Selection at Linked Sites in the Partial Selfer *Caenorhabditis elegans*

Asher D. Cutter and Bret A. Payseur

Department of Ecology and Evolutionary Biology, University of Arizona

Natural selection can produce a correlation between local recombination rates and levels of neutral DNA polymorphism as a consequence of genetic hitchhiking and background selection. Theory suggests that selection at linked sites should affect patterns of neutral variation in partially selfing populations more dramatically than in outcrossing populations. However, empirical investigations of selection at linked sites have focused primarily on outcrossing species. To assess the potential role of selection as a determinant of neutral polymorphism in the context of partial self-fertilization, we conducted a multivariate analysis of single-nucleotide polymorphism (SNP) density throughout the genome of the nematode *Caenorhabditis elegans*. We based the analysis on a published SNP data set and partitioned the genome into windows to calculate SNP densities, recombination rates, and gene densities across all six chromosomes. Our analyses identify a strong, positive correlation between recombination rate and neutral polymorphism (as estimated by noncoding SNP density) across the genome of *C. elegans*. Furthermore, we find that levels of neutral polymorphism are lower in gene-dense regions than in gene-poor regions in some analyses. Analyses incorporating local estimates of divergence between *C. elegans* and *C. briggsae* indicate that a mutational explanation alone is unlikely to explain the observed patterns. Consequently, we interpret these findings as evidence that natural selection shapes genome-wide patterns of neutral polymorphism in *C. elegans*. Our study provides the first demonstration of such an effect in a partially selfing animal. Explicit models of genetic hitchhiking and background selection can each adequately describe the relationship between recombination rate and SNP density, but only when they incorporate selfing rate. Clarification of the relative roles of genetic hitchhiking and background selection in *C. elegans* awaits the development of specific theoretical predictions that account for partial self-fertilization and biased sex ratios.

Introduction

Theory indicates that genome-wide patterns of single-nucleotide polymorphism (SNP) are influenced by the interaction between natural selection and genetic linkage. Purifying selection against deleterious mutations can remove linked neutral polymorphisms, leading to a reduction in variation ("background selection"; Charlesworth, Morgan, and Charlesworth 1993). Positive selection increasing the frequency of and/or fixing beneficial mutations can also cause a decrease in polymorphism levels at linked nucleotides ("genetic hitchhiking"; Maynard Smith and Haigh 1974). The effects of deleterious and beneficial mutations on neutral diversity are expected to be most severe in genomic regions that rarely recombine. Moreover, if the rate of deleterious mutation or selective sweeps (or both) is sufficiently high, background selection (Hudson and Kaplan 1995) and genetic hitchhiking (Wiehe and Stephan 1993) models predict an overall positive correlation between nucleotide variation and recombination rate.

Surveys of nucleotide variation in low-recombination regions in *Drosophila melanogaster* support these predictions. Reduced nucleotide polymorphism has been observed at the tip of the X chromosome (Aguadé, Miyashita, and Langley 1989; Begun and Aquadro 1991) and on the fourth chromosome (Berry, Ajioka, and Kreitman 1991; Jensen, Charlesworth, and Kreitman 2002; but see Wang et al. 2002), both genomic regions experiencing low recombination rates. Additionally, nucleotide polymorphism and recombination rate are positively correlated across the *D. melanogaster* genome (Begun and Aquadro 1992; Aquadro, Begun, and Kindahl 1994; Moriyama and Powell 1996).

Data from a variety of additional species suggest that the positive relationship between nucleotide polymorphism and recombination rate may be taxonomically widespread. Nucleotide diversity is reduced in low-recombination regions in a number of other *Drosophila* species, including *D. ananassae* (Stephan and Langley 1989; Chen, Marsh, and Stephan 2000), *D. simulans* (Begun and Aquadro 1991; Berry, Ajioka, and Kreitman 1991), *D. mauritiana* (Hilton, Kliman, and Hey 1994), and *D. sechellia* (Hilton, Kliman, and Hey 1994). Furthermore, there is evidence for a positive correlation between nucleotide variation and recombination rate in humans (Nachman et al. 1998; Przeworski, Hudson, and Di Rienzo 2000; Nachman 2001), and weaker support for such an association in house mice (Nachman 1997), sea beets (Kraft et al. 1998), tomatoes (Stephan and Langley 1998; Baudry et al. 2001), goatgrasses (Dvorak, Luo, and Yang 1998), and maize (Tenaillon et al. 2001).

The phylogenetic distribution of these patterns raises the question of what factors may be responsible for variation in the role of selection at linked sites in shaping genomic patterns within a species. Among other attributes, the breeding system may affect the dynamics and observed signature of selection at linked sites (Charlesworth and Wright 2001). Selfing reduces the "effective recombination rate" between selected and unselected loci within a genome because the effective recombination rate is controlled by both chromosomal crossovers and outcrossing (Nordborg 1997, 2000). In other words, like restricted recombination in outcrossing species, self-fertilization generates linkage disequilibrium between selected and neutral mutations, increasing the effects of selection on neutral polymorphism. Consequently, the predicted signature of selection at linked sites (e.g., the correlation between nucleotide polymorphism and

Key words: evolutionary genomics, nucleotide polymorphism, SNP, recombination rate, genetic hitchhiking, background selection, *Caenorhabditis elegans*.

E-mail: acutter@email.arizona.edu.

