Drift-barrier hypothesis and mutation-rate evolution

Way Sung, Matthew S. Ackerman, Samuel F. Miller, Thomas G. Doak, and Michael Lynch[1]

Department of Biology, Indiana University, Bloomington, IN 47401

Contributed by Michael Lynch, September 21, 2012 (sent for review September 9, 2012)

Mutation dictates the tempo and mode of evolution, and like all traits, the mutation rate is subject to evolutionary modification. Here, we report refined estimates of the mutation rate for a prokaryote with an exceptionally small genome and for a unicellular eukaryote with a large genome. Combined with prior results, these estimates provide the basis for a potentially unifying explanation for the wide range in mutation rates that exists among organisms. Natural selection appears to reduce the mutation rate of a species to a level that scales negatively with both the effective population size (N_e), which imposes a drift barrier to the evolution of molecular refinements, and the genomic content of coding DNA, which is proportional to the target size for deleterious mutations. As a consequence of an expansion in genome size, some microbial eukaryotes with large N_e appear to have evolved mutation rates that are lower than those known to occur in prokaryotes, but multicellular eukaryotes have experienced elevations in the genome-wide deleterious mutation rate because of substantial reductions in N_e.

random genetic drift | replication fidelity

Mutation is the ultimate source of variation for all evolutionary processes, but like all other traits, the mutation rate itself is subject to evolutionary modification. Unfortunately, because the fidelity of DNA replication and repair is typically very high, mutation-rate estimation is a laborious process, and few comprehensive studies have been performed. However, three phylogenetically general patterns have been suggested. First, for nearly every taxon for which mutations have been cataloged, there is an elevated rate of mutation from G/C to A/T bases relative to the opposite direction (1–3), the only exceptions being derived from indirect polymorphism studies in a few high-GC prokaryotes (3). Second, there is a strong relationship between the mutation rate per nucleotide site per generation (u) and total genome size (4), although the direction of scaling differs dramatically between microbes and multicellular eukaryotes. Third, there appears to be an overall deletion bias in prokaryotes (5, 6), but an overall insertion bias in most eukaryotes because of a predominance of mobile-element activity (7).

Summarizing all studies up to 1990, Drake (8) concluded that u varies inversely with genome size (G) in microbes, such that the total genome-wide mutation rate (the product uG) is an approximate constant 0.003 across taxa. This pattern was derived from data on just three microbes (the bacterium *Escherichia coli*, the budding yeast *Saccharomyces cerevisiae*, and the filamentous fungus *Neurospora crassa*) and three bacteriophage. Subsequent observations continue to uphold the inverse relationship postulated by "Drake's rule" for prokaryotes and DNA viruses, but because of the narrow range of prokaryotic genome sizes, the significance remains borderline unless bacteriophage are included (4).

In contrast, when such an analysis is restricted to eukaryotes (ranging from yeast to invertebrates to human), there is a strong positive scaling of u with genome size (4). Because the two scalings intersect at ∼10 Mb (roughly the upper limit to genome sizes in prokaryotes, and roughly the lower limit for free-living eukaryotes), and because the one unicellular eukaryote with a refined measure of u, *S. cerevisiae*, has a genome size (12.05 Mb) falling near the intersection, it has remained unclear whether Drake's rule strictly applies to bacteriophage and prokaryotes or extends to at least some microbial eukaryotes. Resolution of this issue requires mutation-rate estimates for unicellular eukaryotes with much larger genomes than yeast, and for prokaryotes with much smaller genomes than *E. coli*.

As a first step in this direction, we have pursued a mutation-accumulation (MA) strategy (9, 10), followed by complete genome sequencing of the derived lines (11–13), to estimate the features of spontaneous mutations in the unicellular green alga *Chlamydomonas reinhardtii.* The 121-Mb nuclear genome size of this species (14), is nearly identical to that for the only land plant, *Arabidopsis thaliana* (125 Mb), for which a direct estimate of the mutation rate is available (15). The same strategy was used to estimate the mutational features of the bacterium *Mesoplasma florum*, whose diminutive genome (0.79 Mb) places it at the lower limit for culturable microorganisms.

