tion using complexes formed with ^{32}P-labeled mRNA and found that the mRNA segregated predominantly with 40*S* subunits (fig. S8), consistent with the previously reported dependence of mRNA dissociation on additional initiation factors (*23*). Additionally, no cleavage of mRNA was observed during Dom34:Hbs1-mediated processes (fig. S9), consistent with an earlier report (*8*).

Because Dom34:Hbs1-driven subunit dissociation and peptidyl-tRNA release are correlated events (Fig. 3, A and B), we sought to determine their order by measuring the rate constants of the two reactions using ribosome complexes containing either ^{32}P-labeled mRNA or ^{35}S-Met–labeled peptidyl-tRNA (Fig. 3C). The observed rate constants for the disappearance of 80*S* ribosomes in a native gel were closely matched (~0.15 min^{-1}) when either the labeled mRNA or peptidyl-tRNA was monitored. These data indicate that subunit dissociation and peptidyl-tRNA formation by Dom34:Hbs1 are tightly coupled to one another, with one likely serving as the rate-limiting step for the other [e.g., (*24*)].

If subunit dissociation occurs first, it seemed possible that an intermediate product might be generated in which subunits have dissociated but peptidyl-tRNA has not yet departed (i.e., peptidyl-tRNA:40*S* subunit conjugates). In a comparison of Dom34:Hbs1 activity on an initiation-like ribosome complex (carrying initiator Met-tRNA^{iMet} in the P site) and on an elongated ribosome complex (carrying Met-Phe-tRNA^{Phe} in the P site), free Met-Phe-tRNA^{Phe} was the predominant product from the Dom34:Hbs1-treated elongated ribosome complex, whereas Met-tRNA^{iMet}–bound 40*S* subunit was the predominant product from the Dom34:Hbs1-treated initiation complexes (Fig. 3D and fig. S10A). The appearance of a Met-tRNA^{iMet}–bound 40*S* complex suggests that subunit separation can take place independently of peptidyl-tRNA release. To confirm that the observed Met-tRNA^{iMet}–bound 40*S* complex represented an authentic stable product of the Dom34:Hbs1-catalyzed reaction, rather than reassociation of Met-tRNA^{iMet} with ribosomes after initial dissociation, we repeated the Dom34:Hbs1-catalyzed reaction in the presence of a large excess of unlabeled Met-tRNA^{iMet} and found that the chase had no effect on the reaction products (fig. S10B). These data are broadly consistent with the previously reported high affinity of 40*S* subunits for Met-tRNA^{iMet} (*25*) and show that the Dom34:Hbs1 complex initially promotes subunit dissociation and that peptidyl-tRNA dissociation typically follows.

Given the structural similarities between Dom34:Hbs1 and eRF1:eRF3, we wondered whether the canonical eukaryotic release factors might also promote subunit separation, independent of peptide release, and thereby contribute to ribosome recycling during termination. Treatment of termination complexes (with ^{35}S-Met–labeled peptidyl-tRNA) with catalytically inactive eRF1(AGQ):eRF3 led to the formation of free peptidyl-tRNA in a reaction inhibited by GDPNP. This activity was distinguished from that promoted by Dom34:Hbs1 only by its slower rate (0.012 min^{-1} versus 0.21 min^{-1}) (Fig. 4A). Like the Dom34:Hbs1-catalyzed reaction, the eRF1(AGQ):eRF3 reaction depended on both protein components for full activity (Fig. 4B). Lastly, we found that eRF1(AGQ):eRF3-mediated peptidyl-tRNA release exhibited robust codon specificity, taking place only when a stop codon was presented in the A site (Fig. 4C).