Mol. Biol. Evol. 20(5):665–673. 2003
DOI: 10.1093/molbev/msg072

recombination rate) depends on the level of self-fertilization (Baudry et al. 2001). In a purely selfing population, the effects of recombination are essentially eliminated, so no genomic pattern is expected for variation in neutral polymorphism levels due to selection at linked sites. Highly, but not obligately, selfing populations are expected to exhibit a positive correlation between neutral polymorphism and recombination rate over a greater range of recombination rates than populations with lower degrees of selfing (Baudry et al. 2001). Furthermore, Hedrick (1980) demonstrated that positive selection usually affects neutral variation more dramatically in self-fertilizing species than in regions of reduced recombination in outcrossing species. Provided that initial genotypic frequencies are not at Hardy-Weinberg equilibrium (a reasonable assumption in self-fertilizing species), this result holds for a range of self-fertilization rates. Theoretical work also suggests that background selection may be stronger in self-fertilizing species than in their outcrossing relatives (Charlesworth, Morgan, and Charlesworth 1993; Nordborg, Charlesworth, and Charlesworth 1996). Hence, selection at linked sites is likely to be an important determinant of neutral polymorphism patterns within partially self-fertilizing species.

Empirical studies of the association between nucleotide variation and recombination rate in tomatoes (Baudry et al. 2001), goatgrasses (Dvorak, Luo, and Yang 1998), and maize (Tenaillon et al. 2001) have provided mixed results regarding the potential importance of selection at linked sites in partially self-fertilizing organisms. In goatgrasses, nucleotide variation and recombination rate are weakly positively correlated in five self-fertilizing species, but not in the one outcrossing species investigated (Dvorak, Luo, and Yang 1998). Two self-compatible and three self-incompatible tomato species show a trend toward a correlation between nucleotide diversity and recombination rate, but none of these relationships is significant (although only five genes were surveyed; Baudry et al. 2001). In maize, nucleotide polymorphism and recombination rate appear to be positively correlated (Tenaillon et al. 2001), although this study estimated recombination rates indirectly from observed levels of linkage disequilibria.

Hence, theoretical studies suggest that selection at linked sites should be important in self-fertilizing species, but available evidence provides ambiguous support for this prediction. This incongruence indicates that further investigation of the relationship between nucleotide polymorphism and recombination rate in self-fertilizing species is warranted. To date, this relationship has not been evaluated in any animal that engages in self-fertilization.

The bacteriophagous, soil-dwelling nematode, *Caenorhabditis elegans*, provides a good system in which to assess the effect of reproductive mode on selection at linked sites. First, this androdioecious species reproduces primarily via self-fertilization of hermaphrodites and presumably outcrosses with males only rarely (Fitch and Thomas 1997). Second, availability of dense genetic maps (Barnes et al. 1995) and the complete genomic sequence (The *C. elegans* Sequencing Consortium 1998) allow estimation of recombination rates across the genome. Third, a large-scale study of SNP identification has recently been completed in *C. elegans* (Wicks et al. 2001). Finally, a recent study concluded that selection at linked sites may explain differences in levels of nucleotide polymorphism between Caenorhabditid species (Graustein et al. 2002). Here, we demonstrate that nucleotide polymorphism (as measured by SNP density) and recombination rate correlate strongly and positively across the genome of *C. elegans*. We also suggest that gene-dense regions may harbor lower polymorphism levels than gene-poor regions. Our results indicate that natural selection is an important determinant of genome-wide patterns of neutral DNA sequence variability in *C. elegans*. Finally, we discuss the ability of background selection and genetic hitchhiking models to explain our results, and we suggest that background selection is more compatible with observed patterns than is widespread genetic hitchhiking.

Methods

To estimate SNP density, we utilized the updated version (as of January 2002) of the SNP data set of Wicks et al. (2001), available online at http://genome.wustl.edu/projects/celegans/database/. Polymorphisms were identified based on comparisons of shotgun sequences (5.4 Mbp total) of 11,000 random clones from the Hawaiian *C. elegans* strain CB4856 with the canonical Bristol N2 strain (Wicks et al. 2001). The authors report that these sequences were randomly distributed across the genome (Wicks et al. 2001). From these polymorphism data we excluded 1,574 small insertion or deletion polymorphisms and 782 exonic SNPs, leaving 3,976 noncoding and therefore putatively neutral SNPs. Although not all noncoding nucleotides are strictly neutral (Shabalina and Kondrashov 1999), we restricted our analyses to SNPs in noncoding regions to minimize any potential influence of directly selected sites. Furthermore, preliminary analyses showed no significant correlation across genomic scales of recombination rate with exonic SNP density, which should harbor a greater fraction of directly selected sites than noncoding regions, suggesting that such a phenomenon would be unlikely to influence our analyses. All other genomic information was derived from WormBase release WS62 (January 2002, http://www.wormbase.org), including 21,448 predicted positions and sequences of coding regions, mapped locus positions, and clone GC content from *C. elegans*, and predicted coding locus sequences from ~13 Mbp of the *C. briggsae* genome.

We calculated estimates for genomic statistics (SNP density, gene density, recombination rate, base composition, mean coding region divergence) for nonoverlapping windows of sequence along each chromosome, starting at both the left and right ends of each chromosome. We varied the size of the windows from 500 kbp to 7 Mbp at 500-kbp intervals. We excluded some windows at the ends of chromosomes that were less than half the length of the window size under consideration. The appropriate scale at which to consider relationships between the genomic statistics was unclear, so we arbitrarily chose the forward-oriented 4 Mbp window size for more intensive analysis.