Results

Mutation Rates in Unicellular Species with Extreme Genome Sizes. Whole-genome sequencing of four MA lines of *C. reinhardtii* after an average of 1,730 cell divisions yielded a base-substitutional mutation-rate estimate of $6.76\ (\mathrm{SE} = 1.08) \times 10^{-11}$ per site per cell division, substantially lower than that observed for any other eukaryote except the ciliate *Paramecium tetraurelia*, $0.64\ (0.36) \times 10^{-11}$ (16), and also well below the rates estimated for almost all prokaryotes (Fig. 1*A*). The *Chlamydomonas* mutation rate per cell division is ∼90-times smaller than the per-generation rate for *Arabidopsis* (15); it is also lower than the per-cell division rate in the latter, given that there are ∼40 cell divisions in a seed-to-seed interval in this species (17). On the other hand, the sequencing of 30 *M. florum* lines after 2,000 cell divisions yields the highest known rate of base-substitutional mutation for any unicellular organism, $9.78\ (0.71) \times 10^{-9}$ per site per cell division.

Because of the unexpectedly low mutation rate in *C. reinhardtii*, the harvest of mutations was small. However, as in all prior MA experiments with bottlenecked lines, and consistent with theoretical expectations for experiments of this nature (18), there is no evidence that mutations were systematically eradicated by selection before analysis. Twenty of the *Chlamydomonas* mutations were base substitutions, and 13 were small (< 13 bp) insertion/deletions (Tables S1, S2, S3, and S4). Four of the substitution mutations were exonic, 5 intronic, and 11 intergenic, a distribution that is not significantly different from the null expectation of 3:7:10 based on the genome composition of this species (χ^2 test; $df = 2$, $P = 0.55$). Although all of the small indels occurred in introns and intergenic DNA, this pattern was also not significantly different from the null expectation. For the *Chlamydomonas* base-substitutional mutations, nine were from G/C to A/T, and five were in the opposite direction which, because of the small sample size, is not a significant difference but is still consistent with the mutation bias in the direction of AT seen in all prior studies (1–3).

Author contributions: W.S., M.S.A., S.F.M., T.G.D., and M.L. designed research; W.S., M.S.A., S.F.M., T.G.D., and M.L. performed research; W.S., M.S.A., and M.L. analyzed data; and W.S., M.S.A., S.F.M., and M.L. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: milynch@indiana.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1216223109/-/DCSupplemental.

