

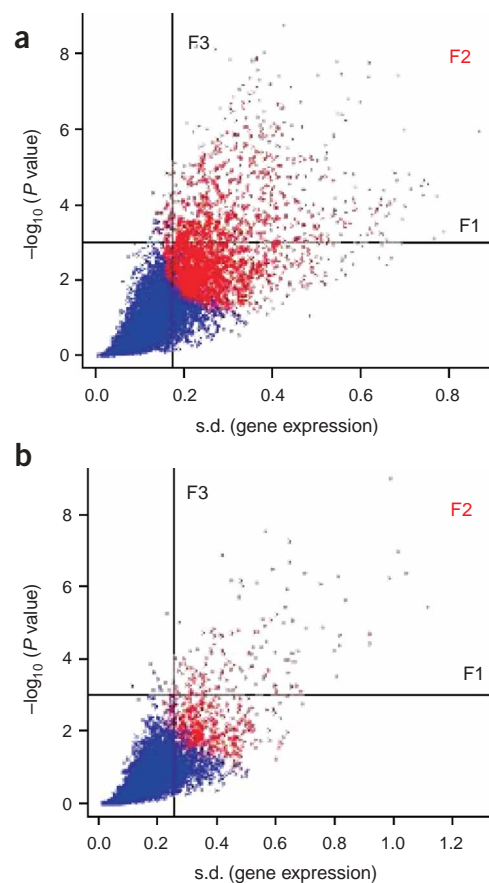
# The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*

Dee R Denver<sup>1,5</sup>, Krystalynne Morris<sup>2</sup>, J Todd Strelman<sup>3</sup>, Stuart K Kim<sup>4</sup>, Michael Lynch<sup>1</sup> & W Kelley Thomas<sup>2</sup>

The evolutionary importance of gene-expression divergence is unclear: some studies suggest that it is an important mechanism for evolution by natural selection<sup>1,2</sup>, whereas others claim that most between-species regulatory changes are neutral or nearly neutral<sup>3</sup>. We examined global transcriptional divergence patterns in a set of *Caenorhabditis elegans* mutation-accumulation lines and natural isolate lines to provide insights into the evolutionary importance of transcriptional variation and to discriminate between the forces of mutation and natural selection in shaping the evolution of gene expression. We detected the effects of selection on transcriptional divergence patterns and characterized them with respect to coexpressed gene sets, chromosomal clustering of expression changes and functional gene categories. We directly compared observed transcriptional variation patterns in the mutation-accumulation and natural isolate lines to a neutral model of transcriptome evolution to show that strong stabilizing selection dominates the evolution of transcriptional change for thousands of *C. elegans* expressed sequences.

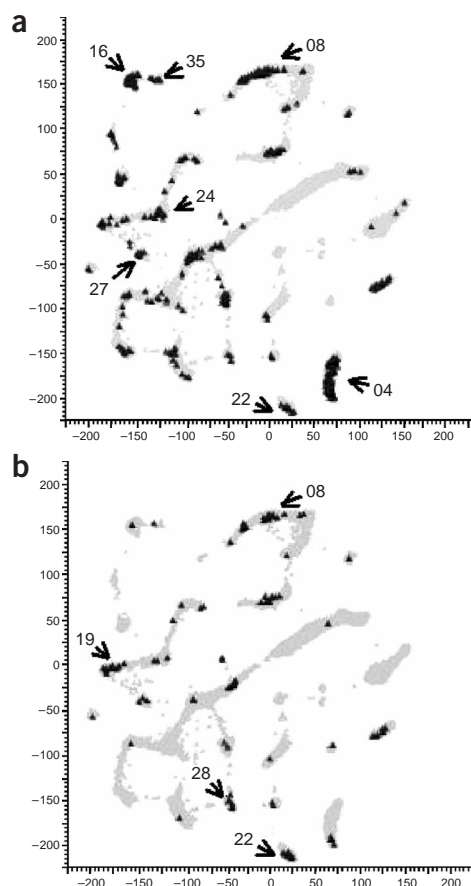
Understanding the relative roles of underlying mutation processes and natural selection in shaping gene-expression divergence has proven difficult, as both of these evolutionary forces affect transcriptional variation in natural populations. To address this problem, we carried out a series of microarray comparisons using a set of *C. elegans* mutation-accumulation (MA) lines and a set of *C. elegans* natural isolate (NI) lines. The MA lines were propagated across 280 generations as single randomly selected hermaphrodites, resulting in an

effective population size ( $N_e$ ) equal to one for each MA line and ensuring that all but the most deleterious mutations accumulated over time in the MA line genomes<sup>4</sup>. For the NI lines, silent-site nucleotide diversity ( $\pi_{si} = 6.3 \times 10^{-4}$ ) and direct base-substitution mutation rate