Unless otherwise noted, our results refer to analyses of this 4 Mbp window size ($n = 24$, one of these points was excluded from analyses involving divergence due to insufficient numbers of homologous loci in *C. briggsae*). We calculated SNP density as the number of noncoding SNPs per Mbp of noncoding DNA and gene density as the number of coding genes (including alternative splicings) per Mbp of total DNA in the window. Because the SNPs were identified from random sequences of ~5% of the worm genome (Wicks et al. 2001), the *absolute* magnitude of SNP density should be ~20-fold higher than the values reported here, although relative density should remain unaffected. We estimated recombination rates (cM/Mbp) for each window based on the total genetic and physical lengths of the windows, using the flanking two mapped loci at each boundary to infer the position of the boundary in cM (from a set of 515 mapped loci among the six *C. elegans* chromosomes in WormBase). We estimated divergence rates across the genome for 1,326 locus pairs between *C. elegans* and *C. briggsae* using a calculation of k_s on nucleotide sequences with Diverge from the Wisconsin Package Version 10.2 (Genetics Computer Group [GCG], Madison, Wis.) software based on the method of Li (Pamilo and Bianchi 1993; Li 1993). We selected the putative homologous loci based on top Blast scores of predicted coding sequences for the full *C. elegans* genome and ~13 Mbp of sequence from *C. briggsae* in WormBase, and we aligned the sequence pairs based on predicted protein sequence. We adjusted the k_s estimates by subtracting the residuals of a linear regression with the codon bias statistic F_{op} (Stenico, Lloyd, and Sharp 1994; Keightley and Eyre-Walker 2000; Marais and Duret 2001). We calculated the mean adjusted k_s value (δ_s) of the available loci in each window as an estimator of local mutation rate. Windows of 4 Mbp size contained a mean of 53.9 loci from which average δ_s was calculated, and windows with fewer than two loci for mean δ_s estimation were excluded from analyses involving divergence. Recombination rate and SNP density were not normally distributed. We employed two approaches to address this issue. Although normality was not completely restored, we used logarithmic transformations of these variables in multiple regression analyses to better satisfy the assumptions of this method. Also, we used nonparametric correlation tests in bivariate analyses that included these variables.

Results

Our inferences about the recombinational landscape of the *C. elegans* genome are consistent with those reported previously (Barnes et al. 1995). For example, recombination rates tend to be higher on chromosomal arms and lower near chromosomal centers. Single-nucleotide polymorphism density is strongly, positively correlated with recombination rate across the *C. elegans* genome (Spearman's ρ= 0.72, $P < 0.0001$; figs. 1 and 2), consistent with the action of selection at linked sites and the qualitative conclusions of Koch et al. (2000). Other bivariate analyses show that SNP density correlates positively with divergence δ_s (Spearman's ρ = 0.84, $P <$ 0.0001) and GC content (Spearman's ρ = 0.51, P = 0.012), but negatively with gene density (Spearman's ρ = –0.42, P = 0.045).

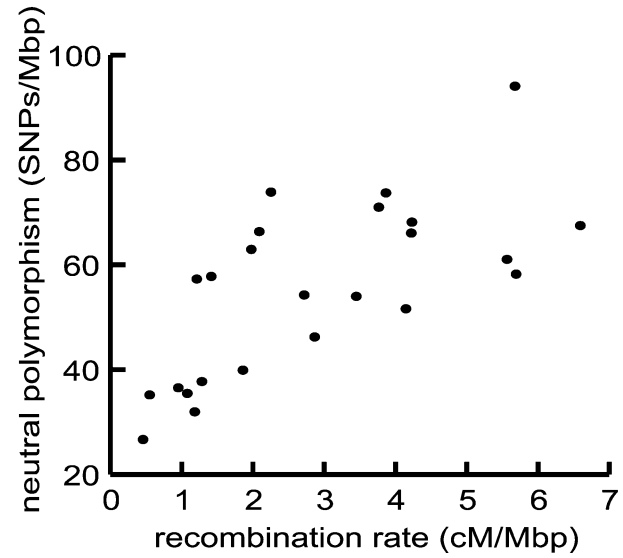

FIG. 1.—The relationship between neutral polymorphism (as estimated by SNP density) and recombination rate is positive.

To ascertain the effect of recombination rate on SNP density independent of other factors, we constructed a multiple linear regression model of SNP density including recombination rate, gene density, mean δ_s, and chromosome identity as covariates. We excluded base composition from the multivariate model because its inclusion did not explain significantly more variation in SNP density at any scale. In this model, recombination rate and chromosome identity contribute significantly to variation in SNP density at nonoverlapping window scales up to 6 Mbp long ($P < 0.037$ in either forward or reverse oriented windows for each scale; fig. 3). Gene density also explains a significant fraction of the variation in SNP density, with a negative correlation independent of recombination rate across many scales, particularly at scales greater than 3 Mbp ($P < 0.042$ in either forward or reverse oriented windows for each scale; fig. 3). Divergence (δ_s) does not consistently explain a significant fraction of the variation in polymorphism across scales when chromosome identity is included as a covariate. The total adjusted r^2 of this model varies from 0.15 (500 kbp windows) to >0.7 (windows 3 Mbp and larger), with recombination rate contributing 45% of the explained variation on average. A somewhat different partitioning of the variance in SNP density results when chromosome identity is excluded from the multivariate analysis: recombination rate continues to act as a significant, strong positive correlate of SNP density across a range of scales, but no significant effect of gene density is observed at any scale, and divergence contributes significantly to variation in SNP density at scales between 3 and 4.5 Mbp, independent of recombination rate ($P < 0.044$ in either forward or reverse oriented windows for each scale; fig. 3).

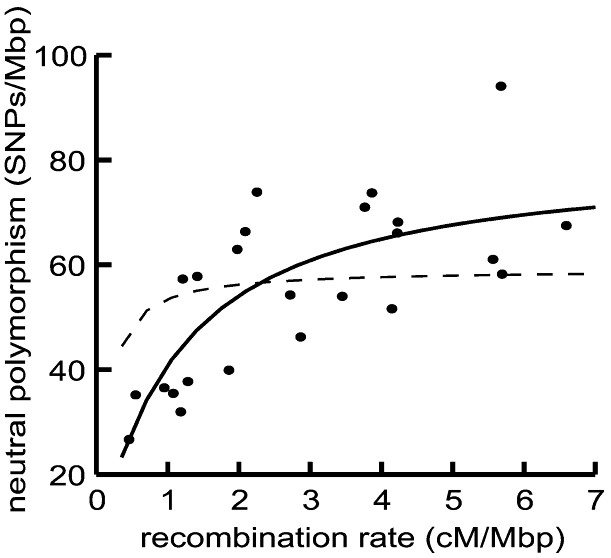

FIG. 2.—The fits of genetic hitchhiking and background selection models with selfing to these data are superimposed on the scatterplot. These curves overlap and are indistinguishable by eye (solid curve). See the tables for parameter values that describe the curve. The dashed curve corresponds to a fit of a version of the background selection model that assumes obligate outcrossing.

