A genome-wide view of the spectrum of spontaneous mutations in yeast

Michael Lynch*†, Way Sung‡, Krystalynne Morris‡, Nicole Coffey*, Christian R. Landry§¶, Erik B. Dopman§, W. Joseph Dickinson‖, Kazufusa Okamoto‡, Shilpa Kulkarni‡, Daniel L. Hartl†§, and W. Kelley Thomas‡

*Department of Biology, Indiana University, Bloomington, IN 47405; ‡Hubbard Center for Genome Studies, University of New Hampshire, Durham, NH 03824; §Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138; and ‖Department of Biology, University of Utah, Salt Lake City, UT 84112

Contributed by Daniel L. Hartl, April 22, 2008 (sent for review March 14, 2008)

The mutation process ultimately defines the genetic features of all populations and, hence, has a bearing on a wide range of issues involving evolutionary genetics, inheritance, and genetic disorders, including the predisposition to cancer. Nevertheless, formidable technical barriers have constrained our understanding of the rate at which mutations arise and the molecular spectrum of their effects. Here, we report on the use of complete-genome sequencing in the characterization of spontaneously arising mutations in the yeast *Saccharomyces cerevisiae*. Our results confirm some findings previously obtained by indirect methods but also yield numerous unexpected findings, in particular a very high rate of point mutation and skewed distribution of base-substitution types in the mitochondrion, a very high rate of segmental duplication and deletion in the nuclear genome, and substantial deviations in the mutational profile among various model organisms.

chromosomal instability | mitochondrion | mutation rate | mutational spectrum | *Saccharomyces cerevisiae*

Despite its relevance to every aspect of genetics and evolution, our understanding of the mutation process and its bearing on organismal fitness remains quite limited (1–4). Owing to the technical difficulties in directly observing very low-frequency events, most estimates of the per-nucleotide mutation rate are derived either from surveys of visible mutations at reporter loci (to enhance the detectability of mutations) or from nucleotide-sequence comparisons of silent sites in distantly related species (to magnify the accumulation of mutations). Neither approach is without problems, the first requiring assumptions about the fraction of mutations with observable phenotypic effects and the second relying on assumptions about interspecific divergence times, generation lengths, and neutrality of the monitored nucleotide sites.

Long-term mutation-accumulation (MA) experiments, whereby replicate lines are taken through regular bottlenecks to minimize the efficiency of selection, have proven to be highly valuable resources for procuring spontaneous mutations in an essentially unbiased fashion (5–8). However, brute-force sequencing of PCR-amplified products constrains the number of mutations that can be detected in a reasonable amount of time. Here, we demonstrate the feasibility of whole-genome sequencing as a means to assay the complete spectrum of mutational effects in a moderately sized eukaryotic genome.

Our analyses are based on an examination of parallel MA lines of a key model system, the yeast *Saccharomyces cerevisiae*. The initially isogenic lines were passed through 200 single-cell bottlenecks on a 3- to 4-day cycle of clonal growth for a total of ≈4,800 cell divisions per line [see supporting information (SI) *Text*]. Although there is some opportunity for the selective elimination of deleterious mutations during daily clonal amplification, this effect is quite small under the imposed bottlenecking procedure. For mutations with a relative selective disadvantage of $s = 0.001$, the bias against mutation fixation is <1%, whereas that for mutations with $s = 0.01$ is <8% (*SI Text* and Fig. S1). Thus, because the fitness effects of most new mutations are on the order of 0.01 or smaller (2), our experiment captures the vast majority of newly arisen mutations. The rate of accumulation of such effectively neutral mutations provides a direct estimate of the mutation rate (9).

We detected mutations by pyrosequencing the complete ≈12-Mb genomes of four MA lines (10). The average depth of sequence coverage for each line was ≈5×, and we restricted our analyses to sites within each genome with at least 3× coverage, which for the observed base-call error rate yields a negligible probability of false-positive consensus sequences (*SI Text*).

Results

Base Substitutions. An average of 4.99×10^6 nucleotide sites (half the genome size) fulfilled our criteria for assayable complex sequence in each line, revealing 33 base-substitutional changes (all of which were subsequently verified by direct sequencing) and an overall base-substitutional mutation rate estimate of $0.33\ (\mathrm{SE} = 0.08) \times 10^{-9}$ per site per cell division. Other than base substitutions, the only small-scale mutations that we observed in complex sequence were a single 1-bp deletion and one 3-bp inversion sandwiched between a palindrome. (For a summary of simple-sequence mutations detected, see Table S1). Given the base-call error rate, our analyses are essentially free of bias from false-negatives (failure to detect true mutations) (*SI Text*). Our base-substitutional mutation-rate estimate is not significantly different from the average estimate from four previous reporter-construct studies, $0.72\ (0.33) \times 10^{-9}$ per site per cell division (11, 12). In addition, the transition/transversion ratio, 0.62, is nearly identical to that obtained from prior studies (0.61, using the data in Table 1); and there is no indication that the observed mutation spectrum is biased by selection: 27% of the mutations were in intergenic DNA, which comprises 25% of the yeast genome; and of the coding-region mutations, 25% were at silent sites, which is essentially the same frequency expected by chance. Taken together, these observations on the mutation rate and mutational spectrum, along with the fact that the average relative fitness of the cell lines had declined only by 5% by the time of sequencing (13), strongly supports the claim that our results are unbiased by selection.

