View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Repositori d'Objectes Digitals per a l'Ensenyament la Recerca i la Cultura

A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes

Dee R. Denver[a,1], Peter C. Dolan[a], Larry J. Wilhelm[a], Way Sung[b], J. Ignacio Lucas-Lledó[c], Dana K. Howe[a], Samantha C. Lewis[a], Kazu Okamoto[b], W. Kelley Thomas[b], Michael Lynch[c], and Charles F. Baer[d]

[a]Department of Zoology and Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331; [b]Hubbard Center for Genome Studies, University of New Hampshire, Durham, NH 03824; [c]Department of Biology, Indiana University, Bloomington, IN 47405; and [d]Department of Zoology, University of Florida, Gainesville, FL 32611

Edited by M. T. Clegg, University of California, Irvine, CA, and approved August 11, 2009 (received for review May 4, 2009)

Knowledge of mutation processes is central to understanding virtually all evolutionary phenomena and the underlying nature of genetic disorders and cancers. However, the limitations of standard molecular mutation detection methods have historically precluded a genome-wide understanding of mutation rates and spectra in the nuclear genomes of multicellular organisms. We applied two high-throughput DNA sequencing technologies to identify and characterize hundreds of spontaneously arising base-substitution mutations in 10 *Caenorhabditis elegans* mutation-accumulation (MA)-line nuclear genomes. *C. elegans* mutation rate estimates were similar to previous calculations based on smaller numbers of mutations. Mutations were distributed uniformly within and among chromosomes and were not associated with recombination rate variation in the MA lines, suggesting that intragenomic variation in genetic hitchhiking and/or background selection are primarily responsible for the chromosomal distribution patterns of polymorphic nucleotides in *C. elegans* natural populations. A strong mutational bias from G/C to A/T nucleotides was detected in the MA lines, implicating oxidative DNA damage as a major endogenous mutagenic force in *C. elegans*. The observed mutational bias also suggests that the *C. elegans* nuclear genome cannot be at equilibrium because of mutation alone. Transversions dominate the spectrum of spontaneous mutations observed here, whereas transitions dominate patterns of allegedly neutral polymorphism in natural populations of *C. elegans* and many other animal species; this observation challenges the assumption that natural patterns of molecular variation in noncoding regions of the nuclear genome accurately reflect underlying mutation processes.

high-throughput DNA sequencing | mutation accumulation

Mutation is the fuel for evolution and the underlying cause of virtually all genetic diseases and cancers. Accurate knowledge of the rate and spectrum of base-substitution mutation is essential to studying and understanding a variety of evolutionary phenomena, including rates of molecular evolution (1), estimating the effective population size from standing levels of neutral genetic variation (2), and evaluating assumptions underlying common tests of selection on DNA sequence (1, 3). Despite the important roles of base-substitution mutations in evolutionary studies and their impact on human health, direct knowledge on genome-wide base-substitution processes remains scarce. Because mutations occur extremely infrequently, the genomic rate and molecular spectrum of mutation have historically been indirectly inferred from either between-species divergence or standing genetic variation at loci thought to be evolving neutrally, or by extrapolation from estimates at a small handful of loci (4, 5). The former approach relies on the assumption of selective neutrality and might produce misleading results if the putatively neutral loci examined are in fact subject to selection or if the estimated times of divergence are inaccurate. The latter approach relies on extending results from one or a few loci to the entire genome and might produce misleading results if there is extensive within-genome mutational heterogeneity.