Discussion

Our results suggest that natural selection plays a role in shaping genome-wide patterns of neutral DNA sequence variability in a primarily self-fertilizing species, *C. elegans*. Across many genomic scales, nucleotide polymorphism is strongly positively correlated with recombination rate, independent of other factors. Evidence from other studies also suggests that patterns of neutral variation in *C. elegans* are affected by selection. In a survey of polymorphism at two nuclear genes and one mitochondrial gene for multiple strains, nucleotide diversity in *C. elegans* was fivefold lower than the neutral expectation based on comparison with the outcrossing congener *C. remanei* (Graustein et al. 2002). Graustein et al. (2002) argue that this difference may be explained by selection at linked sites, consistent with our interpretations. We also observe associations between polymorphism and other variables that must be considered in interpretations of our results.

Gene Density

A prediction of models of selection at linked sites is that genomic regions with more selective targets will exhibit lower levels of polymorphism. If selection is mostly restricted to coding regions, local gene density may provide a useful index of selection intensity. Under this assumption, gene density should be negatively correlated with nucleotide variation. Nucleotide polymorphism and gene density may be negatively correlated in humans (Payseur and Nachman 2002; but see Lercher and Hurst 2002), although there is little evidence for such a relationship in *D. melanogaster* (Hey and Kliman 2002). As predicted by models of selection at linked sites, SNP density in *C. elegans* is negatively correlated with gene density independent of other variables (provided that chromosome identity is included in the multiple regression analysis; see below). An alternative interpretation for the negative association between polymorphism and gene density is that a higher fraction of noncoding SNP sites may experience stabilizing selection in gene-dense regions. This could occur if conserved regulatory elements are represented disproportionately in gene-dense regions. A comprehensive assignment of function to noncoding regions of the worm genome will clarify the relative contribution of these two possible alternatives.

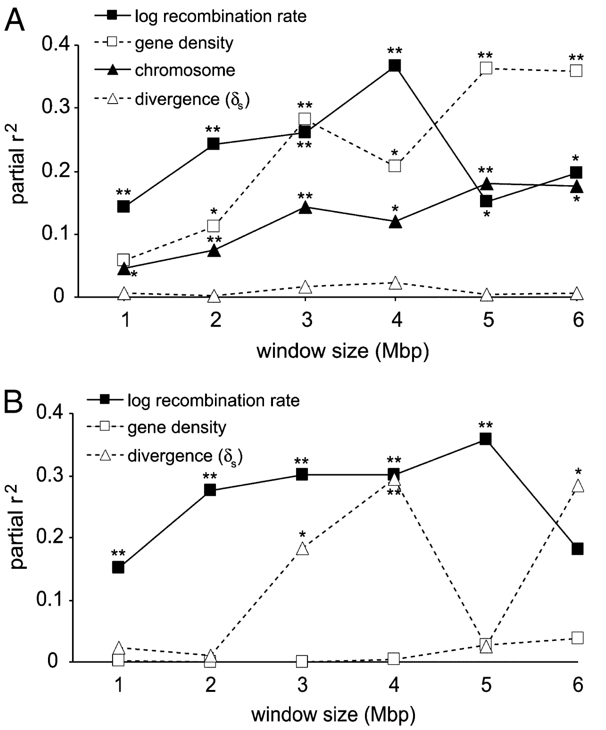

FIG. 3.—Fraction of variation in log SNP density explained at different genomic scales by independent variables in multiple regression analyses. Partial r^2 values are the average from analyses with genomic windows oriented in forward and reverse directions along the chromosomes. Two asterisks (**) indicate a significant effect ($P < 0.05$) with both window orientations, a single asterisk (*) indicates significance with only one orientation. Gene density, but not divergence, consistently contributes significantly to variation in SNP density in (*A*) where a model that includes chromosome identity was used, whereas in (*B*) divergence, and not gene density, is a significant factor at most scales in a model that excludes chromosome. Recombination rate explains a significant fraction of the variance in SNP density regardless of whether chromosome is included as a variable in the multiple regression model.

Divergence

Multiple regression models show that, under some circumstances, the k_s-based measure of divergence δ_s is associated with SNP density independent of other variables. This result suggests that some of the variation in SNP density may be attributable to variation in the neutral mutation rate. However, the robustness of this conclusion depends on the ability of k_s to characterize neutral mutation rates. Several factors affect the quality of

k_s (and therefore δ_s) as a measure of divergence. First, biased usage of codons in *C. elegans* indicates that many synonymous sites experience selection (Stenico, Lloyd, and Sharp 1994; Duret 2000), which complicates the use of k_s as an estimator of the neutral mutation rate. We have attempted to account for this issue by adjusting k_s for codon bias (see *Methods*), although this adjustment will suffice only to the extent to which the F_{op} codon usage statistic accurately captures selection on synonymous sites. Second, a potential problem arises from measuring polymorphism and k_s at different loci. Estimates of divergence from the same noncoding regions where SNP density was measured might provide more appropriate estimates of the neutral mutation rate. Unfortunately, this is not feasible because noncoding regions of *C. elegans* and *C. briggsae* are difficult to align (Shabalina and Kondrashov 1999). Finally, estimating k_s between highly divergent lineages is difficult. *C. elegans* and *C. briggsae* show evidence of saturation at synonymous sites (mean $\delta_s = 1.5$). At this level of divergence, estimates of k_s may be inaccurate. Consequently, the use of δ_s as an indicator of mutational heterogeneity may lead to underestimation of the effect of mutation on variation in SNP density. Additionally, if mutation rates have changed recently, relative to the divergence time between *C. elegans* and *C. briggsae*, k_s may not accurately reflect the mutational environment under which the SNPs arose.