If the yeast genome has achieved nucleotide-composition equilibrium under mutational pressure alone, the numbers of

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

The authors declare no conflict of interest.

†To whom correspondence may be addressed. E-mail: milynch@indiana.edu or dhartl@oeb.harvard.edu.

¶Present address: Département de Biochimie, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, QC, Canada H3C 3J7.

This article contains supporting information online at www.pnas.org/cgi/content/full/0803466105/DCSupplemental.

Table 1. Spectrum of base-substitutional mutations

Method	Transitions		Transversions				*N*
	A:T → G:C	G:C → A:T	A:T → T:A	G:C → T:A	A:T → C:G	G:C → C:G	
SUP4-o*	0.127	0.266	0.044	0.314	0.043	0.207	798
CAN1†‡	0.094	0.358	0.075	0.269	0.066	0.137	212
URA3*‡	0.022	0.241	0.149	0.351	0.057	0.175	228
Summary	0.102	0.277	0.069	0.313	0.050	0.189	1238
Nu, this study	0.118	0.265	0.059	0.265	0.147	0.147	34
Mt, this study	0.476	0.000	0.333	0.000	0.191	0.000	21

The first three rows summarize results from prior studies using single-locus reporter constructs in the nucleus, and the fourth row summarizes these data. Nu, nuclear; Mt, mitochondrial.
*ref. 49.
†ref. 50.
‡ref. 12.

A/T → G/C and G/C → A/T mutations should be equal. However, the latter is 2.0 times greater than the former (Table 1), and on a per-site basis, the rate of G/C → A/T mutation is 2.9 times that in the opposite direction. This biased pattern of mutation is consistent with observations on nucleotide instability: Spontaneous deamination of cytosine or 5-methylcytosine is a major source of C:G → T:A transitions; and the conversion of guanine to 8-oxo-guanine (GO) is a major source of C:G → A:T transversions (14, 15).

These observations imply that the nucleotide composition of the yeast genome cannot be explained by mutation pressure alone. Under a process governed entirely by mutation bias, our data predict an equilibrium A/T composition of 0.74, whereas the data from prior reporter-construct studies (Table 1) imply a somewhat higher expected value of 0.80. Thus, the fact that the fractional A/T content of the yeast nuclear genome in the regions that we have analyzed is 0.59 implies that mutation pressure in the direction of A/T composition has been opposed historically by other forces in natural populations. Although a role for natural selection cannot be ruled out, gene conversion in yeast is strongly biased in the direction of G/C (16, 17). Assuming that the nucleotide composition of yeast is in mutation–drift-conversion equilibrium, then given the observed mutational bias, biased gene conversion must cause mutations toward G/C to fix with a probability 2.0 times that in the opposite direction. Because the latter ratio is equivalent to e^{-2Ns}, where N is the effective population size, and s is the coefficient of gene conversion (box 6.2 in ref. 18), the historical power of biased gene conversion in this species appears to be ≈35% of the power of random genetic drift.

Microsatellites. Repetitive DNAs involving di-, tri-, and tetranucleotide repeats are known to mutate at high rates to novel-length variants as a consequence of replication slippage, with the mutation rate generally increasing dramatically with the number of repeats at the locus (19–23). The yeast genome is relatively devoid of such loci, with 1,092 having 5–10 repeats, 282 having 11–20 repeats, 7 having 30–46 repeats, and none larger (Table S2). To confidently ascertain the number of repeat units at a microsatellite locus by random sequencing, multiple sequences must cover the entire repetitive span, so the average depth of informative coverage will be less than that associated with individual base calls. Across all lines, we were able to assay 50% of the 3,049 loci in length categories 4–14 but found no evidence of mutations. For these loci, we can only place an upper statistical bound on the mutation rate compatible with such observations (Fig. 1 and *SI Text*).

To supplement this analysis, we directly screened by PCR the full set of MA lines for 21 di- and trinucleotide microsatellite loci with various repeat motifs and repeat numbers in the range of 15–36. Of the 42 detected mutations, 81% were insertions and 74% involved single (±1) repeat changes (Table S3). A study of a 16-repeat poly(GT) construct yielded qualitatively similar results—of 370 observed mutations, 71% were insertions, and 99% involved single repeat changes (24). Likewise, a screen of constructed microsatellite tracts with 16 and 33 repeats revealed 64% insertions and 94% single-repeat mutations (19). Although there is a clear bias toward insertion mutations, this appears to be compensated by the relatively large sizes of rare deletions (Table S3).