The mutation-accumulation (MA)-line paradigm provides a system for the study of mutation in which the confounding effects of natural selection are reduced to the maximum extent possible. In MA experiments, organisms are bottlenecked to one (for self-fertile species) or two (for obligately outcrossing species) individuals for many generations to ensure that all but the most deleterious mutations accumulate in an essentially neutral fashion. Molecular mutation screens in *Caenorhabditis elegans* and *Drosophila melanogaster* MA lines have enabled uniquely direct estimates of the base-substitution mutation rate (μ_{bs}; $\approx 10^{-9}$ to 10^{-8} mutations per site per generation) for these species (6, 7). However, because of the low density of mutations in the MA lines and the inherent limitations of conventional molecular mutation detection methods, these estimates relied on small total numbers of mutations (30 for *C. elegans* and 37 for *D. melanogaster*), too few for an effective assay of the spectrum of mutational events around the genome. An analysis of *Saccharomyces cerevisiae* MA lines applied the power of new high-throughput 454 pyrosequencing technology but was still only able to identify 33 nuclear base substitutions because of the small size of the yeast genome, although it did reveal spectra consistent with previous observations using reporter genes (8). A recent analysis of three *D. melanogaster* MA lines using Illumina high-throughput sequencing technology identified 174 nuclear DNA point mutations and characterized genome-wide mutation rates and spectra in this species (9).

Here, we provide a genome-wide analysis of the spontaneous base-substitution mutation process in the *C. elegans* genome that takes advantage of two independent sets of long-term MA lines and two high-throughput DNA sequencing technologies. Hundreds of mutations were detected, enabling a new scale of mutational analysis in the *C. elegans* nuclear genomes.

Results and Discussion

Experimental Overview. We applied both high-throughput sequencing-by-synthesis (Solexa platform; Illumina) and pyrosequencing (454 platform; Roche) technologies to identify and characterize genome-wide base-substitution mutation processes in 10 *C. elegans* MA-line genomes derived from the N2 lab strain. Two different sets of N2 MA lines were used in the analysis. Three genomes from the original set of N2 *C. elegans* MA lines previously used for molecular mutation surveys (6) were analyzed by using 454 technology. Seven additional genomes from a second set of N2 MA lines (10) were surveyed for base substitutions by using Solexa technology. The two independent

Author contributions: D.R.D., W.K.T., M.L., and C.F.B. designed research; D.K.H., S.C.L., and K.O. performed research; D.R.D., P.C.D., L.J.W., W.S., J.I.L.-L., and D.K.H. analyzed data; and D.R.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The DNA sequence reported in this paper has been deposited in the National Center for Biotechnology Information Short-Read Archive (SRA) (accession no. SRA009375).

[1]To whom correspondence should be addressed. E-mail: denver@cgrb.oregonstate.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0904895106/DCSupplemental.

Table 1. Base-substitution patterns in *C. elegans* MA-line genomes

	L41	L83	L99	B523	B526	B529	B538	B545	B553	B574	MA total
Technology	454	454	454	Solexa	Solexa	Solexa	Solexa	Solexa	Solexa	Solexa	454, Solexa
No. of sites surveyed ($\times 10^6$)	12.5	31.6	28.6	64.5	60.5	82.4	83.5	56.2	77.5	86.5	584.0
No. of base substitutions	24	33	18	51	37	38	50	46	56	38	391
Ts/Tv	0.71	0.65	0.64	0.19	0.54	0.41	0.79	0.31	0.44	0.31	0.45
No. of exon synonymous substitutions	2	3	2	1	2	5	2	2	2	3	24
No. of exon nonsynonymous substitutions	5	6	0	7	6	6	9	5	8	4	56
No. of intron substitutions	6	3	4	16	11	11	8	12	20	10	101
No. of UTR substitutions	4	4	5	6	2	5	8	4	4	3	45
No. of intergenic substitutions	7	17	7	21	16	11	23	23	22	18	165

Numbers preceded by "L" (Lynch) and "B" (Baer) indicate specific MA-line genomes; MA total column indicates total values for all 10 MA lines. Technology row indicates the high-throughput DNA sequencing method employed. No. of sites surveyed row shows the total number of base pairs surveyed that met our criteria for consideration of a possible mutation site (e.g., ≥2× and at least one F and one R read for 454; ≥3× and at least one F and one R read for Solexa—see text and *SI Text* for details). No. of base substitutions row shows the total number of base-substitution mutations detected. Ts/Tv indicates the observed ratio of transitions to transversions. No. of exon synonymous substitutions row shows the number of observed synonymous substitutions in exon sequence. No. of exon nonsynonymous substitutions row shows the number of nonsynonymous substitutions (includes one nonsense mutation in each of L83 and B538).

sets of MA lines experienced the same standard nematode MA protocols (11) and were derived from the same N2 progenitor stock.

