

Published in final edited form as:

Science. 2010 October 15; 330(6002): 372–376. doi:10.1126/science.1194208.

Selection at linked sites shapes heritable phenotypic variation in *C. elegans*

Matthew V. Rockman^{1,2}, Sonja S. Skrovaneck^{2,3}, and Leonid Kruglyak^{2,3}

¹Department of Biology and Center for Genomics and Systems Biology, New York University, 100 Washington Square East, New York, New York 10003, USA.

²Lewis-Sigler Institute for Integrative Genomics and Department of Ecology and Evolutionary Biology, Carl Icahn Laboratory, Princeton University, Princeton, New Jersey 08544, USA.

³Howard Hughes Medical Institute, Princeton University, Princeton, New Jersey 08544, USA.

Abstract

Mutation generates the heritable variation that genetic drift and natural selection shape. In classical quantitative genetic models, drift is a function of the effective population size and acts uniformly across traits, while mutation and selection act trait-specifically. We identified thousands of quantitative trait loci (QTL) influencing transcript abundance traits in a cross of two *C. elegans* strains; although trait-specific mutation and selection explained some of the observed pattern of QTL distribution, the pattern was better explained by trait-independent variation in the intensity of selection on linked sites. Our results suggest that traits in *C. elegans* exhibit different levels of variation less because of their own attributes than because of differences in the effective population sizes of the genomic regions harboring their underlying loci.

Some phenotypes exhibit abundant heritable variation and others almost none. As heritable variation is the raw material for adaptation, the forces that shape its distribution across traits are a central concern of evolutionary genetics (1). Among wild strains of the partially selfing nematode *C. elegans*, transcript abundance traits — model quantitative phenotypes (2–7) — differ in their levels of heritable variation (4, 8) and, on the basis of experimental measurements of the rate at which mutation increases their variance, they exhibit lower levels of heritable variation than expected under neutral mutation-drift equilibrium (4). These findings and similar results in other species are consistent with the prediction that trait-dependent stabilizing selection should result in different levels of variation among traits (3–7).

To genetically dissect the causes of different variabilities among *C. elegans* traits, we measured transcript abundances by microarray in developmentally synchronized young adult hermaphrodites of 208 recombinant inbred advanced intercross lines from a cross between the laboratory strain, N2, and a wild isolate from Hawaii, CB4856 (9). These strains, though relatively divergent for *C. elegans*, are closely related, differing at roughly one basepair per 900 (10). Each line was genotyped at 1,455 SNP markers. Interval mapping for each of 15,888 traits identified 2,309 quantitative trait loci at a false discovery rate (FDR) of 5% (Fig. 1A) (11).

The majority of QTLs (65%) are local; that is, these QTLs occur at the genomic locations of the genes whose transcript abundances they influence (the spatial coincidence is defined here by overlap between the 1-lod QTL support interval and the gene). Nearly a quarter of the remaining QTLs (distant QTLs) map to three statistically robust hotspots (11) (Fig 1A, Fig S1). The X-linked hotspot encompasses more than a megabase and probably contains multiple causal variants, one of which may be the known pleiotropic F215V mutation in the

neuropeptide receptor *npr-1* (12). Candidate genes for the other hotspots include a putative SAGA complex component, *Y17G9B.8*, whose transcript abundance maps strongly to a local QTL at its position in the hotspot on the left side of chromosome IV, and *Y105C5A.15*, a putative zinc-finger transcription factor whose transcript abundance maps locally to a QTL at its position in the hotspot on the right side of chromosome IV.

The global distribution of QTLs is strikingly non-uniform. Both local and distant QTLs are strongly enriched in the arms of the chromosomes relative to the centers (Table 1). *C. elegans* lacks heterochromatic centromeres, and the chromosomes are structured in semi-discrete domains that exhibit correlated variation in gene density, evolutionary conservation, repeat sequence density, and recombination rate (9, 13, 14). The chromosomal centers have high gene density and low recombination rates, while chromosome arms have lower gene density and higher recombination rates. Chromosome tips have an intermediate gene density but effectively no recombination (9). Under a simple mutational null model, QTL density is expected to correlate with the density of potentially functional sites and hence to be higher in chromosomal centers than in arms, contrary to the observed pattern. Furthermore, as QTL detection is most favored in low-recombination areas (15, 16), the observed pattern also runs counter to the expected effect of mapping bias.