Canonical recycling, which occurs after termination, involves subunit dissociation, and mRNA and tRNA release, thus allowing for subsequent reinitiation of translation. In bacteria, a specialized ribosome recycling factor, RRF, is central to this GTP-dependent process (*26*). However, in eukaryotes, no RRF has been identified. Our results indicate that eRF1:eRF3 and Dom34:Hbs1 directly destabilize the subunit interface to promote recycling. Although additional factors (including translation elongation factors) appear to promote or accelerate various aspects of recycling in yeast and mammals (*23*, *27*–*29*), our observations could explain why no true RRF homolog is present in eukaryotes where "termination-like" factors instead play the key role in destabilizing the subunit interface. We further argue that Dom34:Hbs1 acts as a specialized recycling factor on malfunctioning ribosome complexes that, for example, do not appropriately engage the next factor in the translation cycle or are inherently defective and thus unable to properly elongate (i.e., NRD) (*6*). Subsequent to Dom34:Hbs1-mediated recycling, kinetic competition between translation reinitiation and mRNA decay (or rRNA decay in the case of NRD) will determine the partitioning of defective RNAs, with the opportunities for degradation accumulating with each passage through the quality-control pathway.

References and Notes

1. S. W. Peltz, A. H. Brown, A. Jacobson, *Genes Dev.* **7**, 1737 (1993).
2. P. A. Frischmeyer *et al.*, *Science* **295**, 2258 (2002).
3. A. van Hoof, P. A. Frischmeyer, H. C. Dietz, R. Parker, *Science* **295**, 2262 (2002).
4. M. K. Doma, R. Parker, *Nature* **440**, 561 (2006).
5. R. Gandhi, M. Manzoor, K. A. Hudak, *J. Biol. Chem.* **283**, 32218 (2008).
6. S. E. Cole, F. J. LaRiviere, C. N. Merrikh, M. J. Moore, *Mol. Cell* **34**, 440 (2009).
7. J. Soudet, J. P. Gélugne, K. Belhabich-Baumas, M. Caizergues-Ferrer, A. Mougin, *EMBO J.* **29**, 80 (2010).
8. D. O. Passos *et al.*, *Mol. Biol. Cell* **20**, 3025 (2009).
9. A. Carr-Schmid, C. Pfund, E. A. Craig, T. G. Kinzy, *Mol. Cell. Biol.* **22**, 2564 (2002).
10. M. Graille, M. Chaillet, H. van Tilbeurgh, *J. Biol. Chem.* **283**, 7145 (2008).
11. G. C. Atkinson, S. L. Baldauf, V. Hauryliuk, *BMC Evol. Biol.* **8**, 290 (2008).
12. H. H. Lee *et al.*, *Mol. Cell* **27**, 938 (2007).
13. M. G. Acker, S. E. Kolitz, S. F. Mitchell, J. S. Nanda, J. R. Lorsch, *Methods Enzymol.* **430**, 111 (2007).
14. L. Frolova *et al.*, *RNA* **2**, 334 (1996).
15. P. Saini, D. E. Eyler, R. Green, T. E. Dever, *Nature* **459**, 118 (2009).
16. M. A. Algire *et al.*, *RNA* **8**, 382 (2002).
17. Z. Vogel, T. Vogel, A. Zamir, D. Elson, *Eur. J. Biochem.* **21**, 582 (1971).
18. E. M. Youngman, J. L. Brunelle, A. B. Kochaniak, R. Green, *Cell* **117**, 589 (2004).
19. E. Z. Alkalaeva, A. V. Pisarev, L. Y. Frolova, L. L. Kisselev, T. V. Pestova, *Cell* **125**, 1125 (2006).
20. L. Y. Frolova *et al.*, *RNA* **5**, 1014 (1999).
21. M. Gartmann *et al.*, *J. Biol. Chem.* **285**, 14848 (2010).
22. K. Si, U. Maitra, *Mol. Cell. Biol.* **19**, 1416 (1999).
23. A. V. Pisarev, C. U. Hellen, T. V. Pestova, *Cell* **131**, 286 (2007).
24. A. Savelsbergh *et al.*, *Mol. Cell* **11**, 1517 (2003).
25. S. E. Kolitz, J. E. Takacs, J. R. Lorsch, *RNA* **15**, 138 (2009).
26. A. V. Zavialov, V. V. Hauryliuk, M. Ehrenberg, *Mol. Cell* **18**, 675 (2005).
27. A. V. Pisarev *et al.*, *Mol. Cell* **37**, 196 (2010).
28. S. Khoshnevis *et al.*, *EMBO Rep.* **11**, 214 (2010).
29. S. Kurata *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10854 (2010).