Our direct survey yielded mutation rate estimates of 7.2 (0.7) × 10^{-7}, 1.7 (0.9) × 10^{-5}, and 7.2 (3.2) × 10^{-5} per locus per cell division, respectively, for loci containing 15 to 17, 20–24, and 31–36 repeats, which are substantially larger than prior estimates derived from two GT-dinucleotide reporter constructs (Fig. 1).

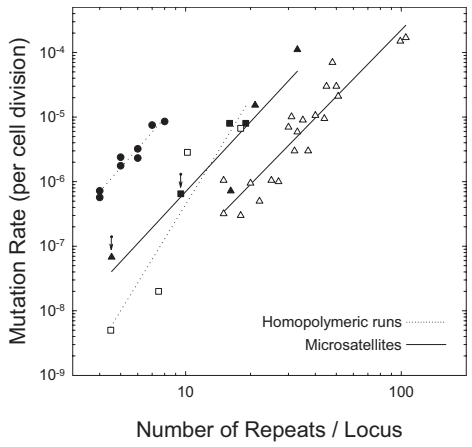

Fig. 1. Microsatellites (triangles, solid lines). The three rightmost filled triangles denote direct estimates of the mutation rates for classes of microsatellites containing large numbers of repeats (n = 10, 8, and 3, respectively); the bottom filled triangle is a rough estimate of the mutation rate for the smallest class, taken to be the midpoint between the upper statistical bound (base of the arrow) and zero; fitted regression to all four points: (1.75 × 10^{-10}) × $L^{3.60}$ (r^2 = 0.88), where L is the number of repeats. Open triangles denote estimates from two earlier studies with reporter constructs (19, 31); fitted regression: (3.24 × 10^{-11}) × $L^{3.42}$ (r^2 = 0.87). Circles, squares, and dotted lines indicate homopolymeric runs. Maximum-likelihood estimates (circles) of the mutation rate of HPR loci as a function of the run length are fitted to the regression line, with each data point representing the average result over all four MA lines, with separate points for A/T and G/C runs. The regression equation is (3.73 × 10^{-9}) × $L^{3.78}$ (r^2 = 0.94). The filled squares denote direct estimates derived from sequence analysis of PCR products in this study, whereas the open squares denote results from other studies (28, 29, 32); the joint regression for these latter two sets of points is (1.50 × 10^{-12}) × $L^{5.47}$ (r^2 = 0.88).

EVOLUTION

This disparity may exist because prior artificial constructs fortuitously experienced low levels of replication slippage or more efficient repair, perhaps because of unique features of GT repeats. Poly(GT)$_{16}$ has been used in a variety of other constructs in *S. cerevisiae*, leading to a very broad range of mutation rate estimates, 3×10^{-7} to 1×10^{-4} (19, 24–28), so it is clear that the context of individual loci can greatly influence the degree of mutagenicity. Despite the elevated level, the scaling of the mutation rate with repeat number that we observed is quite consistent with prior work (Fig. 1), and our rate estimates are also within the range of direct estimates derived from an MA experiment with the nematode *Caenorhabditis elegans*, 6.6×10^{-6} to 1.2×10^{-4} per locus per germ-line cell division for loci with 20–35 repeats (23).

Homopolymeric Repeats. The yeast genome contains large numbers of homopolymeric runs (HPRs) extending up to 20 consecutive bases, and like microsatellites, such repeats can have elevated mutation rates resulting from replication slippage. HPRs present a challenge for pyrosequencing, because the read-error rate in such regions is relatively high compared with single-base misreads. To minimize the likelihood of false positives in the estimation of the mutation rate for HPR loci, we developed a procedure that utilizes the full set of sequence reads across all such loci to obtain joint maximum-likelihood (ML) estimates of the read-error rate and the mutation rate for specific ancestral run lengths (*SI Text*).

The overall results were highly consistent between A:T and G:C HPRs, with the mutation rate scaling with the length of the HP tract in a manner parallel to that for microsatellite loci, from $\approx 7.0 \times 10^{-7}$ / cell division for runs of four to 1.1×10^{-5} for runs of eight (Fig. 1). Surveys of single 16- and 19-bp G/C runs by direct sequence analysis of PCR products across the set of haploid lines each revealed single mutations, for an estimated rate of 8.0×10^{-6} at each locus. However, a survey of 18 9- and 10-bp runs (an even mixture of G/Cs and A/Ts) revealed no mutations, which at the 0.05 probability level, is compatible with a mutation rate no higher than 1.3×10^{-6}, approximately an order of magnitude smaller than the ML estimate for runs of this length. In principle, these discrepancies could again be simple consequences of a high level of variation in the mutation rate among loci. For example, Gragg *et al.* (29) found the average mutation rate of 10-bp mononucleotide runs in artificial constructs to be 3.4×10^{-6}, although the estimates for individual constructs ranged from 2.4×10^{-8} to 1.0×10^{-4}, with >100-fold elevation in rates for G/C homopolymers.