The chromosomal patterning of causal variants is particularly pronounced for local QTLs, which we confirmed in a focused single-marker analysis (17), which increased detection power over our initial genome scan. We identified 2,538 transcripts affected by QTL that are linked to their own genomic locations at a 5% FDR (Fig. 1B). We found that 23.7% of transcripts in chromosome arms and 20.1% of those in chromosome tips have local QTLs, compared to only 10.2% of those in chromosome centers ($\chi^2=495.7, p < 10^{-107}$). The chromosomal patterning is robust to confounding by potential hybridization artifacts, as demonstrated by analysis of only the 7,694 transcripts for which the CB4856 genotype is associated with higher expression than the reference N2 genotype. The 1,057 significant local QTLs among these exhibit the same pattern of enrichment: 20.0% of arm transcripts, 17.9% of tip transcripts, and 9.6% of center transcripts have significant local QTLs ($\chi^2=162.7, p < 10^{-35}$).

We corroborated the results of linkage mapping by estimating the amount of heritable phenotypic variation attributable to each type of chromosomal domain using a genome-partitioning approach that avoids assumptions about the number, location, and effect sizes of QTLs (11, 18). We estimated the amount of genetic variance attributable to chromosomal arms versus centers for each of the 1,191 traits that are significantly heritable by this method (FDR = 0.05; Fig S2), and we observed an excess of both arm-biased and center-biased traits (Fig. S3), consistent with contributions from large-effect or spatially clustered loci. A significant majority of heritable traits are arm-biased (permutation two-tailed $p = 0.0325$). The arm bias remains when the effects of local QTLs are removed by linear regression ($p \leq 0.0025$), and the pattern is not driven by the QTL hotspots (11, Fig S7).

Several non-exclusive models may explain global patterns of variation in the density of functional variants influencing transcript abundance traits (1, 3–7, 19–21). In standard multivariate quantitative genetic models, equilibrium trait variation results from mutation, selection, and drift, the latter governed by effective population size (N_e) and acting uniformly across traits (22). We asked whether mutation and selection could explain why some transcript abundance traits are influenced by their own genomic loci and why others are not. We focused on these local QTLs because they represent largely independent genetic variants, they are precisely localized, and they account for a large fraction of the phenotypic variance in traits with local QTLs (Fig. 1B).

Variation in local QTL density should reflect variation in rates of local mutational input. In *C. elegans*, the rate of spontaneous single-base mutation has been directly measured and is uniform on a chromosomal scale, with no dependence on recombination rate or domain structure (23). Consequently, the rate of mutation that generates local QTLs probably depends on the local mutational target size. Indeed, genes with local QTLs are longer than those without (*t*-test on log-transformed lengths, $p = 0.004$).

Variation in QTL density should also reflect variation in the intensity of purifying selection, which eliminates mutations that adversely affect the phenotype. We used measurable correlates of purifying selection to test this model. Genes that exhibit phenotypes when their expression is knocked down by RNAi (effectively essential genes; nearly all characterized RNAi phenotypes would be lethal in nature; (11)) are less likely to have local QTLs than genes with no RNAi phenotype ($\chi^2 = 55.1$, $p < 2 \times 10^{-13}$). Moreover, we observed fewer evolutionarily constrained nucleotides in genes with local QTLs (11; genes include introns and flanking sequence) than in those of genes without (*t*-test on Box-Cox transformed values, $p < 4 \times 10^{-23}$).

Phenotypic variance not attributable to local QTLs, including measurement error and environmental variance as well as distant genetic effects, does not differ significantly between transcripts with and without local QTLs (*t*-test on log-transformed data, $p = 0.93$). However, traits with local QTLs are more likely than traits without to also map to additional QTLs ($\chi^2 = 63.2$, $p < 2 \times 10^{-15}$). Thus traits that can withstand local genetic variation can also withstand other genetic perturbations, consistent with these transcript abundances experiencing weaker stabilizing selection compared to other genes.

To determine whether the variables associated with mutational target size and strength of selection have independent effects on local QTL probability, we tested their explanatory value in multiple logistic regression models. Gene interval length, number of conserved bases, RNAi phenotype, and presence of distant QTLs are all significant predictors of local QTL probability in a model that includes them all (model M1 in Table 2).