30. We thank J. Lorsch and his lab members R. Parker and M. Moore for valuable discussions, H. Zaher and N. Guydosh for thoughtful comments on the manuscript, and B. Rogers for cloning and purification of Pth and eIF6.

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6002/369/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 and S2
References

18 May 2010; accepted 16 August 2010
10.1126/science.1192430

Selection at Linked Sites Shapes Heritable Phenotypic Variation in *C. elegans*

Matthew V. Rockman,[1,2]* Sonja S. Skrovanek,[2,3] Leonid Kruglyak[2,3]*

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, whereas mutation and selection act trait-specifically. We identified thousands of quantitative trait loci (QTLs) influencing transcript abundance traits in a cross of two *Caenorhabditis 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 *Caenorhabditis 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 1 base pair per 900 (*10*). Each line was genotyped at 1455 single-nucleotide polymorphism (SNP) markers. Interval mapping for each of 15,888 traits identified 2309 quantitative trait loci (QTLs) 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 l-lod (logarithm of the odds ratio for linkage) QTL support interval and the gene]. Nearly a quarter of the remaining QTLs (distant QTLs) map to three statistically robust hotspots (*11*) (Fig. 1A and 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 mutation of phenylalanine to valine at residue 215 in the neuropeptide receptor *npr-1* (*12*). Candidate genes for the other hotspots include *Y17G9B.8*, a putative component of a chromatin regulatory complex 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 markedly nonuniform. 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 semidiscrete 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, whereas 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