Taken together, these data suggest a higher mutation rate for homopolymeric runs than for microsatellites (Fig. 1), consistent with previous observations from *C. elegans*, which indicate an ~140-fold inflation of the mutation rate for mononucleotide runs of 12 bp relative to loci containing the same number of dinucleotide repeats (7, 23). Analyses with human cell lines also suggest much greater instabilities for mononucleotide than dinucleotide repeats (30). Nevertheless, we emphasize that a precise determination of the mutational profile for nuclear HPR loci in yeast remains to be made.

Mitochondrial Mutations. Although shotgun sequencing resulted in ≈44× coverage of the mitochondrial genome, the extraordinarily high A/T composition (0.83) of this genomic compartment results in a very high density of HPRs, and the smaller absolute number of these precludes the type of ML analysis noted above. Nevertheless, an average of 54,762 nucleotide sites suitable for detecting mitochondrial mutations in complex sequence in each line revealed 13 base substitutions and a corresponding mutation-rate estimate of 12.9 (3.6) $\times 10^{-9}$ per site per cell division. A second estimate of the mitochondrial base-substitutional mutation rate, obtained by conventional sequencing of PCR products encompassing a total of 17,786 bp across the full set of lines, is 4.0 (2.2) $\times 10^{-9}$ per site per cell division. The weighted average of these two estimates, based on the number of nucleotides surveyed, is 12.2×10^{-9} per site per cell division, ~37 times that observed in the nuclear genome. This disparity is even more extreme than the ratio of mitochondrial to nuclear base-substitution mutation rates in vertebrates and invertebrates, which average 19 and 8 times, respectively (18), raising questions about the common view that elevated mitochondrial mutation rates are unique to animals.

Remarkably, each of the base substitutional mutations that we observed initiated at an A/T site (67% to G/C), which leads to the prediction that the entire genome would evolve to a G/C composition of 1.00 in the absence of opposing pressures. Because the A/T composition in the regions analyzed is 0.84, it is clear that some force must be strongly opposing the mutation pressure toward G/C. Following the logic outlined above for the nuclear genome, and assuming the minimum fraction of mutations toward A/T compatible with the observed data (0.003, at the 0.05 probability level), then average $2Ns$, where s is the selective advantage of A/T, is equal to 7.47, which implies that the historical power of selection (and/or biased gene conversion) toward A/T in nature must have been at least 3.7 times the power of drift.

In contrast to the situation in the nuclear genome, the insertion/deletion rate in mitochondrial complex sequence determined by random sequencing, 10.4 (3.0) $\times 10^{-9}$ per site per cell division, is nearly as high as the rate for base substitutions. A second estimate of the insertion/deletion rate determined from a PCR survey of the full set of lines, 5.3 (2.1) $\times 10^{-9}$, is compatible with the former, yielding a weighted average of 7.5×10^{-9} per site per cell division. There were nearly equal numbers of insertions ($n = 14$) and deletions ($n = 16$), although all of the former involved single bases, whereas the latter ranged from 1 to 6 bp in length with a mean of 3.2 (0.6). The 0.58:1 ratio of indel to base-substitution mutations in the mitochondrial genome of this species is essentially the same as the ratio of 0.62:1 previously observed with MA lines of *C. elegans* (5).

Large-Scale Changes. To identify large-scale (>1 kb) duplications and deletions, we initially performed pulse-field gel electrophoresis (PFGE) on the full set of 32 lines. This screen revealed detectable changes in chromosome size in eight lines, with the number of apparent size increases ($n = 12$) outnumbering the number of decreases ($n = 2$) (Fig. S7). To further evaluate the nature of the chromosomal modifications driving these changes, we hybridized labeled genomic DNA to oligonucleotide arrays representing 6,388 segments distributed over all chromosomes, using oppositely labeled reference DNA from the progenitor line as a control.

The latter analyses, which were applied to all four focal lines as well as to four others, revealed 11 large-scale insertions and four deletions (Table 2). Four of the inferred changes were consistent with the types (insertion or deletion) and sizes of the alterations observed with PFGE analysis, and most of those that were not revealed by PFGE resided on large chromosomes where detection by PFGE is compromised. Five of these large-scale segmental changes (including four of the nine not identified by PFGE) are flanked on both ends by transposable elements (searched within a window of 3.0 kb). With 89 such elements spread across the 12-Mb genome, the probability of obtaining such a flanking status by chance is $<6.3 \times 10^{-8}$. Thus, our results support the idea that a substantial fraction of chromosomal instabilities in this species result from phenomena associated with mobile elements (33, 34). The few changes that were detected by PFGE but not by genomic hybridization (e.g., the dramatic increases in the sizes of chromosome 9 and 14 in line B6, and the reduction in chromosome 8 in A1) may reflect