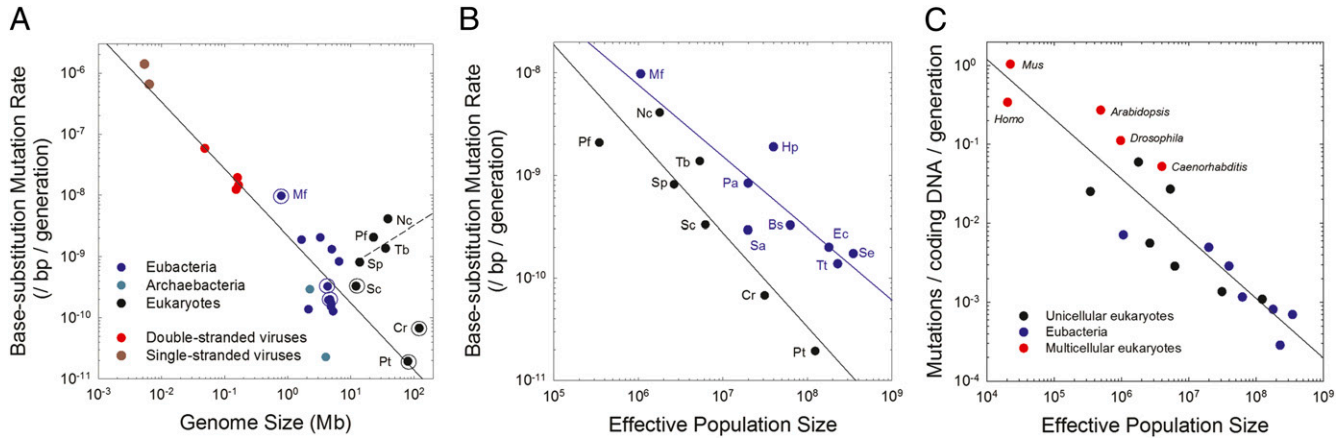

Fig. 1. (*A*) Relationship between the base-substitutional mutation rate/site/cell division and genome size. The regression includes all points except the four uppermost eukaryotes, for which the mutation-rate estimates are based on reporter constructs, $\log_{10}u = -8.663–1.096\log_{10}G$, where u is the mutation rate, and G is the genome size in megabases ($r^2 = 0.872$, $df = 21$). Points surrounded by a circle are based on mutation-accumulation experiments involving whole-genome sequencing; all others are based on reporter constructs. For euykaryotes: Cr, *Chlamydomonas reinhardtii*; Nc, *Neurospora crassa*; Pf, *Plasmodium falciparum*; Pt, *Paramecium tetraurelia*; Sc, *Saccharomyces cerevisiae*; Sp, *Schizosaccharomyces pombe*; Tb, *Trypanosoma brucei*. The prokaryote reported in this study, *Mesoplasma florum*, is denoted as Mf. The dashed regression line to the lower right includes multicellular eukaryotes (not shown) (4). (*B*) Relationship between the base-substitutional mutation rate/site/cell division and the effective population size (N_e) extrapolated from silent-site diversity. Eukaryotic regression (black): $\log_{10}u = -3.145–0.916\log_{10}N_e$ ($r^2 = 0.831$); prokaryotic regression (blue): $\log_{10}u = -3.920–0.699\log_{10}N_e$ ($r^2 = 0.794$). Labeled prokaryotic data points: Bs, *Bacillus subtilis*; Ec, *Escherichia coli*; Hp, *Helicobacter pylori*; Mt, *Mycobacterium tuberculosis*; Pa, *Pseudomonas aeruginosa*; Sa, *Sulfolobus acidocaldarius* (archaea); Se, *Salmonella enterica*; Tt, *Thermus thermophila*. (*C*) Relationship between genome-wide mutation rate/cell division for coding DNA and N_e. Regression: $\log_{10}(uG_e) = 3.109–0.757\log_{10}N_e$ ($r^2 = 0.844$). The data for multicellular eukaryotes (red) are summarized in Tables S8, S9, and S10, which are slight updates from the data previously summarized (4).

The *Mesoplasma* experiment yielded 628 de novo mutations, and hence a much more confident description of the mutation spectrum, which consists of 527 base substitutions and 101 small insertion/deletions (Tables S1, S5, S6, and S7). The mutation rate from G/C sites to A/T was 17.4-times that in the opposite direction. This is the most extreme mutational disparity recorded in any species, and would lead to a predicted equilibrium genome composition of 95% A/T if mutation pressure were the only determinant. The actual A/T composition of the *Mesoplasma* genome is extraordinarily high (73%), but this requires only ~2.7-times inflation in the net flux toward fixed A/T vs. G/C nucleotides, [2.7/(2.7 + 1.0)], so it is clear that there is substantial selective opposition to the accumulation of A/T bases in nature. Again, there is no compelling evidence that mutation accumulation in the experimental lines was opposed by selection; the ratio of base substitutions in synonymous and nonsynonymous sites is 70:417, which is not significantly different from the null expectation of 0.15:1 when the MA-derived mutation spectrum is applied to the codon use in the genome of this species (χ^2 test; $df = 1$, $P = 0.55$). The observed mutations are slightly, but not significantly, biased toward coding regions (χ^2 test; $df = 1, P = 0.08$).