Previous work has suggested that recombination may be mutagenic in *C. elegans* (Marais, Mouchiroud, and Duret 2001). Consistent with this hypothesis, our bivariate analyses uncovered a positive correlation between recombination rate and δ_s (Spearman's $\rho = 0.80$, $P < 0.0001$). Analyses in humans have also indicated that k_s (measured by comparing human and mouse) and recombination rate are correlated (Lercher and Hurst 2002). Along with experimental evidence in yeast (Strathern, Shafer, and McGill 1995; Rattray et al. 2001), these observations collectively provide growing support for the notion that recombination may be mutagenic. Nevertheless, the observation that recombination rate and SNP density are strongly correlated independent of divergence suggests that natural selection shapes genomic patterns of SNP diversity in *C. elegans*.

Chromosome Identity

The inclusion of chromosome identity in multiple regression models does not influence the strong correlation between SNP density and recombination rate; however, its inclusion does affect the relative contribution of gene density and divergence to variation in SNP density (cf. fig. 3*a* and *b*). For example, gene density is not a significant predictor of variation in SNP density when chromosome identity is excluded from multiple regression analyses. It is difficult to construct a biological explanation for the effect of chromosome identity on SNP density. Ascertainment bias at the chromosomal level in the identification of SNPs or mutational differences among chromosomes could account for this effect. We find no evidence, however, that chromosomes differ in SNP density ($F_{5,18} = 0.46$, $P = 0.8$) or δ_s ($F_{5,1374} = 0.88$, $P = 0.5$). If the effect of chromosome identity has a biological basis, then inclusion of this variable in our analyses is appropriate. Without knowledge of the true basis of this effect, interpretation of the relative roles of gene density and divergence as predictors of variation in SNP density requires caution. In contrast, our observation of the relationship between recombination rate and SNP density is unaffected by the inclusion or exclusion of chromosome identity in multiple regression analyses.

Background Selection and Genetic Hitchhiking Models

Is the observed positive correlation between SNP density and recombination rate in *C. elegans* primarily caused by positive or negative selection? A number of approaches have been proposed with the aim of distinguishing between genetic hitchhiking and background selection in outcrossing populations with even sex ratios (Aquadro, Begun, and Kindahl 1994; Andolfatto 2001). However, no theoretical treatments have thoroughly outlined the predictions of these selective models in the context of partial selfing and biased sex ratios. A greater effect of hitchhiking may be expected in partially selfing populations, because selfing rate exerts a much stronger influence on the fixation probability of recessive beneficial mutations than on the fixation probability of deleterious mutations (Charlesworth 1992), although few data are available regarding the average dominance of beneficial mutations. We cannot rigorously evaluate the relative abilities of the two models to explain our results, and both forces likely operate simultaneously (Kim and Stephan 2000). Here, we examine the parameter space that is consistent with each of them.

We can use estimates of the genomic deleterious mutation rate (U) in *C. elegans* to predict the effects of background selection on SNP density. Under background selection, an approximation of the expected level of neutral polymorphism (π) in a genomic region that takes into account partial self-fertilization is

$$\pi = \frac{\pi_0 \cdot \exp\left(-\frac{U}{s_d + r(1-F)}\right)}{1+F}, \qquad (1)$$

where π_0 is the level of polymorphism expected in the absence of selection at linked sites, s_d is the average selection coefficient against deleterious mutations, r is the recombination rate, and the outcrossing rate (c) is related to the inbreeding coefficient as $F = (1 - c)/(1 + c)$ (Charlesworth, Morgan, and Charlesworth 1993; Nordborg 1997, 2000). This relation (as well as the model for hitchhiking described below) demonstrates that in the complete absence of outcrossing ($c = 0$, $F = 1$), neutral polymorphism is not influenced by recombination rate and overall neutral polymorphism should be half that of a purely outcrossing population. Multiple lines of evidence suggest that U is approximately 0.005 to 0.03 mutations per genome per generation in *C. elegans* (Keightley and Caballero 1997; Vassilieva and Lynch 1999; Vassilieva, Hook, and Lynch 2000; Keightley and Bataillon 2000; ADC and BAP, unpublished results). However, these values for U may underestimate the actual genomic

Table 1
Estimated Parameter Values for the Background Selection Model Based on Non-linear Fitting of Equation (1) to SNP Density and Recombination rate Measured at the 4 Mbp Window Size (n = 24)

	Background Selection Model Fit Parameter Estimates (approximate SE)			
Input Parameter	c	s_d	π_0	U
U				
0.50	0.34 (0.35)	0.22 (0.23)	120.7 (49.3)	—
0.10	0.054 (0.043)	0.045 (0.046)	153.8 (29.4)	—
0.05	0.026 (0.021)	0.022 (0.023)	157.9 (27.0)	—
0.01	0.0051 (0.0040)	0.0044 (0.0046)	161.2 (25.0)	—
0.005	0.0026 (0.0020)	0.0022 (0.0023)	161.6 (24.8)	—
0.001	0.00051 (0.00040)	0.00044 (0.00046)	161.9 (24.6)	—
s_d				
0.1	0.13 (0.26)	—	143.5 (52.2)	0.23 (0.23)
0.02	0.023 (0.042)	—	158.3 (29.7)	0.045 (0.047)
0.01	0.012 (0.021)	—	160.2 (27.1)	0.023 (0.023)
0.001	0.0012 (0.0021)	—	161.8 (24.8)	0.0023 (0.0023)