Table 2. Rates of sequence gain and loss by gene-sized processes

	Gains:			Losses:		
Line/Changes	bp	Segments	ORFs	bp	Segments	ORFs
Line						
A1	90,538	2	34	99,724	2	53
A4	79,765	1	28	0	0	0
A7	0	0	0	69,511	1	25
B6	67,612	1	31	81,561	1	44
C3	541,056	8	285	60,633	1	28
C5	74,270	1	28	0	0	0
C8	134,599	3	48	6,076	1	4
D1	501,123	5	285	601,163	12	306
Change per line per cell division:						
Mean	38.78	0.00055	0.01924	23.92	0.00047	0.01198
SE	15.18	0.00019	0.00863	14.76	0.00029	0.00754

All segments are statistically significant at the 0.01 level after Bonferroni correction.

nonduplicative translocations generated by nonhomologous recombination or expansion/contraction of repeats not present on the microarray.

The microarray analyses also provided insight into the dynamics of chromosome copy number within putatively haploid yeast cells. By combining information on the chromosome-wide averages of locus-specific reads from microarrays and depth-of-coverage observations from the random sequencing data, it became clear that almost all chromosomes within the four focal lines had become diploid by the time of analysis, with chromosome 9 in line A4 being monosomic, and chromosome 1 in line C5 being at least trisomic and possibly transiently in a tetrasomic state (Fig. 2). (For the four other lines assayed by both PFGE and genomic hybridization, a single example was found of the addition of the equivalent of chromosome 13, yielding apparent trisomy). Although there is no evidence that diploidization was promoted by selection in these lines (13), such behavior raises the concern that our approach to identifying mutations from novel consensus sequences might lead to downwardly biased mutation-rate estimates (by overlooking heterozygous mutations). However, our near complete failure to find heterozygous mutations in further genomic screens and direct analyses of a temporal series of frozen samples (*SI Text*) is consistent with the hypothesis that most individual chromosomes in these lines experience brief stochastic returns to monosomy, although high rates of mitotic gene conversion could also be involved.

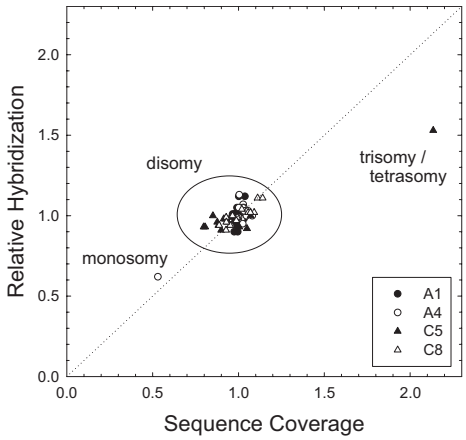

Fig. 2. Inferred ploidy levels for each chromosome within the four surveyed lines. Each chromosome-wide measure is an average over all surveyed nonredundant sequences, scaled by the genome-wide average, so 0.5, 1.0, 1.5, and 2.0 imply monosomy, disomy, trisomy, and tetrasomy, respectively.

These results are concordant with other observations on haploid lines of *S. cerevisiae* transforming to a nearly diploid state within a few hundred generations, with a small probability of individual chromosomes being in transient states of monosomy or trisomy at any point in time (35) and with some very large-scale intrachromosomal changes resulting from massive deletions and/or nonhomologous recombination (36–38). For large segmental deletions/duplications detected by microarray hybridization of genomic DNA, see Table S4. Indeed, both types of changes may be partially interconnected, in that diploidy may be selectively imposed when a key gene is deleted, forcing the cell to depend on suboptimal gene products from paralogous loci (39).

With a large fraction of long-segmental variation identified with microarray analysis exhibiting consistency with the PFGE results, we extended the former to smaller-scale segmental insertions/deletions (Table 2). Our analysis provides the first direct genome-wide estimates of the rates of gene duplication and gene loss, 3.4 (1.5) $\times 10^{-6}$ and 2.1 (1.3) $\times 10^{-6}$ per gene per cell division, respectively (not including whole-chromosome duplications) (from Table 2). Given the restriction of such analysis to segments >1 kb in size and the observation that ~32% of insertion/deletion events >0.5 kb in length in yeast are in the range of 0.5–1.0 kb (40), these estimates are likely to be <68% of the actual rates. Nevertheless, our estimated rate of gene duplication is substantially higher than previous indirect estimates for eukaryotes (derived from duplicate-gene age distributions in sequenced genomes), which are generally in the neighborhood of 10^{-8} per gene per generation (18).

From the above results, we infer that, on a per-cell-division basis, the expected number of gene duplications per genome is 0.019 (0.023 if putative segmental duplications are accepted at probability level 0.05 rather than 0.01; see Table S5), which exceeds the expected number of base substitutions per genome per generation (0.004). Prior work on these MA lines revealed significant correlations between the among-line expression variance of a gene and the number and diversity of upstream genes affecting its expression in the transcriptional cascade (41). Given the very small number of point mutations within individual lines, and because ~25% of nucleotide sites within coding regions are silent and ~27% of the yeast genome consists of noncoding DNA, our results strongly suggest that the correlations revealed in the prior study are largely a consequence of gene deletion and duplication.

