θ; ref. 41) on the two chromosome arms. The corrected, unweighted mean silent heterozygosity (SE) of X-linked loci, 0.021 (0.003), is considerably lower than that of 3R loci, 0.035 (0.004). This difference in heterozygosity of genes on the two arms is statistically significant (Mann–Whitney *U* test; $P = 0.03$; this P value is conservative as it assumes zero recombination within loci).

A possible explanation for different amounts of polymorphism for X-linked vs. 3R genes is that average neutral mutation rates of silent sites is lower for X-linked genes than for 3R genes. Analysis of silent site divergence allows us to address this possibility because if silent fixations are neutral, then the divergence is proportional to the mutation rate (10). Average silent site divergence (SE) is actually slightly higher for the X-linked genes in our sample [0.112 (0.007)] than for the 3R genes in our sample [0.108 (0.008)], although the difference between arms is not statistically significant (Mann–Whitney *U* test; $P = 0.52$). This finding suggests that differences between X and 3R loci in the neutral mutation

Table 1. Silent polymorphism (θ) and divergence (D) in _D. simulans_

Gene (n)	Cyt.	Sil sites	θ	D	θ/D
Chromosome 3R					
Gld (11)*	84D	242	0.037	0.125	0.30
Rh3 (5)*	85F	278	0.057	0.097	0.59
mir (6)	86A	276	0.052	0.120	0.43
nos (7)	86D	163	0.023	0.081	0.28
eld (7)	87B	252	0.032	0.053	0.60
Hsc70 (7)	88E	310	0.011	0.054	0.20
CP190 (7)	88E	251	0.010	0.104	0.10
ry (8)	90A	320	0.067	0.171	0.39
hyd (8)	92E	339	0.010	0.051	0.20
Rel (7)	93D	547	0.026	0.098	0.27
pit (7)	93F	270	0.065	0.108	0.60
AP50 (8)	94B	289	0.058	0.106	0.55
T-cpl (8)	94B	273	0.010	0.125	0.08
fzo (8)	94E	304	0.020	0.225	0.09
AATS† (7)	95CD	321	0.038	0.128	0.30
tld (7)	96A	219	0.039	0.149	0.26
Osbp (8)	96B	242	0.033	0.083	0.40
boss (5)*	96F	366	0.051	0.107	0.48
Tpi (9)*	99E	180	0.024	0.075	0.32
Chromosome I (X)					
runt (8)*	19E	384	0.008	0.081	0.10
G6pd (8)*	18E	363	0.012	0.111	0.11
bnb (8)	17E	224	0.011	0.043	0.26
r (6)	15A	285	0.016	0.173	0.09
mei-218 (8)	15D	257	0.012	0.118	0.10
sog (8)	13D	313	0.013	0.091	0.14
g (7)	12B	242	0.025	0.114	0.22
Yp3 (8)	12BC	257	0.002	0.118	0.02
v (8)*	10A	269	0.032	0.166	0.19
Yp2 (6)*	9A	246	0.004	0.067	0.06
otu (6)	7F	260	0.047	0.145	0.32
sn (8)	7D	301	0.025	0.109	0.23
dec-1 (7)	7C	345	0.020	0.155	0.13
ct (6)	7B	222	0.003	0.038	0.08
sqh (7)	5D	100	0.027	0.096	0.28
X (7)	5D	338	0.035	0.122	0.29
ovo (8)	4E	332	0.012	0.104	0.12
mei-9 (6)	4B	266	0.009	0.097	0.09
per (6)*	3B	397	0.059	0.129	0.46
z (6)*	3A	162	0.039	0.134	0.29
Pgd (7)	2D	219	0.037	0.136	0.27

_n_ is the number of alleles sampled; Cyt. is the cytological location on the polytene map; genes for each arm are arranged from the centromere end (top) to telomere end (bottom) of the chromosome arm. _D_ is the average, pairwise silent site divergence between the _D. simulans_ population sample and a single _D. melanogaster_ allele using the Jukes-Cantor correction. Estimates of θ (41) for X-linked genes have been multiplied by 4/3 to make them directly comparable to estimates from autosomal loci.
*Previously published data (42–50). For _G6pd, runt,_ and _v,_ a haphazardly selected subsample of eight published _D. simulans_ alleles was used.
†AATS-GluPro.