However, when the chromosomal domain of each gene (tip, arm, or center) was included as a factor (model M2), it was by far the most explanatory variable. In fact, chromosomal domain alone (model M3) explained the QTL data better than a model incorporating all of the gene-level attributes, even when all interactions among the variables were included (model M4). Genic point estimates of the recombination rate, although significant if domain type was excluded (model M5), had no significant explanatory value after taking the domains into account (M6). Thus the domain patterning of local QTLs is not explained by gene-level measures of mutation, selection, or recombination.

Although the effective population size (N_e), which governs genetic drift, is shared by all measured traits, natural selection can cause variation in apparent N_e along the genome. Selection — positive or negative — causes alleles in future generations to be descended from a smaller subset of current alleles than would occur without selection, decreasing the effective population size of the linked genomic interval (24, 25, 26). In *C. elegans*, high levels of self-fertilization reduce the effective recombination rate, increasing the effect of selection at linked sites on standing variation at the level of sequence polymorphism (23, 27–29).

In primarily selfing species with small effective population sizes, such as *C. elegans*, background selection, the reduction in neutral variation due to linkage between neutral variants and deleterious mutations undergoing deterministic elimination from the population (26), is likely to be the predominant form of linked selection (28, 30), and it provides a parsimonious explanation for patterns of variation given the certainty that deleterious

mutations arise and are eliminated by selection. Although hitchhiking due to positive selection may also be operating, data from *C. briggsae*, a nematode that shares *C. elegans*' mating system, strongly favor background selection over the alternative models of selection at linked sites (30). Under background selection, the level of neutral variation at a gene is a function of the number of linked sites susceptible to deleterious mutation and the effective rate of recombination between each such site and the gene. We fitted an explicit model of background selection to each gene (26, 31), estimating the physical distribution of deleterious mutations from comparative genomic data and considering a range of values for two poorly constrained parameters, the strength of selection against deleterious mutations and the inbreeding coefficient, F , whose complement ($P = 1 - F$) rescales the meiotic recombination rate to yield the effective rate in partially selfing species (11).

Background selection was a highly significant ($p < 10^{-80}$, model M7) predictor of local QTL probability in logistic regression analyses that include all of the gene-specific mutation and selection variables, and it entirely accounts for the effect of domain type (model M8). Background selection accounts for more of the explained deviance than all gene-specific variables combined, across nearly all of the parameter space of inbreeding and selection intensity (Fig. 2A, Fig. S4).

These results were robust to variation in deleterious mutation rate, alternative treatments of the genetic map and genic variables, different significance thresholds for linkage, alternative modeling methods, and exclusion of all genes susceptible to hybridization artifacts (Fig. S5). Although our model omits the effects of Hill-Robertson interference between linked mutations, such effects are expected to operate primarily as a scaling factor on the expected reduction in variation due to background selection (32). The background selection model that best explains the data predicts high levels of neutral variation on the chromosome arms and low levels in the centers (Fig. 2B). The low-recombination chromosome tips are more similar to the high-recombination arms than the low-recombination centers because they are linked to deleterious mutations only on one side.

Although the effects of selection on linked neutral nucleotide polymorphism are widely recognized, we have shown that such selection at linked sites is also a major factor shaping heritable phenotypic variation. Consequently, quantitative genetic models predicated on uniform effects of genetic drift across traits are not valid in *C. elegans*.

Transcript abundances in *C. elegans*, as in other species, are undoubtedly shaped by trait-specific mutation rates and selection pressures (3–7, 19–21). At the global level, however, the propensity of traits to vary in *C. elegans* is explained by processes independent of the functions of the individual transcripts. These findings provide an alternative explanation for the observed discordance between standing phenotypic variation in *C. elegans* and that predicted from neutral mutation-drift equilibrium (4). It may also explain the fine-scale correlation between *cis*-acting regulatory polymorphism and gene density in humans (20).

Natural selection and quantitative genetic analyses both rely on replicated measurements of the marginal effects of alleles across randomized genetic backgrounds. We have used quantitative genetics in *C. elegans* to show that randomization in this partially selfing species is ineffective, diminishing the ability of natural selection to evaluate individual alleles. Consequently the evolutionary fates of alleles – and hence phenotypes – are determined less by their own effects than by the genomic company they keep.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References and Notes