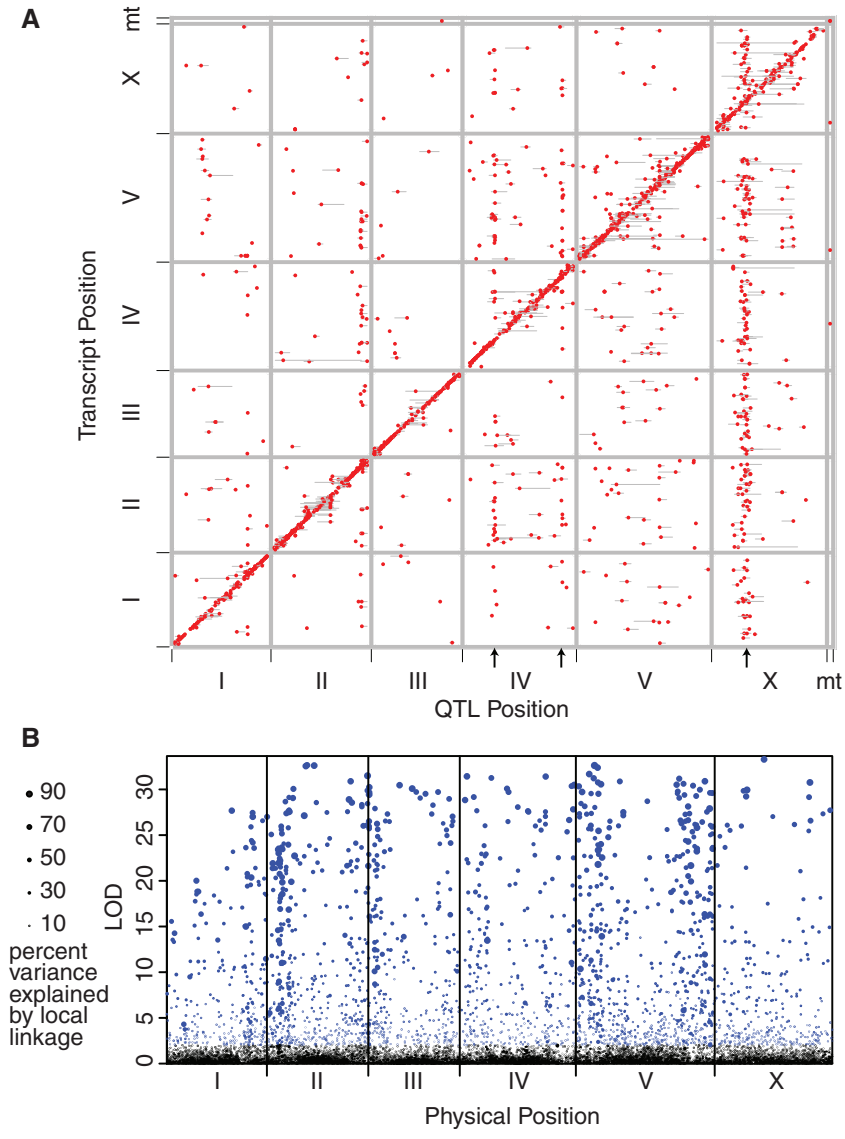

Fig. 1. (**A**) QTLs for each transcript abundance phenotype, significant at an FDR of 5%, are plotted in rows located at the genomic positions of the transcripts. Gray bars represent 1-lod support intervals. The diagonal includes local QTLs, those that colocalize with the transcript they affect. Three 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% FDR according to a single-marker linkage test. Points are scaled according to the fraction of variance in transcript abundance explained by the local QTL.

Table 1. Both distant and local QTLs are overrepresented in chromosome arms relative to centers.

Domain	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

[1]Department of Biology and Center for Genomics and Systems Biology, New York University, 100 Washington Square East, New York, NY 10003, USA. [2]Lewis-Sigler Institute for Integrative Genomics and Department of Ecology and Evolutionary Biology, Carl Icahn Laboratory, Princeton University, Princeton, NJ 08544, USA. [3]Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544, USA.

*To whom correspondence should be addressed. E-mail: mrockman@nyu.edu (M.V.R.); leonid@genomics.princeton.edu (L.K.)

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 2538 transcripts affected by QTLs 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 (χ^2_2 = 495.7, $P < 10^{-107}$). The chromosomal patterning is robust to confounding by potential hybridization artifacts, as demonstrated by analysis of only the 7694 transcripts for which the CB4856 genotype is associated with higher expression than the reference N2 genotype. The 1057 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 (χ^2_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 1191 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 nonexclusive 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

Table 2. Logistic regression models implicate mutation, stabilizing selection, and linked selection in explaining the distribution of local linkages. LRT: likelihood ratio test statistic comparing the logistic regression model in which the specified term has been dropped to the model in which all terms are included. LRT is equivalent to the drop in explained deviance due to excluding the term from the model. The null deviance is 12897.5. The LRT was tested against a χ^2 distribution to yield the associated P values. Δdf: difference in degrees of freedom between the specified model and the null model, including only the intercept. Chromosomal domain is a factor with three levels and hence contributes two degrees of freedom. AIC: Akaike information criterion. Model 4 includes all two-, three-, and four-way interactions among the variables. Consequently the LRT and P values for dropping single terms cannot be calculated.

	Model 1		Model 2		Model 3		Model 4	
	LRT	$-\log_{10}(P)$	LRT	$-\log_{10}(P)$	LRT	$-\log_{10}(P)$	LRT	$-\log_{10}(P)$
Gene size	149	33.5	38.7	9.30	–	–	X	X
RNAi	13.1	3.5	10.8	3.00	–	–	X	X
Conservation	222.6	49.6	81.6	18.77	–	–	X	X
Distant linkage	36.2	8.7	35.1	8.51	–	–	X	X
Interaction terms	–	–	–	–	–	–	X	X
Recombination rate	–	–	–	–	–	–	–	–
Chromosomal domain	–	–	263.1	57.12	444.8	96.59	–	–
Background selection	–	–	–	–	–	–	–	–
Explained deviance	337.8		600.9		444.8		370.6	
Residual deviance	12,559.7		12,296.6		12,452.7		12,526.9	
Δdf	4		6		2		15	
AIC	12,569.7		12,310.6		12,458.7		12,558.9	