For genomes analyzed by Solexa technology, data management and quality control were carried out by using TileQC (12), and sequence read mapping to the reference genome was performed by using the program ELAND, which accompanies the Illumina Solexa platform. ELAND alignments of the short (36-bp) Solexa reads to unique regions (U) of the *C. elegans* genome were used. We imposed conservative thresholds of ≥3X majority consensus coverage of the mutation site, quality scores ≥25, and the requirement that all mutations were identified by at least one read from each strand of DNA. The MA-line N2 progenitor genome was also sequenced with Solexa to distinguish MA-line-specific mutations from potential sequence differences between the specific N2 progenitor used for MA experiments and the published N2 genome sequence. For genomes analyzed by using 454 technology, MA-line-specific reads were assembled against the *C. elegans* reference genome by using AMOS (13); ≥2X majority consensus coverage of mutant bases was required along with identification by at least one read from each strand of DNA. Randomly selected subsets of mutations identified by using these approaches were subsequently analyzed by using conventional PCR and DNA-sequencing approaches: 53 of 54 mutations were confirmed from the Solexa set, and 27 of 29 were confirmed from the 454 set (see *Materials and Methods*), suggesting very low false-positive rates. By using this approach, we identified 391 base-substitution mutations in the 10 *C. elegans* MA-line nuclear genomes (Table 1 and Tables S1 and S2).

Mutation Rates. We first calculated μ_{bs} for each of the 10 MA lines by simply dividing the number of base substitutions identified by using the previously described criteria by the product of the number of nucleotide sites considered and the estimated number of generations experienced by the MA lines (Fig. 1*A*). An average μ_{bs} of 2.7 ($\pm$0.4) $\times$ 10^{-9} per site per generation was estimated for the 10 MA-line genomes by using this approach, and MA-line-specific mutation rate estimates varied only ≈3-fold among genomes with overlapping confidence intervals. This rate is also highly similar to the recently reported *D. melanogaster* rate, 3.5 $\times$ 10^{-9}, based on Solexa analysis of three MA lines (9). The average mutation rate observed in the three *C. elegans* genomes analyzed by 454, μ_{bs} = 3.1 ($\pm$1.9) $\times$ 10^{-9}, was highly similar to that observed in the seven genomes analyzed by Solexa: μ_{bs} = 2.5 ($\pm$0.5) $\times$ 10^{-9}. We also applied a maximum-likelihood method for estimating the mutation rates (14); individual calculations were performed for the each of the 454 and Solexa datasets (see *Materials and Methods*). The maximum-likelihood approach revealed mutation rate estimates that were similar to the previous estimates and were not significantly different between the two datasets: μ_{bs} = 2.0 ($\pm$1.1) $\times$ 10^{-9} for the 454 data, and μ_{bs} = 2.2 ($\pm$0.3) $\times$ 10^{-9} for the Solexa data (Fig. 1*A*). Our current μ_{bs} estimates are slightly (3- to 4-fold) lower than a previous estimate from *C. elegans* MA lines, 9.1 ($\pm$3.0) $\times$ 10^{-9} (same set analyzed here by 454), that used conventional molecular approaches (6). Although we cannot definitively identify