1. Houle D. How should we explain variation in the genetic variance of traits? *Genetica*. 1998; 102–103:241–253.
2. Brem RB, Kruglyak L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci U S A*. 2005; 102:1572–1577. [PubMed: 15659551]
3. Rifkin SA, Kim J, White KP. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat Genet*. 2003; 33:138–144. [PubMed: 12548287]
4. Denver DR, et al. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat Genet*. 2005; 37:544–548. [PubMed: 15852004]
5. Ronald J, Akey JM. The evolution of gene expression QTL in *Saccharomyces cerevisiae*. *PLoS One*. 2007; 2 e678.
6. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. Genetic properties influencing the evolvability of gene expression. *Science*. 2007; 317:118–121. [PubMed: 17525304]
7. Fay JC, Wittkopp PJ. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity*. 2008; 100:191–199. [PubMed: 17519966]
8. Li Y, et al. Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet*. 2006; 2 e222.
9. Rockman MV, Kruglyak L. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet*. 2009; 5 e1000419.
10. Wicks SR, Yeh RT, Gish WR, Waterston RH, Plasterk RH. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nat Genet*. 2001; 28:160–164. [PubMed: 11381264]
11. Information on materials and methods is available on *Science* Online.
12. de Bono M, Bargmann CI. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell*. 1998; 94:679–689. [PubMed: 9741632]
13. Barnes TM, Kohara Y, Coulson A, Hekimi S. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics*. 1995; 141:159–179. [PubMed: 8536965]
14. C. e. S. Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*. 1998; 282:2012–2018. [PubMed: 9851916]
15. Genissel A, McIntyre LM, Wayne ML, Nuzhdin SV. Cis and trans regulatory effects contribute to natural variation in transcriptome of *Drosophila melanogaster*. *Mol Biol Evol*. 2008; 25:101–110. [PubMed: 17998255]
16. Noor MA, Cunningham AL, Larkin JC. Consequences of recombination rate variation on quantitative trait locus mapping studies. Simulations based on the *Drosophila melanogaster* genome. *Genetics*. 2001; 159:581–588. [PubMed: 11606535]
17. Ronald J, Brem RB, Whittle J, Kruglyak L. Local regulatory variation in *Saccharomyces cerevisiae*. *PLoS Genet*. 2005; 1 e25.
18. Visscher PM, et al. Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am J Hum Genet*. 2007; 81:1104–1110. [PubMed: 17924350]
19. Lawniczak MK, Holloway AK, Begun DJ, Jones CD. Genomic analysis of the relationship between gene expression variation and DNA polymorphism in *Drosophila simulans*. *Genome Biol*. 2008; 9:R125. [PubMed: 18700012]
20. Tung J, Fedrigo O, Haygood R, Mukherjee S, Wray GA. Genomic features that predict allelic imbalance in humans suggest patterns of constraint on gene expression variation. *Mol Biol Evol*. 2009; 26:2047–2059. [PubMed: 19506001]
21. Ayroles JF, et al. Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet*. 2009; 41:299–307. [PubMed: 19234471]
22. Arnold SJ, Bürger R, Hohenlohe PA, Ajie BC, Jones AG. Understanding the evolution and stability of the G-matrix. *Evolution*. 2008; 62:2451–2461. [PubMed: 18973631]
23. Denver DR, et al. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc Natl Acad Sci U S A*. 2009; 106:16310–16314. [PubMed: 19805298]

24. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. *Genet Res.* 1966; 8:269–294. [PubMed: 5980116]
25. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res.* 1974; 23:23–35. [PubMed: 4407212]
26. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. *Genetics.* 1993; 134:1289–1303. [PubMed: 8375663]
27. Graustein A, Gaspar JM, Walters JR, Palopoli MF. Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. *Genetics.* 2002; 161:99–107. [PubMed: 12019226]
28. Cutter AD, Payseur BA. Selection at linked sites in the partial selfer *Caenorhabditis elegans*. *Mol Biol Evol.* 2003; 20:665–673. [PubMed: 12679551]
29. Jovelin R. Rapid sequence evolution of transcription factors controlling neuron differentiation in *Caenorhabditis*. *Mol Biol Evol.* 2009; 26:2373–2386. [PubMed: 19589887]
30. Cutter AD, Choi JY. Natural selection shapes nucleotide polymorphism across the genome of the nematode *Caenorhabditis briggsae*. *Genome Res.* 2010
31. Hudson RR, Kaplan NL. Deleterious background selection with recombination. *Genetics.* 1995; 141:1605–1617. [PubMed: 8601498]
32. Kaiser VB, Charlesworth B. The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet.* 2009; 25:9–12. [PubMed: 19027982]
33. We thank E. Andersen, A. Chang, J. Gerke, R. Ghosh, D. Gresham, M. Hahn, A. Paaby, D. Pollard, H. Seidel, and J. Shapiro for comments on the manuscript. We thank the *Caenorhabditis* Genetics Center, funded by the NIH National Center for Research Resources, for strains. Our work was supported by the NIH (R01 HG004321 to LK, R01 GM089972 to MVR, and P50 GM071508 to the Lewis-Sigler Institute), a Jane Coffin Childs Fellowship (MVR), an Ellison Foundation New Scholar Award (MVR), and a James S. McDonnell Foundation Centennial Fellowship (LK). LK is an investigator of the Howard Hughes Medical Institute. Microarray data have been deposited at the Gene Expression Omnibus with accession number pending.