Table 2. Silent polymorphisms within _D. simulans_ and fixed differences to _D. melanogaster_

Gene	Fixed	Polymorphic
Chromosome 3R		
Gld	19 (21)	26 (18)
Rh3	16 (16)	33 (33)
mir	21 (21)	30 (30)
nos	10 (10)	9 (8)
eld	5 (5)	20 (20)
Hsc70	13 (13)	8 (8)
CP190	22 (22)	6 (4)
ry	34 (36)	50 (37)
hyd	14 (14)	10 (8)
Rel	42 (42)	34 (24)
pit	14 (14)	43 (43)
AP50	16 (16)	45 (37)
T-cpl	32 (30)	7 (7)
fzo	55 (56)	15 (12)
AATS	29 (29)	29 (27)
tld	24 (24)	21 (18)
Osbp	14 (15)	21 (18)
boss	23 (23)	39 (39)
Tpi	9 (10)	12 (7)
Chromosome I (X)		
runt	25 (25)	6 (6)
G6pd	35 (11)	11 (11)
bnb	8 (9)	5 (2)
r	42 (42)	5 (5)
mei-218	27 (27)	6 (5)
sog	25 (25)	8 (7)
g	22 (22)	11 (11)
Yp3	28 (28)	1 (1)
v	34 (34)	17 (15)
Yp2	15 (15)	2 (2)
otu	29 (29)	21 (19)
sn	26 (26)	15 (14)
dec-1	45 (45)	13 (11)
ct	9 (8)	1 (0)
sqh	8 (8)	5 (3)
X	29 (30)	23 (18)
ovo	30 (30)	8 (8)
mei-9	24 (24)	4 (4)
per	34 (34)	40 (40)
z	15 (16)	11 (8)
Pgd	21 (21)	15 (15)

Numbers in parentheses are for a sample of five _D. simulans_ and one _D. melanogaster_ allele. Numbers not in parentheses are for the _simulans_ samples in Table 1 and one _D. melanogaster_ allele.

rate at silent sites cannot explain the difference in their silent heterozygosity.

A conservative way to test the hypothesis that X-linked genes are less polymorphic than 3R genes after accounting for interlocus variation in the neutral mutation rate is to compare the mean ratio of polymorphism to divergence (θ/D) for X vs. 3R loci in Table 1. We find that the average ratio of silent polymorphism to divergence for X-linked genes (0.18) is only about half as large as the ratio for 3R genes (0.34). This difference between arms in the ratio of polymorphism to divergence is highly significant (Mann–Whitney U

test; $P = 0.005$), strongly suggesting that different amounts of polymorphism on the two chromosome arms cannot be explained by different neutral mutation rates among genes or among chromosome arms.

Silent mutations may fall into two fitness classes, termed preferred or major (more fit) and unpreferred or minor (less fit) (51, 52). Unpreferred polymorphisms are hypothesized to be slightly deleterious alleles maintained at mutation-selection-drift equilibrium in _D. simulans_ (52). We can ask whether differences in levels of silent polymorphism on X and 3R can be explained by differences among arms in proportions of unpreferred to preferred polymorphisms. There are more unpreferred than preferred polymorphisms (Table 3), as previously described for a smaller sample of _D. simulans_ genes (52). However, there is no hint of a difference between arms in the ratio of unpreferred to preferred polymorphisms (Table 3). Thus, there is no evidence that lower levels of

Table 3. Numbers of unpreferred and preferred polymorphisms in *D. simulans*

Outgroup	Chromosome	Unpreferred	Preferred
D. melanogaster	X (*n* = 21)	112	42
	3R (*n* = 19)	221	85
	G-test, *P* = 0.91		
D. melanogaster/*D. yakuba*	X (*n* = 8)	66	13
	3R (*n* = 12)	131	26
	G-test, *P* = 0.98		

Unpreferred and preferred polymorphisms (51, 52) within *D. simulans* were assigned by using *D. melanogaster* as the outgroup for the entire data set of 40 genes, or by using *D. melanogaster* and *D. yakuba* for a subset of the data set constituting 20 genes (*Rh3, ry, Rel, hyd, Tcp-1, AP50, Osbp, boss, Tpi, mir, CP190, Hsc70, G6pd, sog, v, sn, dec-1, X, per, z*). Only codons for which the outgroup codon(s) was identical to one of the segregating *D. simulans* codons were used. Because sequence evolution in these species approximates the infinite sites model, few errors are made in determining ancestral state by this method; errors in assignment of ancestral state should be random with respect to chromosome arm. *n* is the number of loci surveyed on each arm.

silent heterozygosity on the X chromosome can be explained by increased efficacy of purifying selection against unpreferred X-linked polymorphisms.