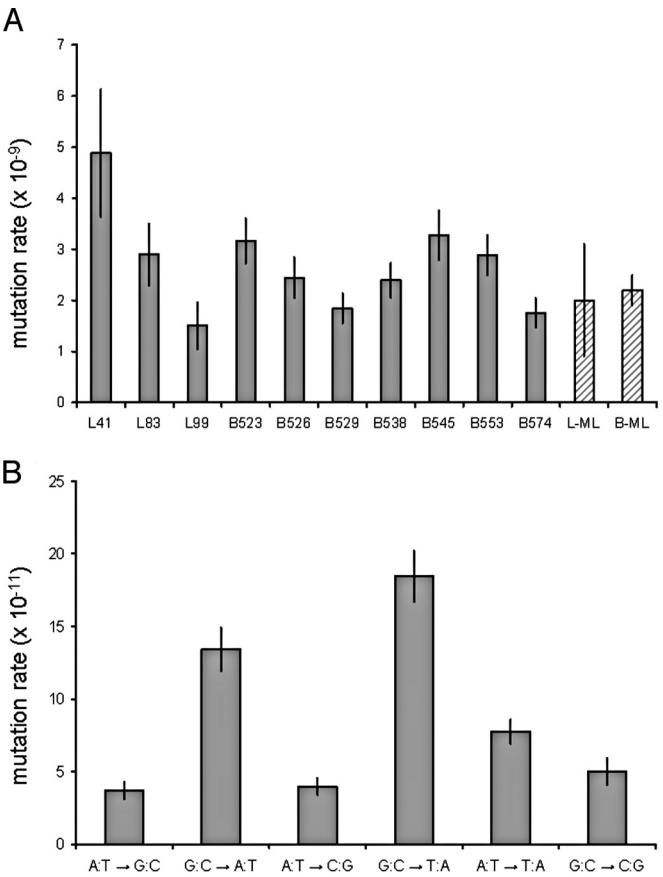

Fig. 1. MA-line mutation rate estimates. (*A*) Mutation rate estimates specific to each of the 10 MA-line genomes analyzed (solid gray bars) and maximum-likelihood mutation rate estimates (striped bars) for the three genomes analyzed by 454 (L-ML) and the seven analyzed by Solexa (B-ML). Error bars for the 10 line-specific estimates show SEM, and those for the maximum-likelihood estimates show SD. (*B*) Conditional mutation rate estimates for each of the six nonstrand-specific base substitution mutation types. Error bars show SEM.

EVOLUTION

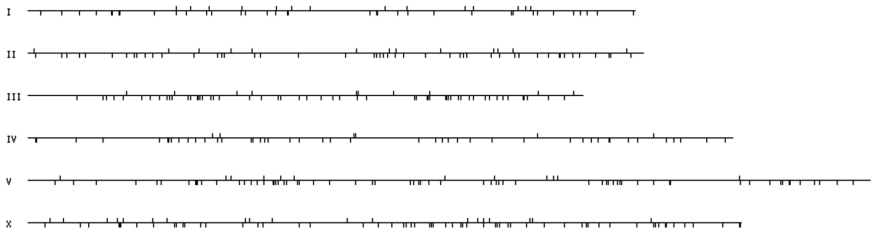

Fig. 2. Distribution of 391 MA-line base substitutions across *C. elegans* chromosomes. Tick marks indicate the physical positions of MA-line mutations across the six *C. elegans* chromosomes; those above the chromosomes derive from the three MA lines analyzed by 454; those below the line are from the seven lines analyzed by Solexa.

the causes underlying the differences between the μ_{bs} values reported here and the previous estimate, repetitive DNA sequences were necessarily excluded in the current analysis, which relied on short-read data. Among the 12 sites in the *C. elegans* genome where base substitutions were observed in the previous analysis, six were in repetitive DNA stretches that were unable to be considered for mutation here. Thus, it is likely that the μ_{bs} estimates for *C. elegans* presented here are slightly downwardly biased, with repetitive DNA sequences experiencing higher mutation rates than the unique genomic regions analyzed here.