	Model 5		Model 6		Model 7		Model 8	
	LRT	$-\log_{10}(P)$	LRT	$-\log_{10}(P)$	LRT	$-\log_{10}(P)$	LRT	$-\log_{10}(P)$
Gene size	121.9	27.62	38.4	9.23	22.6	5.71	22.5	5.68
RNAi	14.4	3.83	11.1	3.06	11.8	3.23	11.9	3.25
Conservation	193.6	38.94	81.6	18.78	44.2	10.54	44	10.49
Distant linkage	35.9	8.68	35.1	8.50	33.4	8.12	33.3	8.11
Interaction terms	–	–	–	–	–	–	–	–
Recombination rate	52.1	12.27	1.2	0.56	–	–	–	–
Chromosomal domain	–	–	212.2	46.07	–	–	0.3	0.06
Background selection	–	–	–	–	363.6	80.32	100.7	22.98
Explained deviance	389.9		602.1		701.4		701.6	
Residual deviance	12,507.6		12,295.4		12,196.1		12,195.9	
Δdf	5		7		5		7	
AIC	12,519.6		12,311.4		12,208.1		12,211.9	

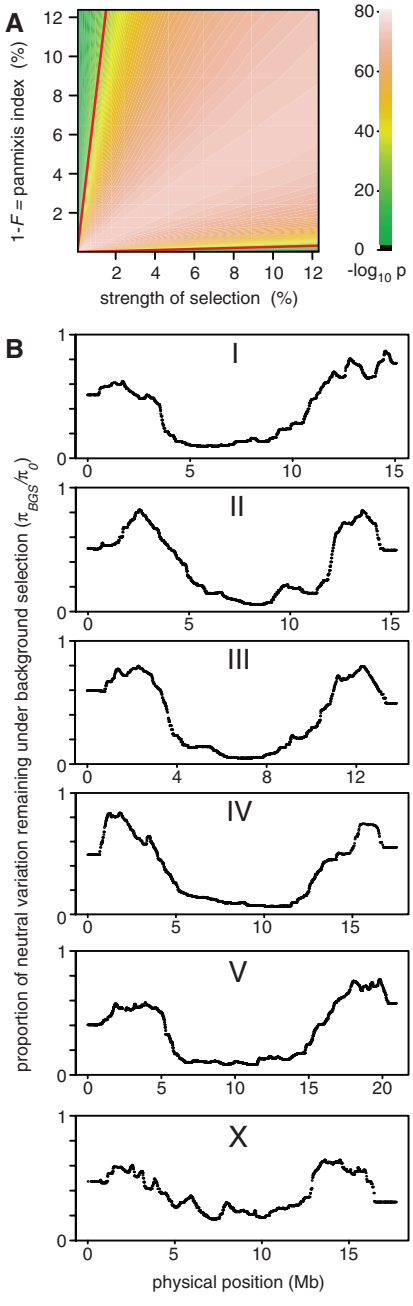

Fig. 2. (**A**) The significance of background selection in a logistic regression model (which includes gene-specific mutation and selection variables) is plotted as a function of the index of panmixis and strength of selection against deleterious mutations. Background selection is significant at $P < 0.01$ 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 fig. S4. (**B**) Effects of background selection on levels of variation along the chromosomes under the best-fitting background selection model.

variation results from mutation, selection, and drift, the last 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, are precisely localized, and 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 RNA interference (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 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. Indeed, 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 N_e of the linked genomic interval (*24*–*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*'s 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 and 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 to 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.

References and Notes

1. D. Houle, *Genetica* **102–103**, 241 (1998).
2. R. B. Brem, L. Kruglyak, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 1572 (2005).
3. S. A. Rifkin, J. Kim, K. P. White, *Nat. Genet.* **33**, 138 (2003).
4. D. R. Denver *et al.*, *Nat. Genet.* **37**, 544 (2005).
5. J. Ronald, J. M. Akey, *PLoS ONE* **2**, e678 (2007).
6. C. R. Landry, B. Lemos, S. A. Rifkin, W. J. Dickinson, D. L. Hartl, *Science* **317**, 118 (2007).
7. J. C. Fay, P. J. Wittkopp, *Heredity* **100**, 191 (2008).
8. Y. Li *et al.*, *PLoS Genet.* **2**, e222 (2006).
9. M. V. Rockman, L. Kruglyak, *PLoS Genet.* **5**, e1000419 (2009).