**Figure 1** Half-volcano plots for MA and NI microarray data. The standard deviations (s.d.) of the adjusted relative expression values for each gene are plotted on the x axes and  $-\log_{10}$ -transformed  $P$  values are plotted on the y axes. The significance threshold cutoff for the F1 test is represented by horizontal lines; the threshold for the F3 test, by vertical lines. Genes considered significant by the hybrid F2 tests are shown in red. (a) MA microarrays. (b) NI microarrays.

<sup>1</sup>Department of Biology, Indiana University, Bloomington, Indiana 47405, USA. <sup>2</sup>Hubbard Center for Genome Studies, University of New Hampshire, Durham, New Hampshire 03824, USA. <sup>3</sup>School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332, USA. <sup>4</sup>Department of Developmental Biology, Stanford University Medical Center, Stanford, California 94305, USA. <sup>5</sup>Present address: Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Auckland, New Zealand. Correspondence should be addressed to D.R.D. (d.denver@massey.ac.nz).



**Figure 2** Distributions of differentially expressed genes in coexpression mounts. Two-dimensional terrain maps of the *C. elegans* gene expression map<sup>14</sup> are shown in gray. *x*-*y* coordinates (arbitrary units) were assigned to group each gene with those with similar expression and to separate it from the other gene clusters. The distributions of differentially expressed genes in the MA (a) and NI (b) lines are represented by black triangles. Numbered arrows indicate mounts where genes are significantly over-represented. Mount 4 is enriched for sperm genes; mount 8, for intestine genes; mount 16, for muscle and collagen genes; mount 19, for amino acid and lipid metabolism genes and cytochrome P450 genes; mount 22, for collagen genes; mount 24, for amino acid and lipid metabolism genes; mount 27, for amino acid metabolism genes; mount 28, for genes of unknown function; and mount 35, for collagen genes.

divergence among the MA lines, a much greater fraction of expressed sequences showed significant transcriptional divergence in the MA lines, where the effects of selection are reduced to a bare minimum. This disparity suggests that stabilizing natural selection has a key role in shaping evolution of the *C. elegans* transcriptome.

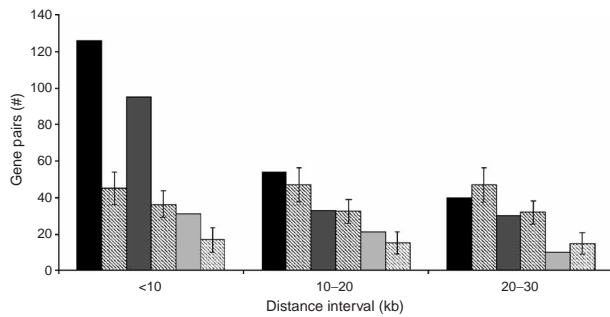
Explaining the molecular sources of gene expression divergence is complicated because regulatory changes can occur through different mechanisms with varying effects. For instance, a point mutation in a promoter region may simply alter the transcript levels of the gene(s) immediately downstream (a *cis*-acting mutation), whereas a structural change in a transcription factor may result in multiple downstream regulatory changes (a *trans*-acting mutation)<sup>12,13</sup>. This is an oversimplification, as many mutations will result in complex indirect transcriptional responses. To obtain a rough estimate of the fraction of differentially expressed sequences from the MA and NI data that may result from *trans*-acting mutations with multiple downstream effects, we examined the distribution patterns of differentially expressed sequences for each of the MA and NI experiments with respect to a *C. elegans* gene coexpression map, composed of multiple mounts of coregulated gene sets<sup>14</sup>. For the MA lines, seven coexpression mounts contained a significant over-representation ( $P < 0.01$ , exact hypergeometric probability test) of differentially expressed MA line genes (Fig. 2a). Most of the genes differentially expressed in the MA lines (447 of 660) were placed into these seven mounts; roughly half of them (266 of 447) were in one mount (mount 4, enriched for sperm genes). For the NI lines, in contrast, only four coexpression mounts had a significant over-representation ( $P < 0.01$ , exact hypergeometric probability test) of differentially expressed NI line genes, and few (37 of 118) of these sequences were placed into these four mounts (Fig. 2b). The over-representation of differentially expressed MA line genes in the coexpression mounts suggests that most of the transcriptional differences observed in the MA lines may be due primarily to a few *trans*-acting mutations with multiple downstream effects. This idea is supported by the observed consistency of MA line-specific relative expression patterns for those sequences over-represented in specific mounts, particularly mount 4 (Supplementary Fig. 2 online). In the NI lines, on the other hand, the relatively small number of differentially expressed genes over-represented in the coexpression mounts suggests that, in natural populations of *C. elegans*, natural selection eliminates most *trans*-acting mutations with multiple downstream effects.