Revisiting Drake's Rule. Combining these new mutation-rate estimates with previous observations demonstrates that most, but not all, microorganisms with estimated mutation rates obey the inverse scaling postulated by Drake (8) (Fig. 1*A*). On the other hand, the data presented here are inconsistent with the 0.003 mutations/genome/generation expected under Drake's rule, as the respective estimates for *Chlamydomonas* and *Mesoplasma* are 0.0082 (0.0013) and 0.0077 (0.0006) for base substitutions alone (with only minor additional contributions from insertion/deletions). Nonetheless, the overall regression suggests average genome-wide mutation rates in the range of 0.001–0.003 for the taxa involved in the range of $G = 0.01–100.00$ Mb.

Comparative analyses of natural isolates imply very high mutation rates on an absolute time scale in two other prokaryotes with small genome sizes, *Buchnera aphidicola* (19) and *Mycoplasma gallisepticum* (20). Because of uncertainties in the numbers of cell divisions per year in natural microbial populations and potential issues of selection, we have elected not to include these studies in our analyses, although it appears that the results would be qualitatively consistent with the pattern in Fig. 1*A*.

The four eukaryotic outliers not included in the regression in Fig. 1*A* involve fairly crude mutation-rate estimates based on a small number of reporter constructs. However, these estimates are also continuous with the apparent V-shaped pattern around $G = 10$ Mb, noted above, which extends to a broad range of metazoans and the land plant *A. thaliana* (4). Although the only two base-substitution mutation rate estimates for Archaebacteria fall below the general regression, and these are again based on reporter constructs, the limited data suggest that these taxa have mutation spectra dominated by insertion/deletions (21, 22), which when included would bring them in greater accord with the overall trend.

The mechanisms responsible for the discontinuity in scaling of the mutation rate with genome size remain unclear. There is no evidence that genome size has a direct, causal influence on the mutation rate. However, the drift-barrier hypothesis predicts that the level of refinement of molecular attributes, including DNA replication fidelity and repair, that can be accomplished by natural selection will be negatively correlated with the effective population size (N_e) of a species (4, 23). Under this hypothesis, as natural selection pushes a trait toward perfection, further improvements are expected to have diminishing fitness advantages. Once the point is reached beyond which the effects of subsequent beneficial mutations are unlikely to be large enough to overcome the power of random genetic drift, adaptive progress is expected to come to a standstill. Because selection is generally expected to favor lower mutation rates as a result of the associated load of deleterious mutations (24–27), and because the power of drift is inversely proportional to N_e, lower mutation rates are expected in species with larger N_e.

A qualitative test of this hypothesis is possible for the subset of species for which estimates of both u and standing levels of variation at neutral nucleotide sites are available, with the latter providing an indirect basis for estimating N_e. At mutation-drift equilibrium, the average nucleotide heterozygosity at silent sites

(π_s) has an expected value equal to $\sim 4N_e u$ for diploids and $2N_e u$ for haploids (28), so the ratio $\pi_s/(ku)$, where $k = 4$ or 2, provides an estimate of N_e. For the species for which such estimates can be made (Tables S8, S9, and S10), there is a strong negative relationship between u and N_e, although the regression for prokaryotes is elevated above that for eukaryotes (Fig. 1*B*). The probabilities that all prokaryotic datapoints will reside above the eukaryotic regression by chance, and vice versa, are 0.0039 and 0.0078, respectively. This finding suggests that for a given N_e, selection is less effective at reducing the per-nucleotide mutation rate in prokaryotes than in eukaryotes, despite the fact that such selection is expected to be substantially stronger in non-recombining genomes (24).