10. S. R. Wicks, R. T. Yeh, W. R. Gish, R. H. Waterston, R. H. Plasterk, *Nat. Genet.* **28**, 160 (2001).
11. Information on materials and methods is available on *Science* Online.
12. M. de Bono, C. I. Bargmann, *Cell* **94**, 679 (1998).
13. T. M. Barnes, Y. Kohara, A. Coulson, S. Hekimi, *Genetics* **141**, 159 (1995).
14. The C. elegans Sequencing Consortium, *Science* **282**, 2012 (1998).
15. A. Genissel, L. M. McIntyre, M. L. Wayne, S. V. Nuzhdin, *Mol. Biol. Evol.* **25**, 101 (2008).
16. M. A. Noor, A. L. Cunningham, J. C. Larkin, *Genetics* **159**, 581 (2001).
17. J. Ronald, R. B. Brem, J. Whittle, L. Kruglyak, *PLoS Genet.* **1**, e25 (2005).
18. P. M. Visscher *et al.*, *Am. J. Hum. Genet.* **81**, 1104 (2007).
19. M. K. Lawniczak, A. K. Holloway, D. J. Begun, C. D. Jones, *Genome Biol.* **9**, R125 (2008).
20. J. Tung, O. Fédrigo, R. Haygood, S. Mukherjee, G. A. Wray, *Mol. Biol. Evol.* **26**, 2047 (2009).
21. J. F. Ayroles *et al.*, *Nat. Genet.* **41**, 299 (2009).
22. S. J. Arnold, R. Bürger, P. A. Hohenlohe, B. C. Ajie, A. G. Jones, *Evolution* **62**, 2451 (2008).
23. D. R. Denver *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16310 (2009).
24. W. G. Hill, A. Robertson, *Genet. Res.* **8**, 269 (1966).
25. J. M. Smith, J. Haigh, *Genet. Res.* **23**, 23 (1974).
26. B. Charlesworth, M. T. Morgan, D. Charlesworth, *Genetics* **134**, 1289 (1993).
27. A. Graustein, J. M. Gaspar, J. R. Walters, M. F. Palopoli, *Genetics* **161**, 99 (2002).
28. A. D. Cutter, B. A. Payseur, *Mol. Biol. Evol.* **20**, 665 (2003).
29. R. Jovelin, *Mol. Biol. Evol.* **26**, 2373 (2009).
30. A. D. Cutter, J. Y. Choi, *Genome Res.* **20**, 1103 (2010).
31. R. R. Hudson, N. L. Kaplan, *Genetics* **141**, 1605 (1995).
32. V. B. Kaiser, B. Charlesworth, *Trends Genet.* **25**, 9 (2009).
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 (grants R01 HG004321 to L.K., R01 GM089972 to M.V.R., and P50 GM071508 to the Lewis-Sigler Institute), a Jane Coffin Childs Fellowship (M.V.R.), an Ellison Foundation New Scholar Award (M.V.R.), and a James S. McDonnell Foundation Centennial Fellowship (L.K.). L.K. is an investigator of the Howard Hughes Medical Institute. Microarray data have been deposited at the Gene Expression Omnibus with accession number GSE23857.

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6002/372/DC1
Materials and Methods
Figs. S1 to S7
References

24 June 2010; accepted 31 August 2010
10.1126/science.1194208

Rapid Construction of Empirical RNA Fitness Landscapes

Jason N. Pitt and Adrian R. Ferré-D'Amaré*

Evolution is an adaptive walk through a hypothetical fitness landscape, which depicts the relationship between genotypes and the fitness of each corresponding phenotype. We constructed an empirical fitness landscape for a catalytic RNA by combining next-generation sequencing, computational analysis, and "serial depletion," an in vitro selection protocol. By determining the reaction rate constant for every point mutant of a catalytic RNA, we demonstrated that abundance in serially depleted pools correlates with biochemical activity (correlation coefficient $r = 0.67$, standard score $Z = 7.4$). Therefore, enumeration of each genotype by deep sequencing yielded a fitness landscape containing ~10^7 unique sequences, without requiring measurement of the phenotypic fitness for each sequence. High-throughput mapping between genotype and phenotype may apply to artificial selections, host-pathogen interactions, and other biomedically relevant evolutionary phenomena.