*C. elegans* gene-rich autosomal core regions are poor in repetitive elements and have low SNP densities, whereas gene-poor autosomal arm regions have relatively high densities of repetitive DNA and SNPs<sup>15,16</sup>. Differentially expressed MA line genes did not have a significant distributional bias toward core or arm regions (286 observed in cores, 301 observed in arms; 307 and 280 expected, respectively;  $0.1 < P < 0.5$ ), suggesting that gene regulation is equally

estimates ( $\mu_{bs} = 9.0 \times 10^{-9}$  base substitutions per site per generation) for *C. elegans* indicate an  $N_e$  of  $\sim 17,500$  (refs. 5–7), assuming that  $\pi_{si}$  is equal to its expected equilibrium value<sup>8</sup> and that the base-substitution mutation rate in the MA lines is equal to the evolutionary rate of silent-site substitution in natural *C. elegans* populations. Hence, selection is expected to have been much more efficient in the evolutionary history of the NI lines than in that of the MA lines (where  $N_e = 1$ ).

Using a loop design for our microarray comparisons (Supplementary Fig. 1 online)<sup>9</sup>, we analyzed transcriptional variation among four randomly selected MA lines (MA24, MA41, MA83 and MA99) and in their N2 (common laboratory strain) ancestor. We also analyzed five NI lines (AB1, CB4856, N2, PB303 and PB306), chosen to represent a maximal diversity of genotypes as determined by intraspecific *C. elegans* phylogenetic analyses<sup>10</sup>. We hybridized dye-labeled cDNA samples to microarray slides containing representative spots for virtually all coding sequences in the genome, along with multiple control spots. Using R/Microarray Analysis of Variance (R/MAANOVA) software<sup>11</sup>, we carried out three F tests with each of the adjusted sets of MA and NI microarray data to identify genes that had significant differential expression patterns.

Among the 7,014 unique expressed sequences considered in the MA lines, 660 ( $\sim 9\%$ ; Supplementary Table 1 online) had significant ( $P < 0.001$ ) differential expression patterns in all three F tests (Fig. 1a). By contrast, only 118 of 5,588 unique expressed sequences in the NI lines ( $\sim 2\%$ ; Supplementary Table 2 online) were differentially expressed under the same significance parameters (Fig. 1b). Despite the fact that the NI lines have experienced many thousands of generations of divergence ( $\sim 1\%$  mitochondrial DNA divergence among NI lines<sup>10</sup>), compared with the 280 known generations of



**Figure 3** Physical distances between differentially expressed genes in the MA lines. Distance intervals (in kb) are plotted on the x axis, and numbers of gene pairs are plotted on the y axis. Solid bars indicate observed values; striped bars indicate expected values. Black bars represent the total set of differentially expressed sequences in the MA lines; dark gray bars indicate differentially expressed gene sets over-represented in coexpression mounts; and light gray bars represent differentially expressed genes not over-represented in mounts. Distributions of genes in distance intervals >30 kb were not significantly different than random expectations. Error bars for expected values indicate 95% confidence intervals.