One potential issue with respect to the analyses in Fig. 1*B* is that the estimation of N_e involves the division of estimates of π_s by estimates of u. As a consequence of sampling variance in the estimates of u, such a procedure can induce some negative correlation between the estimates of N_e and u. For example, if there was no true interspecific variation in π_s, sampling variance in u would cause an inverse relationship between estimates of u and N_e. Two observations indicate that the general patterns outlined in Fig. 1*B* are not a consequence of such a statistical artifact. First, there is essentially no relationship between u and π_s in microbial species (Fig. S1*A*). Because these two measures are independent statistically, given that π_s is expected to be proportional to $N_e u$, this absence of correlation alone is consistent with increases in u being accompanied by decreases in N_e. Second, the joint distributions of estimates of u and N_e based on the sampling variances of u and π_s within species are far too narrow to account for the among-species patterns noted above (Fig. S1*B*). Thus, we conclude that the inverse scaling between u and N_e is indeed an evolved biological feature.

A plausible explanation for the different intercepts for the prokaryotic and eukaryotic regressions in Fig. 1*B* is that the magnitude of selection operating to reduce the mutation rate is not simply a function of the per-site mutation rate but of the genome-wide deleterious mutation rate (24–27), with prokaryotic genomes imposing a smaller target for the origin of such mutations than eukaryotic genomes. Most prokaryotic genomes contain <5,000 protein-coding genes, whereas most eukaryotic genomes encode for 7,000–30,000, with the much smaller amounts of DNA associated with tRNAs, rRNAs, and so forth, scaling accordingly (7).

As selection operates on the total number of deleterious mutations incurred by a genome, the magnitude of selection operating to reduce the mutation rate per nucleotide site is expected to scale positively with the effective size of a genome (G_e), which we approximate as the sum of protein-coding DNA. Consistent with this hypothesis, the effective genome-wide mutation rate (uG_e) scales inversely with N_e in a way that is largely independent of phylogenetic background, eliminating the distinct eukaryotic and prokaryotic patterns seen for u alone (Fig. 1*C*). Moreover, this general relationship appears to extend beyond microbes, as the genome-wide mutation rate continues to increase as N_e declines to lower levels in multicellular eukaryotes. This fit, which accounts for 84% of the variation in uG_e across taxa, is all the more striking when one considers the intrinsic difficulties in obtaining accurate estimates of both u and N_e, and the fact that the latter is expected to vary somewhat over evolutionary time.

Discussion

Our results provide a potentially unifying explanation for the wide range of variation in mutation rates observed across the tree of life. First, the existence of extremely low mutation rates in some microbial eukaryotes, even compared with prokaryotes, appears to be consistent with the drift-barrier hypothesis in that some unicellular eukaryotes appear to have effective population sizes approaching the highest levels observed in prokaryotes, while harboring a larger number of nucleotides in the genome subject to selection. Second, although Drake's rule for a constant number of mutations per genome per cell division may hold for a subset of microbial species, this generalization does not apply to all taxa. Among unicellular species alone, the total number of base substitutions in coding DNA per cell division ranges from $\sim 10^{-4}$ to $\sim 10^{-1}$, and for vertebrates approaches 1.0 on a generational time scale (Fig. 1*C*). This wide range of variation becomes understandable only after the joint distribution of N_e and G_e is accounted for. Finally, the results in Fig. 1*B* provide a simple explanation for Lewontin's (29) "paradox of variation": the weak response of standing levels of molecular variation across species with a wide range of population sizes. As suggested previously (4, 7), and demonstrated here in further detail (Fig. 1*B*), this pattern is at least in part a simple consequence of the evolved, negative relationship between u and N_e.

Could it be that rather than N_e driving the evolution of the mutation rate, the direction of causality is reversed? The recurrent introduction of deleterious mutations results in selection on linked sites across the genome (background selection), inducing variation in family sizes that in turn causes a reduction in N_e (30). However, a substantial body of theory predicts a linear decline of $\log(N_e)$ with the genome-wide deleterious mutation rate under this background-selection model (30), which is at variance with the strong log-log relationship that is actually observed (Fig. 1*C*). Thus, although there seems little doubt that N_e can be influenced by background selection, the broad phylogenetic patterns suggested here imply that other, presumably ecological, factors are the primary determinants of N_e.