Transformation of our current per-generation μ_{bs} value for *C. elegans* to a per-germ-line cell division mutation rate, assuming 8.5 germ-line cell divisions per generation (15), yields an estimate of 3.2×10^{-10} per site per cell division. This value is highly similar to the current per-cell division rate estimate for *S. cerevisiae*, 3.3×10^{-10} (8), and is only 2- to 3-fold higher than current estimates for *D. melanogaster*, 1.5×10^{-10} (7), and humans, 1.0×10^{-10} (16–18). The nominal rate variability observed in this phylogenetically disparate group of eukaryotic model organisms suggests that despite extensive variability in the DNA metabolism and repair machinery encoded by eukaryotic genomes (19, 20), there might be substantial per-cell division mutation rate stasis among many distantly related eukaryotic taxa.

Mutational Bias. We analyzed mutation spectra in the MA lines by considering the six nonstrand-specific base-substitution types, expressed as the type-specific mutation rate conditioned on the underlying number of sites that fit our criteria for evaluation (i.e., numbers of considered G:C and A:T sites for each respective set of associated mutation types). We found that G:C→A:T transitions and G:C→T:A transversions experienced much higher average conditional mutation rates among the 10 MA-line genomes compared with the other four mutation types (Fig. 1*B*). Although the particular sources of DNA damage or error leading to the mutations detected here are unknown, the observed pattern of mutational bias is suggestive of oxidative DNA damage playing a major role in MA-line mutagenesis; 5-hydroxyuracil (resulting from the oxidative deamination of cytosine) and 8-oxoguanine are the two most common types of oxidative DNA damage in animal genomes (21, 22), and they cause G:C→A:T transitions and G:C→T:A transversions, respectively, the two most common base-substitution types observed here in the *C. elegans* MA lines. These two mutation types were also observed to be the most prevalent among *S. cerevisiae* MA-line nuclear mutations analyzed by 454 pyrosequencing (8). If the *C. elegans* genome has achieved base-composition equilibrium through mutational pressure alone, we would expect to see equal numbers of G or C→A or T, and A or T→G or C mutations in the MA lines. The former mutation types were 2.2-fold more frequent than the latter, however, suggesting that underlying base-substitution processes alone have not shaped *C. elegans* base composition.

Transition bias, i.e., transition/transversion (Ts/Tv) base-substitution ratios greater than the random expectation of 0.5, is widely observed in the natural genomic variation of animals, although it remains unclear whether this bias is caused by mutation or selective forces (23, 24). We compared the observed relative abundances of transition and transversion in the MA lines to previously published Ts/Tv ratios in *C. elegans* natural-isolate nuclear genomes to investigate the relative roles of mutation and selection in shaping Ts/Tv ratios observed in nature. The average Ts/Tv ratio among the MA lines examined here was 0.45, with line-specific values ranging from 0.19 to 0.79 (Table 1). Although the MA-line average was very close to the random expectation of 0.5, our observations do not fit a null expectation of completely random mutation, because different transition and transversion subtypes occurred at different rates (Fig. 1*B*). Sequences frequently considered to be neutral (pseudogenes, silent codon positions, intron and intergenic sequences) in nuclear genomes of *C. elegans* natural isolates display Ts/Tv ratios that are much higher, ranging from 1.2 to 3.0, depending on the analysis (25–28), and are typical of patterns observed in many other animal species. The much lower Ts/TV ratios observed in MA-line genomes suggest that, genome-wide, transversions might be more susceptible to selective purging than transitions in *C. elegans* natural populations. However, this pattern might also result if segregating polymorphisms at putatively neutral sites in the *C. elegans* genome are not at mutational equilibrium. Another possible explanation that cannot be ruled out is that basic differences in the mutational environment of *C. elegans* evolving in nature versus that experienced by the MA lines are responsible for the observed differences between MA-line mutations and natural patterns of polymorphism in presumably neutral regions. Although mutation processes are reported to not significantly differ between microbes grown in the lab and those evolving in nature (29), the extent to which environmental differences between natural and laboratory conditions affect *C. elegans* mutagenesis remains unknown.