mutable for arm and core genes. In contrast, differentially expressed NI line genes were significantly ( $P < 0.005$ ) biased toward autosomal arm regions over cores (34 observed in cores, 66 observed in arms; 56 and 44 expected, respectively). This difference could be explained by either stronger or more efficient selection against transcriptional change for core genes relative to arm genes. Average recombination rates are much lower in autosomal core regions ( $0.0007 \text{ cM kb}^{-1}$ ) than in arm regions ( $0.0047 \text{ cM kb}^{-1}$ )<sup>15</sup>, indicating that more efficient selection associated with Hill-Robertson effects (hitchhiking and background selection)<sup>17</sup> is not responsible for the low relative incidence of core gene transcriptional divergence in the NI lines. In autosomal core regions (as compared with arms), there is a higher incidence of genes involved in fundamental metabolic or information-management processes that have yeast and human orthologs<sup>18</sup>; consequently, core genes may, on average, be subject to stronger selection against transcriptional change than are arm genes.

To analyze patterns of chromosomal clustering more specifically, we considered the physical distances between differentially expressed sequences with respect to physical chromosome positions. In the MA lines, we found a strong bias toward gene pairs spaced less than 10 kb apart (**Fig. 3**); the number of gene pairs separated by <10 kb in the NI lines (four observed pairs) was much closer to the expected range of one to three pairs. We then considered the physical clustering patterns of those differentially expressed MA line genes that were over-represented in the coexpression mounts versus those that were not, as certain classes of coexpressed genes show chromosomal clustering in *C. elegans*<sup>19,20</sup>. The clustering bias was much stronger for those gene pairs over-represented in mounts (**Fig. 3**). Nonetheless, the number of differentially expressed MA line gene pairs spaced by <10 kb that cannot be explained by their over-representation in coexpression mounts was roughly twice that expected. These instances may reflect local mutations with large *cis*-acting transcriptional effects.

Our approach provides an opportunity to compare observed gene-expression divergence patterns in *C. elegans* with neutral expectations directly on a gene-by-gene basis by considering the gene-specific ratios of transcriptional genetic variance ( $V_g$ ) observed in the NI data to the transcriptional mutational variance ( $V_m$ ) observed in the MA data. Under neutrality, the  $V_g/V_m$  ratio is expected to be equal to  $4N_e$  in primarily self-replicating diploid organisms such as *C. elegans*<sup>10,21</sup>. Purifying selection against transcriptional change will cause the ratio to become increasingly smaller than the neutral expectation. We compared the  $V_g/V_m$  ratios for the 3,696 expressed sequences considered in both the MA and NI loops to the ratio expected under neutrality (70,000, assuming  $N_e$  for *C. elegans* is 17,500) and found that all of the observed gene-specific  $V_g/V_m$  ratios were far below the neutral expectation (**Supplementary Fig. 3** online); the largest observed value was 9,436. Even if we upwardly biased the estimate of  $4N_e$  for *C. elegans* by a full order of magnitude, all but two of the observed gene-specific ratios would still be below the neutral expectation. This suggests that the intraspecific evolution of transcriptional variation for most *C. elegans* genes is dominated by intense stabilizing selection. A key role for stabilizing selection in primate transcriptome evolution has also been reported in mixed-model reanalyses<sup>22</sup> of oligonucleotide array-based expression data<sup>23</sup>.