deleterious mutation rate by a factor of ~25 (Davies, Peters, and Keightley 1999). Using nonlinear regression, we fit this model to our estimates of SNP density and recombination rate to obtain estimates of the expected neutral polymorphism level (π_0), the average strength of selection against deleterious mutations (s_d), and the outcrossing rate (c), given the rate of deleterious mutation (U). The background selection model appears to explain the variation in polymorphism due to recombination rate (r^2 = 0.57; fig. 2) to a comparable extent as log-transformed values in a simple linear regression model (r^2 = 0.60). This background selection model, which includes partial selfing, explains significantly more variation in SNP density than a version of the model that excludes partial selfing (extra sum-of-squares F-test $F_{1,21} = 17.6$, $P = 0.0004$; fig. 2). We have summarized in table 1 the estimated parameter values from the background selection model fits. Note that all values of π_0 should be scaled upward by a factor of ~20 to represent the actual SNP density across the entire genome (see *Methods*). Overall, higher rates of deleterious mutation result in higher predicted levels of outcrossing and a higher strength of selection against deleterious mutations (table 1). The levels of outcrossing and strength of selection against deleterious mutations predicted by the background selection model, based on the SNP density and recombination rate data, appear to be reasonable, provided that the deleterious mutation rate is not too high.

We also applied this method to a model of genetic hitchhiking that accounts for partial self-fertilization. Neutral polymorphism under hitchhiking is expected to be

$$\pi = \frac{\pi_0 \cdot r(1-F)}{(1+F) \cdot [r(1-F)+\beta]}, \qquad (2)$$

where $\beta = 2 \cdot N_e \cdot s_a \cdot k \cdot v_0$, N_e is the effective population size, s_a is the average selection coefficient for beneficial mutations, k is approximately 0.075 (Stephan 1995), and v_0 is the expected number of advantageous mutations (Wiehe and Stephan 1993; Nordborg 1997; Kim and Stephan 2000; Nordborg 2000). This model provides a similar fit to the data as the background selection model (fig. 2), but unfortunately it does not allow separate estimation of c, π_0, and β. We assume that the estimate of $2 \cdot N_e \cdot s_a \cdot v_0 \sim 4.6 \times 10^{-8}$ from *D. melanogaster* (Stephan 1995) applies to *C. elegans* (and therefore $\beta \sim 3.73 \times 10^{-9}$). This assumption seems a reasonable first approximation because estimates of N_e are in rough accordance (calculated as $N_e = \pi/4\mu = 0.5 \times 10^6 - 1 \times 10^6$; based on these SNPs, the SNPs of Koch et al. [2000], or the π estimate of Graustein et al. [2002], and the mutation rate estimate μ of Drake et al. [1998]). When partial selfing is excluded from the hitchhiking model ($F = 0$, $c = 1$) for $\beta \sim 3.73 \times 10^{-9}$, the fit of the model is significantly worse (extra sum-of-squares F-test $F_{1,21} = 29.0$, $P < 0.0001$). We summarize in table 2 the predicted parameter values from the genetic hitchhiking model fits for a range of input estimates for β. The hitchhiking model predicts an extremely low estimate for the rate of outcrossing, given the estimate of β derived from *D. melanogaster* (table 2). The value of β must be several orders of magnitude larger than the *D. melanogaster* estimate before the levels of outcrossing predicted by the hitchhiking model approach the levels expected from laboratory and theoretical studies (Hedgecock 1976; Chasnov and Chow 2002; Stewart and Phillips 2002; Cutter, Avilés, and Ward 2003). Because the rate of outcrossing must be at least as high as the frequency of males in populations (Hedgecock 1976; Chasnov and Chow 2002; Stewart and Phillips 2002; Cutter, Avilés, and Ward 2003), this suggests that widespread selection for adaptive mutations may be an unlikely candidate as a general explanation for patterns of neutral polymorphism in *C. elegans*.

Examination of differences between autosomes and sex chromosomes in obligate outcrossers forms the basis for some means of distinguishing between background selection and genetic hitchhiking models (Begun and Whitley 2000). However, a high level of selfing and hermaphrodite-biased sex ratio renders the effective sizes of the X and autosomes essentially identical (E. Pollack and ADC, unpublished results) and the X will spend very little time in a hemizygous state. These effects remove the

Table 2
Estimated Parameter Values for Genetic Hitchhiking Model Based on Non-linear Fitting of Equation (2) to SNP Density and Recombination Rate Measured at the 4 Mbp Window Size (*n* = 24)

	Genetic Hitchhiking Model Fit Parameter Estimates (approximate SE)		
Input Parameter	*c*	β	π_0
β			
3.73×10^{-8}	1.9×10^{-8} (6.5×10^{-9})	—	161.0 (15.3)
3.73×10^{-9}	1.9×10^{-9} (6.5×10^{-10})	—	161.0 (15.3)
3.73×10^{-10}	1.9×10^{-10} (6.5×10^{-11})	—	161.0 (15.3)
c			
0.01	—	0.019 (0.0064)	159.4 (15.1)
0.001	—	0.0019 (0.00065)	160.8 (15.2)

theoretical basis for X–autosome differences in species like *C. elegans*. Here, we observe no significant difference between the X chromosome and the autosomes in SNP density ($P = 0.7$). In light of the equivalent effective sizes of autosomes and the sex chromosome that are expected in *C. elegans*, we suspect that X–autosome comparisons will not aid in discriminating between background selection and genetic hitchhiking models in this species.

Conclusions

Outcrossing occurs with sufficient frequency within *C. elegans* to yield a significant signature of selection across the genome. In the absence of outcrossing, we would not expect selection at linked sites to induce a correlation between neutral polymorphism and recombination rate. However, we observe a clear correlation. If background selection accurately describes the process underlying the relationship between neutral polymorphism and recombination rate, then outcrossing may occur with a frequency of >1%. Two independent sources of evidence for the operation of outcrossing among populations come from mixed SNP profiles among 11 *C. elegans* strains (Koch et al. 2000) and mosaic distributions of transposable elements among strains (Egilmez, Ebert, and Reis 1995). These observations suggest that males may deserve a more prominent role in our understanding of the evolution of *C. elegans* populations. The intriguing question of how even a low level of outcrossing is maintained in this species, given that reproduction does not require the presence of males, remains open.