Discussion

Given that selection against deleterious mutants is expected to leave autosomes less polymorphic than X chromosomes at linked neutral sites, the finding that 3R is significantly more variable than the X chromosome suggests that background selection is not a major cause of perturbations at silent sites in *D. simulans* populations. The spread of an initially rare beneficial allele to fixation causes a reduction of heterozygosity at linked, neutral sites (5, 6). More complex forms of positive selection are also usually expected to reduce variation at linked sites (53). We should always expect the magnitude of these hitchhiking effects to reflect the balance between the strength of selection and the amount of recombination between selected sites and linked, neutral sites (5, 6, 53). What are the circumstances under which positive selection might cause a reduction of linked variation on X chromosomes?

Let us assume that the recombination rate per nucleotide per generation is the same for X chromosomes and autosomes in *D. simulans* populations. Then for hitchhiking effects to be greater on X chromosomes than on autosomes requires either (*i*) that average selection coefficients are greater for X-linked than for autosomal mutants, and/or (*ii*) that on average, there are fewer recombination events between positively selected sites and neutral sites on X chromosomes than on autosomes. There is no obvious biological rationale for the former, so we will focus on the latter. We expect fewer recombination events between positively selected sites and neutral sites on X chromosomes if greater numbers of beneficial mutants per physical distance enter populations on X chromosomes than on autosomes (because neutral sites are more likely to come under the influence of a linked, selected site as the number of sites under positive selection increases), or if rates of allele-frequency change of beneficial mutants are faster for X chromosomes than for autosomes. In general, we might categorize factors that contribute to differences in the relative importance of positive selection on X chromosomes vs. autosomes as (*i*) dominance properties of beneficial mutants, (*ii*) fixation probabilities, fixation rates, and sojourn times of X-linked vs. autosomal mutants, and (*iii*) male heterogamety.

Dominance. If adaptive evolution resulted from selection on new mutants that were, on average, partially recessive, then one would expect greater numbers of X-linked than autosomal adaptive fixations (39, 54). Such a phenomenon would result in greater

hitchhiking effects, and thus lower levels of neutral variation on X chromosomes. When the substitution process is mutation limited, the expected number of substitutions per generation is the product of the number of new mutants produced each generation and the fixation probability for each mutant. Although there are fewer, new X-linked mutants each generation (because there are three X chromosomes for each four autosomes in a population with equal numbers of males and females), the higher fixation probabilities of recessive X-linked, compared with recessive autosomal mutants "outweighs" the smaller number of mutants; the result is that there are greater numbers of X-linked beneficial fixations per generation (54). The main problem with this model is that it is difficult to understand why the average beneficial mutant should be recessive. This is especially true if one believes that most deleterious mutants are recessive (20), for then one would be forced to posit that on average, both deleterious and beneficial mutants are recessive. This hardly seems likely.

New mutants that are deleterious when heterozygous over the ancestral allele (i.e., dominant negative mutants), but beneficial when homozygous (e.g., ref. 55) are expected to fix much more often when X-linked than when autosomal, thereby causing greater hitchhiking effects on X chromosomes. In fact, the difference in substitution rate for X chromosomes vs. autosomes is expected to be very large for such underdominant mutants (54). Results from recent analyses of duplicated genes or genomes have been taken as evidence that dominant negative mutants are common (56, 57). Even if this interpretation were correct, it is not clear how often such mutants might have beneficial effects when homozygous. Nevertheless, the strong theoretical results for underdominant mutants is intriguing. Empirical evidence on whether such mutants contribute to evolution is needed before we can begin to speculate on whether or not they could quantitatively account for reduced polymorphism of X chromosomes in *D. simulans*.

For both recessive and underdominant mutants, increased hitchhiking effects are expected for X chromosomes because there are more nucleotides per physical distance fixing under positive selection on X chromosomes than on autosomes.

Fixation Probabilities and Fixation Rates. *Selection on new mutants.* The frequency of a new X-linked mutant is higher than the frequency of a new autosomal mutant. Although the frequencies for all new mutants is very small in large populations, the relative frequencies of new mutants on the two types of chromosomes is quite different. In a population with equal numbers of males and females the initial frequency of autosomal mutants is $1/2N$, whereas the initial frequency of X-linked mutants is $2/3N$, a third higher than the initial autosomal frequency. Mutants with higher initial frequencies have higher fixation probabilities (4). Thus, for large N and small s, the fixation probability of an individual X-linked mutant is one-third greater than that of an equivalent autosomal mutant (39).