**Table 1**  $V_g/V_m$  intervals and functional gene categories

Functional category	$V_g/V_m$ interval				Significance
	<100	100–200	200–300	>300	
DEG ( $n = 134$ )	39 (36.9)	29 (30.9)	15 (21.3)	51 (44.9)	$0.1 < P < 0.5$
DNA ( $n = 85$ )	34 (23.4)	18 (19.6)	11 (13.5)	22 (28.5)	$0.05 < P < 0.1$
MET ( $n = 355$ )	90 (97.8)	63 (81.9)	45 (56.4)	157 (118.9)	$P < 0.005$
PRO ( $n = 55$ )	11 (15.1)	10 (12.7)	11 (8.7)	23 (18.4)	$0.1 < P < 0.5$
PSE ( $n = 2$ )	1 (0.6)	1 (0.5)	0 (0.3)	0 (0.7)	$0.5 < P < 0.9$
RNP ( $n = 70$ )	25 (19.3)	12 (16.2)	11 (11.1)	22 (23.4)	$0.1 < P < 0.5$
RTP ( $n = 18$ )	8 (5.0)	6 (4.2)	1 (2.9)	3 (6.0)	$0.1 < P < 0.5$
SGN ( $n = 280$ )	110 (77.1)	69 (64.6)	41 (44.5)	60 (93.8)	$P < 0.005$
STR ( $n = 173$ )	36 (47.6)	48 (39.9)	39 (27.5)	50 (57.9)	$0.01 < P < 0.05$
TRC ( $n = 144$ )	46 (39.7)	46 (33.2)	20 (22.9)	32 (48.2)	$0.005 < P < 0.01$
TRL ( $n = 98$ )	12 (27.0)	19 (22.6)	20 (15.6)	47 (32.8)	$0.005 < P < 0.01$
TSP ( $n = 170$ )	52 (41.9)	37 (35.1)	21 (27.0)	60 (56.9)	$0.1 < P < 0.5$
UNK ( $n = 2,112$ )	554 (581.8)	490 (487.4)	352 (335.4)	716 (707.4)	$0.5 < P < 0.9$

DEG, protein degradation; DNA, DNA metabolism; MET, amino acid, carbohydrate and lipid metabolism; PRO, protein processing; PSE, pseudogene; RNP, RNA processing; RTP, (retro)transposon; SGN, signaling; STR, structural; TRC, transcription; TRL, translation; TSP, transport; UNK, unknown. The 3,696 genes for which  $V_g/V_m$  ratios were calculated were assigned into functional categories (**Supplementary Note** online); their distribution patterns with respect to four  $V_g/V_m$  intervals are shown. The observed values are given, followed by the values expected on the basis of a random distribution of the total number of genes in a given functional category among the four  $V_g/V_m$  intervals (in parentheses). Significance values were determined by  $\chi^2$  test.

We next characterized those sequences for which  $V_g/V_m$  ratios were estimated in terms of functional gene categories (Table 1). We defined 13 broad functional categories (Supplementary Note online) and placed genes into these categories using Gene Ontology terms<sup>24</sup> and genome annotations from WormBase<sup>25</sup>. We divided the observed gene-specific  $V_g/V_m$  ratios into four intervals ( $<100$  (strongest selection), 100–200, 200–300 and  $>300$  (weakest selection)) and analyzed the distributions of genes from the 13 functional categories with respect to these four intervals. For 5 of the 13 functional categories, the distributions of genes among the four  $V_g/V_m$  intervals deviated significantly ( $P < 0.05$ ,  $\chi^2$  test) from null expectations. Genes involved in signal transduction pathways were significantly over-represented in the  $<100$   $V_g/V_m$  interval and under-represented in the  $>300$  interval, suggesting that transcriptional stability is under more intense stabilizing selection for this particular set of genes than for most other genes considered. In contrast, genes involved in carbohydrate, amino acid and lipid metabolism pathways were significantly over-represented in the  $>300$   $V_g/V_m$  interval and under-represented in the other three intervals, indicating that regulation is under less stringent stabilizing selection in these genes than in most other gene classes considered. Distribution patterns of genes encoding structural proteins as well as those involved in transcription and translation also deviated significantly from expectations (Table 1).

We show that strong stabilizing selection dominates intraspecific transcriptome evolution in *C. elegans* by comparing global transcriptional divergence patterns in the MA lines to that in the NI lines, and by directly testing thousands of observed gene-specific  $V_g/V_m$  ratios for transcriptional divergence against neutral expectations. Further studies are required to determine the role of natural selection in shaping between-species gene-expression divergence patterns and the extent to which stabilizing selection dominates intraspecific transcriptome evolution in other species.