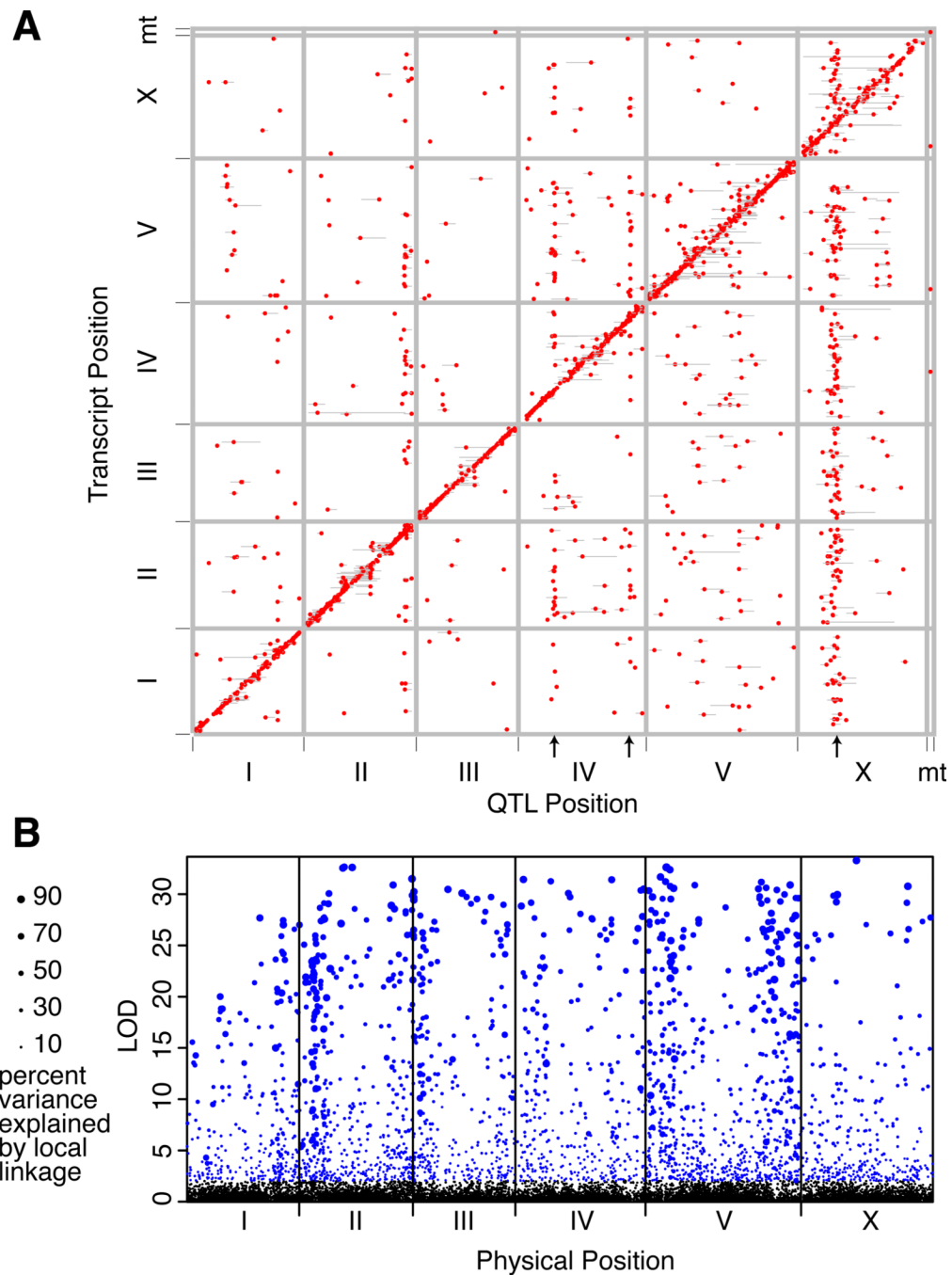


Figure 1.