However, the fixation rate (i.e., number of mutants fixing per generation) is the more relevant quantity for us, because it is the beneficial alleles that actually spread through populations that cause hitchhiking effects. The fixation rate depends on the number of new mutants produced per generation and the fixation probability for each mutant. There are fewer X chromosomes than autosomes in a population with equal numbers of males and females, and thus, fewer new X-linked mutants per generation. The higher fixation probability for beneficial X-linked mutants (accruing from their higher initial frequency) is exactly counterbalanced by the fact that there are fewer such mutants introduced into a population each generation; the result is that there is no difference in the number of beneficial mutants fixed per generation for the two kinds of chromosomes for additive mutants (39, 54). This would seem to predict identical fixation rates for beneficial mutants with additive effects on X chromosomes and autosomes, and thus no

difference in effects of hitchhiking on reducing linked variation on the two types of chromosomes.

Selection on standing variation. The above model is explicitly about selection on new mutants (i.e., it posits that adaptive evolution is mutation limited). A model of adaptive evolution in which the supply of beneficial mutations does not depend on new mutants might, however, predict greater fixation rates for X chromosomes. Suppose that beneficial mutants are drawn from a pool of preexisting neutral mutants that become favored subsequent to a change of the environment, and suppose that each selected locus has only one beneficial allele spreading through the population at any time. Then subsequent to environmental change, there will be a ready source of beneficial mutants that only weakly depends on the population size (for large populations). In this case the number of beneficial mutants fixing over time should not strongly depend on the number of new mutants per generation, but rather should depend more on the rate of change of the environment (which we presume to be the same for X-linked and autosomal mutants) and the fixation probabilities of beneficial mutants on X chromosomes vs. autosomes. In other words, if the influx of new beneficial mutants into a population through mutation is no longer a limitation, then the fixation probability determines the fixation rate.

Recall that the fixation probability is higher for any particular new X-linked mutant than it is for any particular new autosomal mutant because the initial allele frequency is higher for the X-linked mutant. In large populations most mutants are young, occur at low frequency, and are lost in very few generations (4); among these mutants, the average frequency is expected to be higher on X chromosomes than on autosomes because the initial frequencies are higher for the X chromosome. Therefore, fixation rates for such mutants may be higher when they are X-linked than when they are autosomal. This, in turn, would predict greater hitchhiking effects for X chromosomes than for autosomes. Whether or not such a model could quantitatively explain our observations remains a matter for speculation. A potential problem with this idea comes from our interpretation of the *D. simulans* data presented here as supporting the notion that there are fewer neutral mutants on X chromosomes than on autosomes. This raises some question as to the overall effect of linked, directional selection on previously neutral, standing variation for X chromosomes vs. autosomes. The problem is complex, and may well depend on the details of the model of natural selection. Further theoretical studies of these issues are needed.

An alternative model of selection on standing variation invokes positive selection on alleles that were deleterious and at mutation-selection balance before a change of the environment. If most deleterious mutants are partially recessive, then the equilibrium frequency is lower for X-linked mutants. This would result in smaller fixation probabilities for X-linked compared with autosomal mutants. However, mutants that are more recessive have proportionally greater fixation probabilities for X-linked compared with autosomal genes; differences in initial frequency and in dominance work in opposite directions as far as fixations and their associated hitchhiking effects for X-linked vs. autosomal mutants initially at mutation-selection balance. Perhaps the most unappealing feature of the model, regardless of the quantitative details of the dynamics, is that as is the case for models that posit recessivity of new beneficial mutants, this model too posits that both deleterious and beneficial mutants are partially recessive.