## METHODS

***C. elegans* culturing, mRNA purification and microarrays.** We grew developmentally synchronized *C. elegans* cultures on RNGM agarose plates in accordance with standard protocols<sup>26</sup>. We purified mRNAs from synchronous worm populations at the young adult stage using the MESSAGING mRNA purification kit (Invitrogen). We synthesized Cy3- and Cy5-labeled dUTP cDNA and carried out microarray hybridizations and scanning as previously described<sup>27</sup>. Each microarray contained  $>21,000$  spots with representative sequences from annotated *C. elegans* genes, predicted genes and various controls (intergenic regions, vector sequences and duplicates of gene-specific sequences). Each of the 20 comparisons in each of the MA line and NI line loops constitute distinct biological replicates (we grew independent worm cultures for each comparison). We carried out four replicates for each pairwise comparison in each of the loops (Supplementary Fig. 1 online), with full dye swapping (sample A was labeled with Cy5-dUTP and sample B was labeled with Cy3-dUTP in two arrays, whereas sample A was labeled with Cy3-dUTP and sample B was labeled with Cy5-dUTP in the other two arrays).

**Microarray data analysis.** We downloaded microarray data files from the Stanford Microarray Database and subjected them to preliminary data filtering procedures. Data corresponding to spots that were flagged or that had raw intensity signal to noise ratios  $<1.5$  were omitted from subsequent analyses. We consolidated spot-specific raw intensity values for each of the 40 samples analyzed in each of the MA and NI loops into a single spreadsheet and removed all spots that had missing data from any of the 20 microarrays (required for downstream analyses). We then analyzed these filtered raw intensity data sets (7,014 unique expressed sequences for the MA loop; 5,588 sequences for the NI loop) using the R/MAANOVA microarray analysis program<sup>11</sup>. We smoothed the data with the joint loess method and then fit the data to a mixed-effects ANOVA model in two stages (normalization modeling followed by gene-specific effects modeling) with the array effects term modeled as random. We

then carried out hypothesis testing using F permutation tests (1,000 permutations for each analysis) with a null hypothesis of no differential expression. We carried out three F tests: the F1 test computes gene-specific error variances; the F3 test assumes common error variances for all genes; and the F2 test is a hybrid that uses a weighted combination of common and gene-specific variance estimates in the denominator. All these procedures were carried out independently for the MA loop and NI loop data sets. Genes whose expression differed significantly at  $P < 0.001$  in all three F tests we designated as differentially expressed. More explicit details regarding ANOVA modeling and the F tests, including specific equations, are provided in ref. 11.

**Chromosomal distributions of differentially expressed genes.** We obtained 5' and 3' gene boundaries (with respect to chromosome position) for all *C. elegans* genes from WormBase<sup>25</sup>. To determine the number of nucleotides intervening between pairs of differentially expressed genes in the MA and NI data, we grouped the genes according to chromosome and chromosome position and then subtracted the beginning (5') position of a given gene from the ending position (3') of the preceding gene. For a null model, we gathered random sets of genes from the complete set of genes considered in the ANOVAs (7,014 for the MA loop; 5,588 for the NI loop) and determined physical distances between the random genes set (we used the same number of random genes as we did true differentially expressed genes). For each loop, we repeated this process 1,000 times to generate null expectations (Fig. 3). We designated genes as being in autosomal arm or core regions in accordance with the boundaries previously defined<sup>15</sup>. We determined null expectations for arm and core distributions using the distribution patterns of the total analyzed set of 7,014 and 5,588 genes for the MA and NI data, respectively.

**Genetic and mutational variance.** We calculated the gene-specific  $V_g$  estimates for transcription observed in the NI lines directly from the adjusted relative expression values resulting from the analyses in R/MAANOVA. We calculated the gene-specific  $V_m$  estimates for transcription observed in the MA lines by dividing the gene-specific among-line variances by the product of the number of bottleneck generations experienced by each MA line (280) and 1.6 (to account for the total number of MA line comparisons; the progenitor N2 was included in the MA loop). We calculated gene-specific  $V_g/V_m$  ratios for each of the 3,696 genes that were present in both the MA and NI data sets. We compared the observed gene-specific  $V_g/V_m$  ratios to the neutral expectation value (under neutrality, this ratio is expected to equal  $4N_e$  in self-replicating organisms<sup>21</sup>), calculated using the  $N_e$  estimate for *C. elegans* from silent-site nucleotide variation<sup>6</sup> and the base-substitution mutation rate<sup>7</sup>. We defined functional gene categories using *C. elegans* Gene Ontology terms<sup>24</sup> and WormBase annotations<sup>25</sup>. We determined random expectations for gene distributions among the above categories in a fashion similar to the null expectations for chromosomal distribution patterns described above: we collected random gene sets of sizes equal to observed sets and recorded their distribution patterns among the four  $V_g/V_m$  intervals (1,000 permutations per analysis; Table 1).