Genomic Distributions of Mutations. The physical distribution of mutations was uniform across the six chromosomes (Fig. 2); the observed numbers of chromosome-specific mutations did not deviate significantly from expectations based on the genome-wide mutation totals and chromosome-specific numbers of sites that met our criteria for consideration ($P = 0.18$, χ^2 test, 5 df), and the distribution of mutations on the X chromosome versus autosomes in the MA lines was also not significantly different from null expectations ($P = 0.91$, χ^2 test, 1 df). The five *C. elegans* autosomes are by convention each subdivided into external arm regions (a left arm and a right arm for each chromosome) and a central core region; the arms have higher recombination rates, higher repetitive-element loads, and lower gene densities compared with the core regions (30, 31). Although previous analyses of *C. elegans* natural-isolate SNP patterns showed that SNPs occur in much greater densities in autosomal arms compared with core regions (26, 27), whether mutational or selective forces are responsible for this pattern has remained unknown. It has been proposed that recombination is a mutagenic force (32, 33) that might lead to higher mutation rates on the arms versus core regions, but genetic hitchhiking and/or background selection

might also be primarily responsible for reduced SNP densities in the cores relative to the arms (34). We compared the observed distributions of MA-line base substitutions in arm and core regions to null expected values based on uniform distributions in these regions and the associated numbers of sites that fit our criteria for analysis, and we saw no significant mutation distributional biases with respect to chromosomal arm/core boundaries ($P = 0.31$, χ^2 test, 8 df). We also tested for correlations between recombination rates and mutation densities and observed no significant correlations (Pearson's correlation coefficient $r = 0.32$, $P = 0.24$, two-tailed t test). These findings suggest that increased genetic hitchhiking and/or background selection in core versus arm regions is the more likely explanation for lower natural-isolate SNP densities in the cores versus arms rather than a mutational bias associated with recombination.

The distributions of mutations in coding sequence categories were considered by mapping mutations with respect to the boundaries of four annotated functional categories in the *C. elegans* genome: exon, intron, untranslated regions (both 5′ and 3′), and intergenic regions. Highly similar mutation rates were observed in these four functional sequence categories (Fig. S1), suggesting equal mutational susceptibilities for each of these categories in the MA-line genomes and further corroborating that selection has minimally affected the pool of mutations analyzed here. We observed 56 nonsynonymous mutations (including two mutations resulting in premature termination codons) and 24 synonymous mutations in exon sequence. These numbers were not significantly different from expected values based on the genetic code, *C. elegans* codon usage patterns, and mutational bias patterns observed in the MA lines ($P = 0.08$, χ^2 test, 1 df), although the number of nonsynonymous mutations observed ($n = 56$) was lower than the null expectation ($n = 62.5$), which might reflect selection against some highly deleterious nonsynonymous changes in the MA lines.

Conclusions. This study provides a genome-wide view of the spontaneous base-substitution mutation spectrum in an animal nuclear genome. Our high-throughput DNA sequencing approaches yielded an order of magnitude more mutation data than previous MA-line mutation surveys, which relied on conventional molecular mutation detection approaches (6, 7). The distribution of detected base substitutions in MA-line genomes was highly uniform across chromosomes and with respect to functional coding categories of DNA sequence and that nonsynonymous and synonymous sites accumulated mutations at their expected relative rates in the MA lines. Further, we observed highly similar mutation rates and patterns in two different particular sets of *C. elegans* MA lines derived from the N2 lab strain, each analyzed by using a different DNA sequencing technology, suggesting that mutation processes in the MA-line genomes analyzed have been minimally affected by mutations accumulated during the course of MA experiments. The strong discordance between our observed pattern of base-substitution types in the MA lines (transversions predominant) and natural nucleotide polymorphism patterns in *C. elegans* natural isolates (transitions predominant) challenges the widely held assumption that patterns of natural variation in presumed neutral genomic regions accurately reflect underlying spontaneous mutation processes, although alternative explanations cannot be ruled out. We are optimistic that the rapidly declining costs and increasing read lengths of high-throughput sequencing technologies will enable more comprehensive MA-line-based mutational studies that include repetitive regions and large-scale changes in the near future.