Sojourn Times. The above arguments comparing fixation probabilities and fixation rates for X chromosomes vs. autosomes suffers from the fact that the X chromosome population is considered as equivalent to an autosomal population that is only three-fourths as large. Although true for neutral mutants, this is an oversimplification for selected mutants because autosomal mutants are always found in diploids, whereas X-linked mutants are found as a mixture of haploids and diploids. The fact that X-linked mutants are

"partial" haploids leads to the theoretical prediction that the rate of allele-frequency change of beneficial mutants is greater for X chromosomes than for autosomes (39, 58). This is true for all frequencies and all dominance coefficients (39). The intuitive explanation for this result is that the variance in fitness is greater for haploids than for diploids, and that the rate of change of allele frequencies is directly proportional to the variance in fitness (59). Thus, the rate of spread of a beneficial allele is greater for an X-linked gene because it spends a third of its time as a haploid in males; X-linked alleles with additive effects on fitness are expected to spread at a rate one-third greater than the rate for equivalent autosomal alleles in large populations (39, 58). If the recombination rate per generation between a selected site and a neutral site were the same for X chromosomes and autosomes, then we would expect fewer recombination events between loci on the X chromosome during the fixation process. This leads to the expectation of greater hitchhiking effects, and thus reduced levels of neutral variation on X chromosomes.

Selection in Males. Greater hitchhiking effects on X chromosomes also could be attributable to the existence of beneficial mutations that invade populations or fix more readily when X-linked than when autosomal. For example, meiotic drive alleles acting in males might invade populations more readily when X-linked than when autosomal (60, 61). Evidence for X-chromosome meiotic drive in several *Drosophila* species (53) including *D. simulans* (63, 64), suggests that such drive alleles may be very common, at least in Diptera (65). Another example of selected alleles that might enter populations more readily when X-linked are those that are sexually antagonistic such that they benefit one sex, but are detrimental to the other (66). Recent experimental results from *D. melanogaster* (37, 67) provide evidence that such alleles are common in *Drosophila* populations. We should point out that fixation might be an unlikely outcome for both meiotic drive and sexually antagonistic alleles (62, 66). One might question whether alleles that start out as very rare mutants and rapidly spread, but do not fix, are candidates for causing hitchhiking effects. Although this problem has not been studied in detail, some theoretical results suggest that such mutants are expected to cause reductions in linked variation (53), although as one would expect, mutants that rapidly fix cause greater hitchhiking effects than mutants that do not fix.

The "faster males" hypothesis for Haldane's rule in *Drosophila* posits that genes of male function may be under stronger or more frequent directional selection compared with other types of genes (68, 69). If this were true, then the dominance coefficient of the class of male-specific beneficial X-linked mutants is irrelevant, and we might expect X-linked adaptive fixations to outnumber autosomal adaptive fixations.

All of our discussions of the relative effects of background selection and hitchhiking effects have assumed that the density of nucleotides that are potential targets of selection is the same for X chromosomes and autosomes. If gene density were higher for the X chromosome, then the numbers of deleterious mutants per kb might be higher for the X chromosome. In this case the background selection model would predict less variation on the X chromosome than on the autosomes. However, greater gene density on the X also would be expected to result in increased numbers of beneficial mutants per kb, greater hitchhiking effects on the X chromosome, and reduced variation on X chromosomes. Data from the *Drosophila* Genome Project will soon tell whether heterogeneity in gene density across the *D. melanogaster* genome may have to be accounted for in any explanation of the *D. simulans* data presented here. The fact that X chromosomes have an additional regulatory mechanism, dosage compensation (70, 71), also raises the possibility that X chromosomes have a higher density of selected nucleotides than autosomes, although the number of X-linked nucleotide sites participating in regulating dosage compensation is unclear. Finally, there are few theoretical treatments of how natural selec-

tion acts differently on X chromosomes vs. autosomes when numbers of males and females in a population are very different. The potential effects of all of these complications on the population genetics of X chromosomes vs. autosomes remain to be explored.

Conclusions

Surveys of variation in mice, humans, and *D. melanogaster* have not provided evidence of reductions of X chromosome variation of the magnitude that we have observed in *D. simulans* (reviewed in ref. 16). Why might this be the case? One explanation is that reduced X-linked variation is actually widespread, but that compared with our *D. simulans* data, data from other taxa have provided less power to reject the null hypothesis of equal levels of variation on the X chromosome and autosomes. Another possibility is that the phenomenon we have observed is not general, but instead is restricted to certain taxa. For example, perhaps the distribution of beneficial mutant effects is such that elevated X-linked hitchhiking effects with observable effects on average levels of X-linked variation depends on the population size; perhaps meiotic drive is an important cause of reduced variation on X chromosomes, and such drive alleles are abundant in some species but not others.