In vitro selection of RNA has led to the isolation of aptamers and ribozymes expanding our understanding of the biochemical capabilities of this nucleic acid (*1*). Systematic evolution of ligands by exponential enrichment (SELEX) methodology also allows the study of the process of evolutionary adaptation (*2*), which has been conceptualized as an optimizing walk through a fitness landscape (*3*). In molecular evolution, such a landscape represents the sequence space of a macromolecule and the fitness associated with each genotype (*4*, *5*). Two major difficulties have been identified in the empirical construction of macromolecular fitness landscapes. First, even for macromolecules of modest length, the sequence space is vast; a 20-mer RNA or protein has ~10^{12} or ~10^{26} possible sequences, respectively. Second, to characterize the landscape, the phenotypic fitness of each individual genotype needs to be measured, or an indirect measure of fitness needs to be validated. Therefore, although fitness landscapes have been constructed through computer simulations [see, for example, (*6*–*8*)], experimental analyses of fitness landscapes have typically been limited to dozens to hundreds of genotypes [reviewed in (*9*)].

Advances in DNA sequencing methodology allow us to sequence ~10^8 individual molecules of lengths up to ~50 nucleotides (nt) with high accuracy, producing larger accessible experimental sequence spaces (*10*). During an in vitro selection experiment, the representation of RNA molecules that are more active should increase. Thus, we hypothesize that the abundance (or "fecundity") of a particular sequence in an RNA pool undergoing selection, or its rate of increase, may serve as a surrogate fitness metric. Therefore, deep sequencing of earlier stages of an in vitro RNA selection experiment (before the pool became dominated by the most active species) and enumeration of the frequency of each genotype may directly provide an experimental fitness landscape.

To establish the validity of this approach, we mutagenized a well-characterized in vitro selected RNA ligase ribozyme and subjected the resulting pool to reselection. We chose the class II ligase (*11*), because its length (54 nt) is comparable to those of reliable individual reads by the Illumina genome analyzer and because the wild type [isolate a4-11, hereafter "master sequence;" fig. S1 and (*11*)] of this catalytic RNA is extremely active, likely being near-optimal. The mutagenized pool (45 nt at a degeneracy of 21% per position; expected frequency of master sequence = 0.0025%) can be analogous to a viral population arising because of error-prone replication (*12*). Our class II ligase construct catalyzes the formation of a 2′-5′ phosphodiester bond between its 5′ triphosphate and the 3′ terminus of a substrate RNA that is immobilized on a magnetic bead (fig. S1) (*13*). Additional sequence pools served as negative controls: One was composed of ~10^{13} random RNAs; the other was made up of mutants of a sequence variant of the class II ligase that has been engineered to favor 3′-5′ bond formation, but which is crippled in catalytic activity (fig. S1) (*14*).

Our control pools were flanked by constant sequences allowing their members to be distinguished from those of the master-sequence pool, and neither was expected to be substantially enriched during the course of selection. The three RNA pools were mixed to generate a starting pool with a complexity of ~6×10^{13} sequences (*13*), with the genotypes of members of each pool centered on a characteristic sequence distance (number of mutations) from the master sequence (Fig. 1A).

To separate individual sequences by their reactivity, we incubated the RNA pool with substrate RNA linked to beads, depleting the most active species from the population. RNAs were allowed to react for 24 hours. Then, we sequenced species capable of ligation to the beads, and we analyzed their genotypes and frequencies (Fig. 1 and table S1) (*13*). This one-step selection experiment resulted in an enrichment of the master-sequence pool at the detriment of the random and engineered pools. The most abundant sequences in the starting pool are polymerase chain reaction artifacts from the random pool (Fig. 1B). These disappear in the 24-hour selected pool (Fig. 1D). Moreover, the representation of the master sequence (Hamming distance = 0) increased markedly, from 0.0015 to 0.28% of the total population (Fig. 1, C and E).

Because the enrichment of a sequence in the selection experiment should be proportional to its catalytic activity, we successively depleted an

Howard Hughes Medical Institute and Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109–1024, USA.

*To whom correspondence should be addressed. E-mail: aferre@fhcrc.org