Note: Supplementary information is available on the Nature Genetics website.

## ACKNOWLEDGMENTS

We thank K. Duke and M. Jiang for cDNA syntheses and microarray hybridizations and scanning, H. Wu for help with R/MAANOVA, T.D. Kocher for insightful comments and the *Caenorhabditis* Genetics Center for providing the *C. elegans* natural isolates. This work was supported by a University of Missouri Research Board grant (to W.K.T.) and US National Institutes of Health grant (to M.L. and W.K.T.). D.R.D. was supported by a US National Institutes of Health National Research Service Award fellowship, and J.T.S. was supported by a postdoctoral fellowship from the Alfred P. Sloan Foundation.

## COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 24 November 2004; accepted 28 March 2005  
Published online at <http://www.nature.com/naturegenetics/>

- King, M.-C. & Wilson, A.C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Oleksiak, M.F., Churchill, G.A. & Crawford, D.L. Variation in gene expression within and among natural populations. *Nat. Genet.* **32**, 261–266 (2002).

3. Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol.* **2**, 682–689 (2004).
4. Vassilieva, L.L., Hook, A.M. & Lynch, M. The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* **54**, 1234–1246 (2000).
5. Lynch, M. & Conery, J.S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
6. Graustein, A., Gaspar, J.M., Walters, J.R. & Palopoli, M.F. Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. *Genetics* **161**, 99–107 (2002).
7. Denver, D.R., Morris, K., Lynch, M. & Thomas, W.K. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430**, 679–682 (2004).
8. Charlesworth, B. Effective population size. *Curr. Biol.* **12**, R716–R717 (2002).
9. Kerr, M.K. & Churchill, G.A. Experimental design for gene expression microarrays. *Biostatistics* **2**, 183–201 (2001).
10. Denver, D.R., Morris, K. & Thomas, W.K. Phylogenetics in *Caenorhabditis elegans*: an analysis of divergence and outcrossing. *Mol. Biol. Evol.* **20**, 393–400 (2003).
11. Cui, X. & Churchill, G.A. Statistical tests for differential expression in cDNA microarray experiments. *Genome Biol.* **4**, 210 (2003).
12. Wittkopp, P.J., Haerum, B.K. & Clark, A.G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).
13. Wray, G.A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419 (2003).
14. Kim, S.K. *et al.* A gene expression map for *Caenorhabditis elegans*. *Science* **293**, 2087–2092 (2001).
15. Barnes, T.M., Kohara, Y., Coulson, A. & Hekimi, S. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* **141**, 159–179 (1995).
16. Koch, R., van Luenen, H.G., van der Horst, M., Thijssen, K.L. & Plasterk, R.H. Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. *Genome Res.* **10**, 1690–1696 (2000).
17. Hill, W.G. & Robertson, A. The effects of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
18. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
19. Roy, P.J., Stuart, J.M., Lund, J. & Kim, S.K. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature* **418**, 975–979 (2002).
20. Lercher, M.L., Blumenthal, T. & Hurst, L.D. Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. *Genome Res.* **13**, 238–243 (2003).
21. Lynch, M. & Hill, W.G. Phenotypic evolution by neutral mutation. *Evolution* **40**, 915–935 (1986).
22. Hsieh, W., Chu, T., Wolfinger, R.D. & Gibson, G. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* **165**, 747–757 (2003).
23. Enard, W. *et al.* Intra- and interspecific variation in primate gene expression patterns. *Science* **296**, 340–343 (2002).
24. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
25. Harris, T.W. *et al.* WormBase: a multi-species resource for nematode biology and genomics. *Nucleic Acids Res.* **32**, D411–D417 (2004).
26. Fabian, T.J. & Johnson, T.E. Production of age-synchronous mass cultures of *Caenorhabditis elegans*. *J. Gerontol.* **49**, B145–B156 (1994).
27. Jiang, M. *et al.* Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **98**, 218–223 (2001).