Although we can be fairly confident of our rejection of background selection as a sufficient explanation for the polymorphism and divergence data from *D. simulans*, we are not yet in a position to strongly favor a particular alternative model. The widespread effect of reduced X-linked polymorphism in *D. simulans* leads us to speculate that mutants causing hitchhiking effects are common, but individually have relatively small effects on linked variation; this does not necessarily imply, however, that the beneficial mutants themselves are of very small effect. Uncertainties regarding the generality of the pattern we have observed and the explanations proffered certainly motivate additional empirical and theoretical studies of X chromosome vs. autosome population genetics, and haploid vs. diploid population genetics.

1. Begun, D. J. & Aquadro, C. F. (1992) *Nature (London)* **356,** 519–520.
2. Aquadro, C. F., Begun D. J. & Kindahl, E. C. (1994) in *Non-Neutral Evolution: Theories and Molecular Data*, ed. Golding, B. (Chapman and Hall, NY), pp. 46–56.
3. Ashburner, M. (1989) *Drosophila: A Laboratory Handbook* (Cold Spring Harbor Lab. Press, Plainview, NY).
4. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge).
5. Maynard Smith, J. & Haigh, J. (1974) *Genet. Res.* **23,** 23–35.
6. Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989) *Genetics* **123,** 887–899.
7. Wiehe, T. H. E. & Stephan, W. (1993) *Mol. Biol. Evol.* **10,** 842–854.
8. Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993) *Genetics* **134,** 1289–1303.
9. Hudson, R. R. & Kaplan, N. L. (1995) *Genetics* **141,** 1605–1617.
10. Charlesworth, B. (1996) *Genet. Res.* **68,** 131–149.
11. Birky, C. W., Jr. & Walsh, J. B. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6414–6418.
12. Stephan, W. (1995) *Mol. Biol. Evol.* **12,** 959–962.
13. Aguadé, M., Miyashita, N. & Langley, C. H. (1989) *Genetics* **122,** 607–615.
14. Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan W. (1995) *Genetics* **140,** 783–796.
15. Stephan, W., Xing, L., Kirby, D. A. & Braverman, J. M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 5649–5654.
16. Hedrick, P. W. & Parker, J. D. (1997) *Annu. Rev. Ecol. Syst.* **28,** 55–83.
17. Crow, J. F. & Kimura, M. (1970) *An Introduction to Population Genetics Theory* (Harper and Row, New York), pp. 258–262.
18. Langley, C. H., Voelker, R. A., Leigh Brown, A. J., Ohnishi, S., Dickson, B. & Montgomery, E. (1981) *Genetics* **99,** 151–156.
19. Eanes, W. F., Hey, J. & Houle, D. (1985) *Genetics* **111,** 831–844.
20. Crow, J. F. & Simmons, M. J. (1983) in *Genetics and Biology of Drosophila*, eds. Ashburner, M., Carson, H. L. & Thompson, J. N., Jr. (Academic, London), pp. 1–35.
21. Moriyama, E. N. & Powell, J. R. (1996) *Mol. Biol. Evol.* **13,** 261–277.
22. Sturtevant, A. H. (1929) *Contributions to the Genetics of Drosophila simulans and Drosophila Melanogaster* (Carnegie Institute, Washington, DC), No. 399.
23. Ohnishi, S. & Voelker, R. A. (1979) *Jpn. J. Genet.* **54,** 203–209.
24. True, J. R., Mercer, J. M. & Laurie, C. C. (1996) *Genetics* **142,** 507–523.
25. Gubenko, I. S. & Evgen'ev, M. B. (1984) *Genetica* **65,** 127–139.
26. Dobzhansky, T. (1950) *J. Hered.* **41,** 156–158.
27. Spassky, B. & Dobzhansky, T. (1950) *Heredity* **4,** 201–215.
28. Sturtevant, A. H. & Tan, C. C. (1937) *J. Genet.* **34,** 415–432.
29. Orr, H. A. (1995) *Drosophila Inf. Serv.* **76,** 127–128.