Reduced levels of genetic variation overall in partially or fully selfing species compared to outcrossing relatives have been observed in several plant clades (Miyashita, Innan, and Terauchi 1996; Liu, Zhang, and Charlesworth 1998; Liu, Charlesworth, and Kreitman 1999; Savolainen et al. 2000; Baudry et al. 2001), but this study provides one of the first unambiguous examples of a relationship between neutral polymorphism and recombination rate within the genome of a partially selfing organism (Dvorak, Luo, and Yang 1998; Baudry et al. 2001; Tenaillon et al. 2001). The strength of this relationship and the wealth of genetic information in *C. elegans*, together, suggest that the genus *Caenorhabditis*, in which species vary in mode of reproduction, may provide a more fertile system than previously recognized for studying the influence of breeding system on patterns of genomic diversity.

Acknowledgments

We thank the Sanger Institute and the Genome Sequencing Center, Washington University, St. Louis for making *C. briggsae* sequence available. We also thank T. Harris and L. Stein for assistance with data retrieval. This manuscript was improved by the constructive comments of P. Andolfatto, C. Birky, A. Eyre-Walker, M. Nachman, M. Nordborg, W. Stephan, J. Storz, and two anonymous reviewers. A Department of Defense National Defense Science and Engineering Graduate Fellowship to A.D.C. facilitated this research. Support for B.A.P. was provided by an Intergrative Graduate Education and Research Training Grant in Biology, Mathematics, and Physics.

Literature Cited

Aguadé, M., N. Miyashita, and C. H. Langley. 1989. Reduced variation in the yellow-achaete-scute region in natural-populations of *Drosophila melanogaster*. Genetics **122**:607–615.

Andolfatto, P. 2001. Adaptive hitchhiking effects on genome variability. Curr. Opin. Genet. Dev. **11**:635–641.

Aquadro, C. F., D. J. Begun, and E. C. Kindahl. 1994. Selection, recombination and DNA polymorphism in *Drosophila*. Pp. 46–56 *in* B. Golding, ed. Non-neutral evolution: theories and molecular data. Chapman and Hall, New York.

Barnes, T. M., Y. Kohara, A. Coulson, and S. Hekimi. 1995. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. Genetics **141**:159–179.

Baudry, E., C. Kerdelhue, H. Innan, and W. Stephan. 2001. Species and recombination effects on DNA variability in the tomato genus. Genetics **158**:1725–1735.

Begun, D. J., and C. F. Aquadro. 1991. Molecular population-genetics of the distal portion of the X-chromosome in *Drosophila*—evidence for genetic hitchhiking of the yellow-achaete region. Genetics **129**:1147–1158.

———. 1992. Levels of naturally-occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. Nature **356**:519–520.

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

Berry, A. J., J. W. Ajioka, and M. Kreitman. 1991. Lack of polymorphism on the *Drosophila* 4th chromosome resulting from selection. Genetics **129**:1111–1117.

Charlesworth, B. 1992. Evolutionary rates in partially self-fertilizing species. Am. Nat. **140**:126–148.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics **134**:1289–1303.

Charlesworth, D., and S. I. Wright. 2001. Breeding systems and genome evolution. Curr. Opin. Genet. Dev. **11**:685–690.

Chasnov, J. R., and K. L. Chow. 2002. Why are there males in the hermaphroditic species *Caenorhabditis elegans*? Genetics **160**:983–994.

Chen, Y., B. J. Marsh, and W. Stephan. 2000. Joint effects of natural selection and recombination on gene flow between *Drosophila ananassae* populations. Genetics **155**: 1185–1194.

Cutter, A. D., L. Avilés, and S. Ward. 2003. The proximate determinants of sex ratio in *C. elegans* populations. Genet. Res. (in press).

Davies, E. K., A. D. Peters, and P. D. Keightley. 1999. High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. Science **285**:1748–1751.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics **148**:1667–1686.

Duret, L. 2000. tRNA gene number and codon usage in the C-elegans genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. **16**:287–289.

Dvorak, J., M. C. Luo, and Z. L. Yang. 1998. Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. Genetics **148**:423–434.

Egilmez, N. K., R. H. Ebert, and R. J. S. Reis. 1995. Strain evolution in *Caenorhabditis elegans*—transposable elements as markers of interstrain evolutionary history. J. Mol. Evol. **40**:372–381.

Fitch, D. H. A., and W. K. Thomas. 1997. Evolution. Pp. 815–850 *in* D. L. Riddle, T. Blumenthal, B. J. Meyer, and J. R. Priess, eds. *C. elegans* II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Graustein, A., J. M. Gaspar, J. R. Walters, and M. F. Palopoli. 2002. Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. Genetics **161**:99–107.

Hedgecock, E. M. 1976. Mating system of *Caenorhabditis elegans*: evolutionary equilibrium between self-fertilization and cross-fertilization in a facultative hermaphrodite. Am. Nat. **110**:1007–1012.

Hedrick, P. W. 1980. Hitchhiking—a comparison of linkage and partial selfing. Genetics **94**:791–808.

Hey, J., and R. M. Kliman. 2002. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. Genetics **160**:595–608.

Hilton, H., R. M. Kliman, and J. Hey. 1994. Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* species complex. Evolution **48**:1900–1913.

Hudson, R. R., and N. L. Kaplan. 1995. Deleterious background selection with recombination. Genetics **141**:1605–1617.

Jensen, M. A., B. Charlesworth, and M. Kreitman. 2002. Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. Genetics **160**:493–507.

Keightley, P. D., and T. M. Bataillon. 2000. Multigeneration maximum-likelihood analysis applied to mutation-accumulation experiments in *Caenorhabditis elegans*. Genetics **154**: 1193–1201.