A. QTLs for each transcript abundance phenotype, significant at a False Discovery Rate of 5%, are plotted in rows located at the genomic positions of the transcripts. Gray bars represent 1-ld support intervals. The diagonal includes local QTLs, those that co-localize with the transcript they affect. Three robust QTL hotspots are indicated with arrows. **B.** Local lod score is plotted for each probe at its physical position along the chromosomes. Points in blue are significant at a 5% False Discovery Rate according to a single-marker linkage test. Points are scaled according to the fraction of variance in transcript abundance explained by the local QTL.

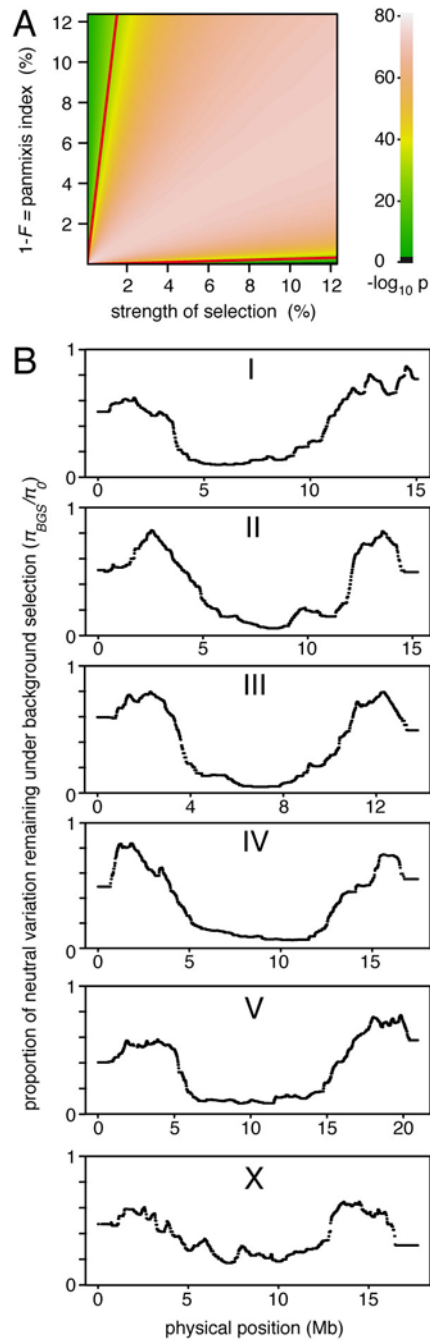


Figure 2.

A. The significance of background selection in a logistic regression model (which includes gene-specific mutation and selection variables) is plotted as a function of the index of panmixis and strength of selection against deleterious mutations. Background selection is significant at $p < 0.01$ across all but a small slice of parameter space corresponding to very low rates of outcrossing (black). The red lines bracket the region of parameter space over which background selection explains more of the local linkage probability than any other variable in the model. See Figure S4. **B.** Effects of background selection on levels of variation along the chromosomes under the best fitting background selection model.

Table 1

Both distant and local QTLs are overrepresented in chromosome arms relative to centers.

	Distant QTLs	Local QTLs	Genome	Probes
Tip	25 (3.1%)	93 (6.2%)	7.4 Mb (7.3%)	987 (6.2%)
Arm	508 (62.5%)	935 (62.5%)	45.9 Mb (45.8%)	6049 (38.1%)
Center	277 (34.1%)	466 (31.1%)	47.0 Mb (46.9%)	8843 (55.7%)
Mitochondrial	3 (0.4%)	2 (0.1%)	0.014 Mb (0.01%)	9 (0.1%)
Total	813	1496	100.3 Mb (100%)	15888

	MODEL 1		MODEL 2		MODEL 3		MODEL 4	
	LRT	$-\log_{10}(p)$	LRT	$-\log_{10}(p)$	LRT	$-\log_{10}(p)$	LRT	$-\log_{10}(p)$
background selection					363.6	80.32	100.7	22.98
explained deviance	389.9		602.1		701.4		701.6	
residual deviance	12507.6		12295.4		12196.1		12195.9	
Δ df	5		7		5		7	
AIC	12519.6		12311.4		12208.1		12211.9	