30. Sorsa, V. (1988) *Chromosome Maps of Drosophila* (CRC, Boca Raton, FL).
31. Schultz, J. & Redfield, H. (1951) *Cold Spring Harbor Symp. Quant. Biol.* **16,** 175–197.
32. Sniegowski, P. D., Pringle, A. & Hughes, K. A. (1994) *Genet. Res.* **63,** 57–62.
33. Ashburner, M. & Lemeunier, F. (1976) *Proc. R. Soc. London* **193,** 137–157.
34. Bateman, A. J. (1948) *Heredity* **2,** 349–368.
35. Wilkinson, G. S. (1987) *Evolution* **41,** 11–21.
36. Gromko, M. H. & Markow, T. A. (1993) *Anim. Behav.* **45,** 253–262.
37. Rice, W. R. (1996) *Nature (London)* **381,** 232–234.
38. Gromko, M. H., Gilbert, D. G. & Richmond, R. C. (1984) in *Sperm Competition and the Evolution of Animal Mating Systems*, ed. Smith, R. L. (Academic, London), pp. 371–426.
39. Avery, P. J. (1984) *Genet. Res.* **44,** 321–341.
40. Rozas, J. & Rozas, R. (1999) *Bioinformatics* **15,** 174–175.
41. Watterson, G. A. (1975) *Theor. Popul. Biol.* **7,** 256–276.
42. Hamblin, M. T. & Aquadro, C. F. (1996) *Mol. Biol. Evol.* **13,** 1133–1140.
43. Ayala, F. J., Chang, B. S. W. & Hartl, D. L. (1993) *Genetica* **92,** 23–32.
44. Ayala, F. J. & Hartl, D. L. (1993) *Mol. Biol. Evol.* **10,** 1030–1040.
45. Hasson, E., Wang, I. N., Zeng, L. W., Kreitman, M. & Eanes, W. F. (1998) *Mol. Biol. Evol.* **15,** 756–769.
46. Eanes, W. F., Kirchner, M., Yoon, J., Biermann, C. H., Wang, I. N., McCartney, M. A. & Verrelli, B. C. (1996) *Genetics* **144,** 1027–1041.
47. Begun, D. J. & Aquadro, C. F. (1995) *Genetics* **140,** 1019–1032.
48. Hey, J. & Kliman, R. M. (1993) *Mol. Biol. Evol.* **10,** 804–822.
49. Kliman, R. M. & Hey, J. (1993) *Genetics* **133,** 375–387.
50. Labate, J. A., Biermann, C. H. & Eanes, W. F. (1999) *Mol. Biol. Evol.* **16,** 724–731.
51. Sharp, P. M. & Lloyd, A. T. (1993) in *An Atlas of Drosophila Genes: Sequences and Molecular Features*, ed. Maroni, G. (Oxford Univ. Press, New York), pp. 378–397.
52. Akashi, H. (1995) *Genetics* **139,** 1067–1076.
53. Gillespie, J. H. (1997) *Gene* **205,** 291–299.
54. Charlesworth, B., Coyne, J. & Barton, N. H. (1987) *Am. Nat.* **130,** 113–146.
55. Roughgarden, J. (1979) *Theory of Population Genetics and Evolutionary Ecology* (Macmillan, New York).
56. Hughes, M. K. & Hughes, A. L. (1993) *Mol. Biol. Evol.* **10,** 1360–1369.
57. Gibson, T. & Spring, J. (1998) *Trends Genet.* **14,** 46–49.
58. Hartl, D. L. (1972) *Am. Nat.* **106,** 516–524.
59. Fisher, R. A. (1958) *The Genetical Theory of Natural Selection* (Dover, New York).
60. Hurst, L. D. & Pomianowski, A. (1991) *Genetics* **128,** 841–858.
61. Wu, C.-I. & Hammer, M. F. (1991) in *Evolution at the Molecular Level*, eds. Selander, R., Clark, A. & Whittam, T. (Sinauer, Sunderland, MA), pp. 177–203.
62. Jaenike, J. (1996) *Am. Nat.* **148,** 237–254.
63. Cazemajor, M., Landre, C. & Montchamp-Moreau, C. (1997) *Genetics* **147,** 635–642.
64. Atlan, A., Mercot, H., Landre, C. & Montchamp-Moreau, C. (1997) *Evolution* **51,** 1886–1895.
65. Jiggins, F. M., Hurst, G. D. D. & Majerus, M. E. N. (1999) *Am. Nat.* **154,** 481–483.
66. Rice, W. R. (1984) *Evolution* **38,** 735–742.
67. Holland, B. & Rice, W. R. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 5083–5088.
68. Wu, C.-I. & Davis, A. W. (1992) *Am. Nat.* **142,** 187–212.
69. Orr, H. A. (1989) *Heredity* **63,** 231–237.
70. Stuckenholz, C., Kageyama, Y. & Kuroda, M. I. (1999) *Trends Genet.* **15,** 454–458.
71. Baker, B. S. (1994) *Annu. Rev. Genet.* **28,** 491–521.

EVOLUTION