Keightley, P. D., and A. Caballero. 1997. Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **94**: 3823–3827.

Keightley, P. D., and A. Eyre-Walker. 2000. Deleterious mutations and the evolution of sex. Science **290**:331–333.

Kim, Y., and W. Stephan. 2000. Joint effects of genetic hitch-hiking and background selection on neutral variation. Genetics **155**:1415–1427.

Koch, R., H. G. A. M. van Luenen, M. van der Horst, K. L. Thijssen, and R. H. A. Plasterk. 2000. Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. Genome Res. **10**:1690–1696.

Kraft, T., T. Sall, I. Magnusson-Rading, N. O. Nilsson, and C. Hallden. 1998. Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). Genetics **150**:1239–1244.

Lercher, M. J., and L. D. Hurst. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet. **18**:337–340.

Li, W. H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36**:96–99.

Liu, F., D. Charlesworth, and M. Kreitman. 1999. The effect of mating system differences on nucleotide diversity at the phosphoglucose isomerase locus in the plant genus *Leavenworthia*. Genetics **151**:343–357.

Liu, F., L. Zhang, and D. Charlesworth. 1998. Genetic diversity in *Leavenworthia* populations with different inbreeding levels. Proc. R. Soc. Lond. Ser. B **265**:293–301.

Marais, G., and L. Duret. 2001. Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. J. Mol. Evol. **52**:275–280.

Marais, G., D. Mouchiroud, and L. Duret. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. Proc. Natl. Acad. Sci. USA **98**:5688–5692.

Maynard Smith, J., and J. Haigh. 1974. Hitch-hiking effect of a favorable gene. Genet. Res. **23**:23–35.

Miyashita, N. T., H. Innan, and R. Terauchi. 1996. Intra- and interspecific variation of the alcohol dehydrogenase locus region in wild plants *Arabis gemmifera* and *Arabidopsis thaliana*. Mol. Biol. Evol. **13**:433–436.

Moriyama, E. N., and J. R. Powell. 1996. Intraspecific nuclear DNA variation in *Drosophila*. Mol. Biol. Evol. **13**: 261–277.

Nachman, M. W. 1997. Patterns of DNA variability at X-linked loci in *Mus domesticus*. Genetics **147**:1303–1316.

———. 2001. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. **17**:481–485.

Nachman, M. W., V. L. Bauer, S. L. Crowell, and C. F. Aquadro. 1998. DNA variability and recombination rates at X-linked loci in humans. Genetics **150**:1133–1141.

Nordborg, M. 1997. Structured coalescent processes on different time scales. Genetics **146**:1501–1514.

———. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics **154**:923–929.

Nordborg, M., B. Charlesworth, and D. Charlesworth. 1996. The effect of recombination on background selection. Genet. Res. **67**:159–174.

Pamilo, P., and N. O. Bianchi. 1993. Evolution of the *zfx* and *zfy* genes—rates and interdependence between the genes. Mol. Biol. Evol. **10**:271–281.

Payseur, B. A., and M. W. Nachman. 2002. Gene density and human nucleotide polymorphism. Mol. Biol. Evol. **19**:336–340.

Przeworski, M., R. R. Hudson, and A. Di Rienzo. 2000. Adjusting the focus on human variation. Trends Genet. **16**:296–302.

Rattray, A. J., C. B. McGill, B. K. Shafer, and J. N. Strathern. 2001. Fidelity of mitotic double-strand-break repair in *Saccharomyces cerevisiae*: a role for SAE2/COM1. Genetics **158**:109–122.

Savolainen, O., C. H. Langley, B. P. Lazzaro, and H. Freville. 2000. Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. Mol. Biol. Evol. **17**:645–655.

Shabalina, S. A., and A. S. Kondrashov. 1999. Pattern of selective constraint in *C. elegans* and *C. briggsae* genomes. Genet. Res. **74**:23–30.

Stenico, M., A. T. Lloyd, and P. M. Sharp. 1994. Codon usage in *Caenorhabditis elegans*—delineation of translational selection and mutational biases. Nucleic Acids Res. **22**:2437–2446.

Stephan, W. 1995. An improved method for estimating the rate of fixation of favorable mutations based on DNA polymorphism data. Mol. Biol. Evol. **12**:959–962.

Stephan, W., and C. H. Langley. 1989. Molecular genetic-variation in the centromeric region of the X-chromosome in 3 *Drosophila ananassae* populations. 1. Contrasts between the vermilion and forked loci. Genetics **121**:89–99.

———. 1998. DNA polymorphism in *Lycopersicon* and crossing-over per physical length. Genetics **150**:1585–1593.

Stewart, A. D., and P. C. Phillips. 2002. Selection and maintenance of androdioecy in Caenorhabditis elegans. Genetics **160**:975–982.

Strathern, J. N., B. K. Shafer, and C. B. McGill. 1995. DNA-synthesis errors associated with double-strand-break repair. Genetics **140**:965–972.

Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley, and B. S. Gaut. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp *mays* L.). Proc. Natl. Acad. Sci. USA **98**:9161–9166.

The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282**:2012–2018.

Vassilieva, L. L., A. M. Hook, and M. Lynch. 2000. The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. Evolution **54**:1234–1246.

Vassilieva, L. L., and M. Lynch. 1999. The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. Genetics **151**:119–129.

Wang, W., K. Thornton, A. Berry, and M. Y. Long. 2002. Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. Science **295**:134–137.

Wicks, S. R., R. T. Yeh, W. R. Gish, R. H. Waterston, and R. H. A. Plasterk. 2001. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. Nat. Genet. **28**:160–164.

Wiehe, T. H. E., and W. Stephan. 1993. Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. Mol. Biol. Evol. **10**:842–854.

Adam Eyre-Walker, Associate Editor

Accepted December 12, 2002