EVOLUTION

Table 3. Comparisons of mutation rates per cell division

	Base substitution		Small insertion/deletion		Nu microsatellite		Nu homopolymer	
Organism	Nu	Mt	Nu	Mt	10 rpt	25 rpt	5 bp	15 bp
S. cerevisiae	0.33	12.23	0.02	7.48	0.70	18.86	4.07	104.07
C. elegans	1.07	11.58	1.16	6.51	0.31	6.82	0.04	10.75
D. melanogaster	0.16	0.38	0.04	—	0.56	2.02	—	—
Homo sapiens	0.10	2.09	0.01	—	4.78	52.30	27.39	483.62

Data are in units of 10^{-9} per germ-line cell division for base substitutions and small insertion/deletions, and as 10^{-6} per germ-line cell division for repetitive DNAs. Small insertion/deletions are 1–3 bp in length. *C. elegans*: mutation rate estimates (5–7, 23); assumes 8.5 germ-line cell divisions per generation (51). *D. melanogaster*: mutation rate estimates (8, 52–55); assumes 36 germ-line cell divisions per generation (56). *H. sapiens*: mutation rate estimates (22, 54, 56–58); rates for homopolymeric runs assume an average 35 times inflation over that for microsatellites of the same size (the observed inflation in yeast); assumes 200 germ-line cell divisions per generation (3). The rates for homopolymers in yeast are derived from the regression involving ML estimates, and reduce to 0.01 and 1.64 if the regression involving large loci is relied on.

Discussion

Because nonmammalian systems, and yeast in particular, are widely relied on in studies of mutational processes deemed relevant to human health, the extent to which mutational rates and profiles of effects deviate among model organisms is of considerable interest. Our results bear on this issue in several ways when compared with the germ-line mutation rates for several animal species on a per-cell-division basis (Table 3). First, the nuclear mutation rates to both base substitutions and small insertion/deletions are within a factor of four across yeast, *Drosophila*, and human, but are substantially higher in the nematode *C. elegans*. Second, mitochondrial mutation rates in yeast and nematode are nearly equivalent and dramatically higher than those in *Drosophila* and human. Third, nuclear mutation rates at homopolymeric runs and microsatellite loci are much higher in humans than in either yeast or nematode.

Although these estimates are only approximate, it appears clear that there are significant differences in the internal cellular environments of different eukaryotes in terms of both replication fidelity and the nature of mutation. Such variation should perhaps not be too surprising, because many examples are known in which the components of various DNA-damage control pathways differ among yeast and animals and even among animal phyla (15). Clearly, the evolution of multicellularity was not accompanied by a reduction in the mutation rate, even on a per-cell-division basis, and because the number of germ-line cell divisions is ≈9 in nematodes, 36 in flies, and 200 in humans, the per-generation rates of mutation in these organisms is one to nearly three orders of magnitude greater than that in yeast, depending on the nature of the genetic material.

Drawing from prior estimates of the fitness effects of spontaneous mutations, it is further possible to derive some inferences with respect to the relative consequences of mutations in these varied systems. Three prior MA studies of *S. cerevisiae* have yielded indirect estimates of the haploid genome-wide deleterious mutation rate in the range of 0.000048–0.0011 (42–44), which averages to 0.00041. In contrast, our molecular results (Table 2) imply a nuclear genome-wide mutation rate per cell division of ≈0.32 (0.0041 base substitutions, 0.0002 small insertion/deletions in complex sequence, 0.0019 microsatellite mutations, and 0.3094 homopolymer mutations). This suggests that only ≈0.1% of the mutations in this species have fitness effects discernible in laboratory experiments (although this fraction would rise to ≈1.8% if it were assumed that our ML estimates of yeast HPR mutation rates are too high, and the lower regression in Fig. 1 is relied on instead). Although roughly 10% of mutations with discernible effects on fitness might be beneficial in *S. cerevisiae* (13, 44), this does not substantially alter this conclusion.

A comparable analysis yields a similar conclusion for *C. elegans*. This species incurs an estimated 2.9 mutations per haploid genome per generation (2.1 base substitutions and small insertion/deletions, 0.8 homopolymeric run changes, and 0.01 microsatellite mutations), whereas the genome-wide deleterious mutation rate has been consistently estimated as ≈0.015 (45, 46), implying that ≈0.5% of mutations in this species have discernible effects on fitness. In contrast, the total haploid genomic mutation rate in *D. melanogaster* is estimated to be ≈0.99 per generation (8), and although considerable disagreement exists on the genome-wide deleterious rate in this species, most plausible estimates fall in the range of 0.05–0.5 per haploid genome (2, 47, 48). Thus, even if the lower of the above values is deemed closer to reality, a much higher fraction of mutations (≈5%) appears to influence fitness in *Drosophila* than in other species.