Materials and Methods

DNA Preparation for High-Throughput Sequencing. For *C. elegans* MA-line genomes analyzed by using Solexa, DNA was extracted by using a Qiagen DNeasy tissue miniprep kit, according to the manufacturer's protocol, and then prepared according to standard Illumina protocols for genomic DNA samples. A total of 2.5 pmol of prepared DNA sample was loaded into each lane of a Solexa flow cell for analysis, and seven lanes were used for each genome analyzed. The Illumina Genome Analyzer operated by the Oregon State University Center for Genome Research and Biocomputing was used. More detail on the Solexa sequencing process can be found at www.illumina.com/pages.ilmn?ID=203. Similarly, DNA for *C. elegans* genomes analyzed by using 454 was extracted by using a Qiagen genomic tip 500/G kit, according to the manufacturer's protocol. The DNA quality control for each line was done by using a spectrophotometer assay (OD A260/280 $\approx$ 1.8) and then sent to The Center for Genomics and Bioinformatics at Indiana University for 454 processing. Additional details on the 454 sequencing process can be found at www.454.com/products-solutions/how-it-works/index.asp.

Mutation Identification Procedures. Technical details of the computational analyses performed to identify candidate mutations and molecular procedures applied to confirm mutation subsets are described in *SI Text*. Lists of mutations identified and used in analyses are provided in Tables S1 and S2.

Mutation Rate Analyses. Individual MA-line-specific mutation rates were calculated with the equation $\mu_{bs} = m/(nT)$, where μ_{bs} is the base substitution mutation rate (per nucleotide site per generation), m is the number of observed mutations, n is the number of nucleotide sites, and T is the time in generations, as described (6). The standard errors for individual mutation rates were calculated as $[\mu_{bs}/(nT)]^{1/2}$, as described (6). Values for n are shown in Table 1, reflecting the total number of base pairs surveyed that met our criteria for consideration of a possible mutation site (*SI Text*).

We also applied a maximum-likelihood method (14) to estimating mutation rates. For this approach, we assumed that all reads covering a site must call either the true base in that site or an erroneous base due to the inherent error frequency of the sequencing technology. We further assumed that a site must be either nonmutant (and hold the same base in all MA lines) or mutant in no more than one MA line. The likelihood of a specific configuration of reads in a site across several MA lines is a function of the error frequency (ε) and the mutation rate per site per generation (μ; equation 14 in ref. 14). Newton-Raphson's method of optimization was implemented in custom-made C++ scripts and was used to obtain the estimates of ε and μ that maximized the sum of log likelihoods across sites. The initial values of å and μ to start the optimization process were taken from a consensus approach. The sites used to get the estimates had: all MA lines covered by at least four reads, no more than one MA line with two or more bases read, a consensus base supported by at least 76% of reads, and both forward and reverse reads present in all MA lines. A total of 1,317,950 sites matched these requirements in the 454 data; 14,836,133 sites matched these criteria in the Solexa data.

ACKNOWLEDGMENTS. We thank Mark Dasenko, Chris Sullivan, and Scott Givan at the Oregon State University Center for Genome Research and Biocomputing for Solexa support. This work was supported by National Institutes of Health Grants GM087678 (to C.F.B. and D.R.D.) and GM036827 (to M.L. and W.K.T.).

1. Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: Causes and consequences. *Nat Rev Genet* 8:619–631.
2. Charlesworth B (2009) Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet* 10:195–205.
3. Hurst LD (2002) The Ka/Ks ratio: Diagnosing the form of sequence evolution. *Trends Genet* 18:486.
4. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. *Genetics* 148:1667–1686.
5. Li WH (1997) *Molecular Evolution* (Sinauer, Sunderland, MA).
6. Denver DR, Morris K, Lynch M, Thomas WK (2004) High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* 430:679–682.
7. Haag-Liautard C, et al. (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445:82–85.
8. Lynch M, et al. (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105:9272–9277.
9. Keightley PD, et al. (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res* 19:1195–1201.
10. Baer CF, et al. (2005) Comparative evolutionary genetics of spontaneous mutations affecting fitness in rhabditid nematodes. *Proc Natl Acad Sci USA* 102:5785–5790.
11. Vassilieva LL, Lynch M (1999) The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* 151:119–129.
12. Dolan PC, Denver DR (2008) TileQC: A system for tile-based quality control of Solexa data. *BMC Bioinformatics* 9:250.
13. Pop M, Phillippy A, Delcher AL, Salzberg SL (2004) Comparative genome assembly. *Brief Bioinform* 5:237–248.

14. Lynch M (2008) Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. *Mol Biol Evol* 25:2409–2419.
15. Wood W (1988) *The Nematode Caenorhabditis elegans* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
16. Podlutsky A, Osterholm AM, Hou SM, Hofmaier A, Lambert B (1998) Spectrum of point mutations in the coding region of the hypoxanthine-guanine phosphoribosyltransferase (hprt) gene in human T-lymphocytes in vivo. *Carcinogenesis* 19:557–566.
17. Giannelli F, Anagnostopoulos T, Green PM (1999) Mutation rates in humans. II. Sporadic mutation-specific rates and rate of detrimental human mutations inferred from hemophilia B. *Am J Hum Genet* 65:1580–1587.
18. Kondrashov AS (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* 21:12–27.
19. Eisen JA, Hanawalt PC (1999) A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* 435:171–213.
20. Denver DR, Swenson SL, Lynch M (2003) An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. *Mol Biol Evol* 20:1603–1611.
21. Hagen TM, et al. (1994) Extensive oxidative DNA damage in hepatocytes of transgenic mice with chronic active hepatitis destined to develop hepatocellular carcinoma. *Proc Natl Acad Sci USA* 91:12808–12812.
22. Thiviyanathan V, et al. (2008) Base-pairing properties of the oxidized cytosine derivative, 5-hydroxy uracil. *Biochem Biophys Res Commun* 366:752–757.
23. Gojobori T, Li WH, Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. *J Mol Evol* 18:360–369.
24. Keller I, Bensasson D, Nichols RA (2007) Transition-transversion bias is not universal: A counter example from grasshopper pseudogenes. *PLoS Genet* 3:e22.
25. Cutter AD (2006) Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. *Genetics* 172:171–184.
26. Denver DR, Morris K, Thomas WK (2003) Phylogenetics in *Caenorhabditis elegans*: An analysis of divergence and outcrossing. *Mol Biol Evol* 20:393–400.
27. Koch R, van Luenen HG, van der Horst M, Thijssen KL, Plasterk RH (2000) Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. *Genome Res* 10:1690–1696.
28. Witherspoon DJ, Robertson HM (2003) Neutral evolution of ten types of mariner transposons in the genomes of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *J Mol Evol* 56:751–769.
29. Ochman H (2003) Neutral mutations and neutral substitutions in bacterial genomes. *Mol Biol Evol* 20:2091–2096.
30. *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* 282:2012–2018.
31. Barnes TM, Kohara Y, Coulson A, Hekimi S (1995) Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* 141:159–179.
32. Hurles M (2005) How homologous recombination generates a mutable genome. *Hum Genomics* 2:179–186.
33. Watanabe Y, Takahashi A, Itoh M, Takano-Shimizu T (2009) Molecular spectrum of spontaneous de novo mutations in male and female germline cells of *Drosophila melanogaster*. *Genetics* 181:1035–1043.
34. Cutter AD, Payseur BA (2003) Selection at linked sites in the partial selfer *Caenorhabditis elegans*. *Mol Biol Evol* 20:665–673.