Mutations initially arise when base misincorporations or insertion/deletions remain after proofreading by the replicating DNA polymerase, and become established when mismatch repair (MMR) pathways fail to correct such errors. Numerous studies have estimated the magnitude by which the mutation rate is inflated when MMR is knocked out, yielding averages of 27.0 (1.9) for yeast, 54.0 (8.1) for mammals, and 41.9 (5.8) for *C. elegans* (M.L., unpublished work) for base substitutions and small insertion/deletions in complex sequence. Although these results suggest that MMR operates more efficiently in animals than in yeast, they also imply that MMR eliminates ≈96.3% of pre-mutations in yeast and 98.1% and 97.6% in mammals and *C. elegans*, respectively. Thus, the large apparent differences in the mutational profiles among these species are not likely to be consequences of differences in MMR but, rather, of variation at the level of cellular mutagenicity and/or replication proofreading. In any event, given the disparities in the mutational features of the best studied model species, it is clear that some caution is needed in extrapolating from such systems to human biology.

Materials and Methods

Mutation-Accumulation Lines. Spontaneous mutations were accumulated in 32 independent sets of haploid lines, 8 in each of four series, A–D, all initiated from the same clonal parent derived from a single laboratory stock, FY10 (leu2Δ1, ura3–52, MATα) (60). These four sets of lines are entirely isogenic except for differences at a single selectively neutral marker and for the allele at the mating-type locus; the manipulations at these loci were performed to

facilitate downstream competitive fitness assays, and have no bearing on the results of this study. From each of the four sets of lines, eight replicate sublines were passed through single-cell bottlenecks on a 3- to 4-day cycle for 200 cycles. Each cycle was initiated by picking a random colony of each line and streaking on rich YPD (yeast extract/peptone/dextrose) medium. Random propagation was imposed by overlaying each plate on a template with a marked target and selecting the colony closest to the target. The following analyses are derived from four haploid lines: A1, A4, C5, and C8.

Molecular Analysis. Technical aspects of the molecular analysis are outlined in the *SI Text*; and see Figs. S2–S7 for further details.

ACKNOWLEDGMENTS. This work was supported by National Institutes of Health Grants GM36827 (to M.L. and W.K.T.) and GM068465 (to D.L.H.). C.R.L. was supported by the Natural Sciences and Engineering Research Council of Canada, Fonds Québécois de la Recherche sur la Nature et les Technologies, and the Frank Knox Memorial Fellowship Program.

1. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. *Genetics* 148:1667–1686.
2. Lynch M, *et al.* (1999) Spontaneous deleterious mutation. *Evolution (Lawrence, Kans)* 53:645–663.
3. Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* 1:40–47.
4. Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: Causes and consequences. *Nat Rev Genet* 8:619–631.
5. Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK (2000) High direct estimate of the mutation rate in the mitochondrial genome of *C. elegans*. *Science* 289:2342–2344.
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. Denver DR, *et al.* (2004) Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of *Caenorhabditis elegans*. *J Mol Evol* 58:584–595.
8. Haag-Liautard C, *et al.* (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445:82–85.
9. Kimura M (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ Press, Cambridge, UK).
10. Margulies M, *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380.
11. Lynch M (2006) The origins of eukaryotic gene structure. *Mol Biol Evol* 23:450–468.
12. Lang GI, Murray AW (2008) Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics* 178:67–82.
13. Dickinson WJ (2008) Synergistic fitness interactions and a high frequency of beneficial changes among mutations accumulated under relaxed selection in *Saccharomyces cerevisiae*. *Genetics* 178:1571–1578.
14. Maki H (2002) Origins of spontaneous mutations: Specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Ann Rev Genet* 36:279–303.
15. Friedberg EC, *et al.* (2006) *DNA Repair and Mutagenesis* (ASM Press, Washington, DC).
16. Birdsell JA (2002) Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Mol Biol Evol* 19:1181–1197.
17. Marais G (2003) Biased gene conversion: implications for genome and sex evolution. *Trends Genet* 19:330–338.
18. Lynch M (2007) *The Origins of Genomic Architecture* (Sinauer, Sunderland, MA).
19. Wierdl M, Dominska M, Petes TD (1997) Microsatellite instability in yeast: Dependence on the length of the microsatellite. *Genetics* 146:769–779.
20. Ellegren H (2000) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat Genet* 24:400–402.
21. Vigouroux Y, *et al.* (2002) Rate and pattern of mutation at microsatellite loci in maize. *Mol Biol Evol* 19:1251–1260.
22. Dettman JR, Taylor JW (2004) Mutation and evolution of microsatellite loci in *Neurospora*. *Genetics* 168:1231–1248.
23. Seyfert AL, *et al.* (2008) The rate and spectrum of microsatellite mutation in *Caenorhabditis elegans* and *Daphnia pulex*. *Genetics* 178:2113–2121.
24. Hawk JD, Stefanovic L, Boyer JC, Petes TD, Farber RA (2005) Variation in efficiency of DNA mismatch repair at different sites in the yeast genome. *Proc Natl Acad Sci USA* 102:8639–8643.
25. Henderson ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Mol Cell Biol* 12:2749–2757.
26. Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365:274–276.
27. Johnson RE, Kovvali GK, Prakash L, Prakash S (1996) Requirement of the yeast *MSH3* and *MSH6* genes for MSH2-dependent genomic stability. *J Biol Chem* 271:7285–7288.
28. Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD (1997) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol Cell Biol* 17:2851–2858.
29. Gragg H, Harfe BD, Jinks-Robertson S (2002) Base composition of mononucleotide runs affects DNA polymerase slippage and removal of frameshift intermediates by mismatch repair in *Saccharomyces cerevisiae*. *Mol Cell Biol* 22:8756–8762.
30. Yamada NA, Parker JM, Farber RA (2003) Mutation frequency analysis of mononucleotide and dinucleotide repeats after oxidative stress. *Environ Mol Mutagen* 42:75–84.
31. Legendre M, Pochet N, Pak T, Verstrepen KJ (2007) Sequence-based estimation of minisatellite and microsatellite repeat variability. *Genome Res* 17:1787–1796.
32. Tran HT, Keen JD, Kricker M, Resnick MA, Gordenin DA (1997) Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol Cell Biol* 17:2859–2865.
33. Kupiec M, Petes TD (1988) Allelic and ectopic recombination between Ty elements in yeast. *Genetics* 119:549–559.
34. Rachidi N, Barre P, Blondin B (1999) Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of *Saccharomyces cerevisiae*. *Mol Gen Genet* 261:841–850.
35. Gerstein AC, Chun HJ, Grant A, Otto SP (2006) Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. *PLoS Genet* 2:e145.
36. Hiraoka M, Watanabe K, Umezu K, Maki H (2000) Spontaneous loss of heterozygosity in diploid *Saccharomyces cerevisiae* cells. *Genetics* 156:1531–1548.
37. Dunham MJ, *et al.* (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 99:16144–16149.
38. Putnam CD, Pennaneach V, Kolodner RD (2005) *Saccharomyces cerevisiae* as a model system to define the chromosomal instability phenotype. *Mol Cell Biol* 25:7226–7238.
39. Hughes TR, *et al.* (2000) Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat Genet* 25:333–337.
40. Schacherer J, *et al.* (2007) Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS ONE* 2(3):e322.
41. Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL (2007) Genetic properties influencing the evolvability of gene expression. *Science* 317:118–121.
42. Wloch DM, Szafraniec K, Borts RH, Korona R (2001) Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics* 159:441–452.
43. Zeyl C, DeVisser JA (2001) Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics* 157:53–61.
44. Joseph SB, Hall DW (2004) Spontaneous mutations in diploid *Saccharomyces cerevisiae*: More beneficial than expected. *Genetics* 168:1817–1825.
45. Vassilieva L, Hook AM, Lynch M (2000) The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution (Lawrence, Kans)* 54:1234–1246.
46. Baer CF, *et al.* (2006) Cumulative effects of spontaneous mutations for fitness in *Caenorhabditis*: Role of genotype, environment and stress. *Genetics* 174:1387–1395.
47. Fry JD (2001) Rapid mutational declines of viability in *Drosophila*. *Genet Res* 77:53–60.
48. Fry JD (2004) On the rate and linearity of viability declines in *Drosophila* mutation-accumulation experiments: Genomic mutation rates and synergistic epistasis revisited. *Genetics* 166:797–806.
49. Kunz BA, Ramachandran K, Vonarx EJ (1998) DNA sequence analysis of spontaneous mutagenesis in *Saccharomyces cerevisiae*. *Genetics* 148:1491–1505.
50. Ohnishi G, *et al.* (2004) Spontaneous mutagenesis in haploid and diploid *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun* 325:928–933.
51. Wilkins AS (1992) *Genetic Analysis of Animal Development* (Wiley–Liss, New York), 2nd ed.
52. Schug MD, *et al.* (1998) The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Mol Biol Evol* 15:1751–1760.
53. Schlötterer C, Ritter R, Harr B, Brem G (1998) High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Mol Biol Evol* 15:1269–1274.
54. Vázquez JF, Pérez T, Albornoz J, Domínguez A (2000) Estimation of microsatellite mutation rates in *Drosophila melanogaster*. *Genet Res* 76:323–326.
55. Lynch M, Koskella B, Schaack S (2006) Mutation pressure and the evolution of organelle genome architecture. *Science* 311:1727–1730.
56. Drost JB, Lee WR (1995) Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among *Drosophila*, mouse, and human. *Environ Mol Mut* Suppl 26:48–64.
57. 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.
58. 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.
59. Kondrashov AS (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum Mutat* 21:12–27.
60. Thatcher JW, Shaw JM, Dickinson WJ (1998) Marginal fitness contributions of nonessential genes in yeast. *Proc Natl Acad Sci USA* 95:253–257.