There are some caveats with respect to the preceding analyses. First, although the general scaling of u with N_e and G_e is qualitatively consistent with existing evolutionary genetic theory, an explanation for the magnitude of u remains to be developed. One potential issue is that mutation-rate estimates are almost always obtained under optimal laboratory growth conditions, raising questions as to whether harsher environmental conditions would result in lower (or higher) levels of replication fidelity. Experimental work in this area has led to mixed results (31–33), and there is considerable room for further research.

Second, as discussed further in the *SI Text*, the estimation of species-wide N_e is fraught with difficulties. On the one hand, because a regression coefficient is inversely proportional to the variance of the explanatory variable, sampling error in N_e will cause the regressions in Fig. 1 *B* and *C* to be flatter than the true values. On the other hand, a tendency to underestimate N_e to a greater degree in high-N_e taxa could bias the slopes in Fig. 1 *B* and *C* toward more negative values. Because silent sites are known to be under weak selection for translational accuracy, with the presumed efficiency of such selection increasing with N_e (7, 34), the estimates of N_e used here are likely downwardly biased for the highest-N_e taxa, in which case the true slope of the regression in Fig. 1*C* could be less extreme than depicted. Others have argued that boom-and-bust cycles may result in species having phases of short-term of N_e that exceed longer-term values reflected in heterozygosity data (35). Although such conditions would enhance the efficiency of selection during periods of peak densities, this would not alter the scaling that we have presented unless the magnitudes of population-density fluctuations systematically varied with standing levels of heterozygosity. In addition, we acknowledge that criteria for species delineations can be arbitrary, and we have simply adhered to the identifications made by those previously involved in species-variation studies. Culling particular genotypes based on absolute sequence divergence is arbitrary and circular; the existence of deep branches in a gene genealogy does not provide a compelling criterion for species delineation, as the expected final branch depth in even a neutral coalescent is equal half the total depth of the genealogy (36), and most species show significant levels of population subdivision resulting from isolation by distance. Despite all of these issues, it

is difficult to see how their cumulative effects would be sufficient to create a significantly negative association between u and N_e out of a lack of (or even positive) relationship.

Finally, it is well-known that the magnitude of selection on the mutation rate is a function of the recombination rate, and in recombining species on the deleterious effects of mutations as well (24–27). Although only a subset of the species relied upon in this study regularly engage in meiotic sexual reproduction, it also appears that despite the absence of meiosis, the average rate of recombination (scaled to the mutation rate) in prokaryotic species is not greatly different from that in eukaryotes (7), so it should perhaps not be a great surprise that the data from these two domains are comingled in Fig. 1*C*. There is the additional issue that our analyses also derive from a mixture of haploid and diploid species, although the limited data suggest that the average fitness effects of mutations are not greatly different in these two contexts (9, 18, 37).

In summary, if the drift-barrier hypothesis for the evolution of mutation rates is correct, we will have arrived at a fairly general explanation for the evolution of a trait that not only encompasses all of life, but that in turn dictates the pace and mechanisms of evolutionary change that are possible in different phylogenetic lineages. Moreover, should microbial eukaryotes exist with even larger G_e and N_e than those included in this study, our results predict that even more refined mechanisms of replication fidelity/DNA repair than exhibited by *Chlamydomonas* and *Paramecium* await discovery.

Materials and Methods

Line Development. Seventy-five independent *M. florum* L1 (ATCC #33453) MA lines were initiated from a single colony. Lines were grown on 100 × 15-mm Petri dishes containing *Mycoplasma* medium (heart infusion broth, horse serum, yeast extract, sucrose, agar). Every 4 d, for ~2,300 cell divisions, a single isolated colony from each MA line was transferred by streaking to a new plate, ensuring that each line regularly passed through a single-cell bottleneck (17).

One-hundred lines of *C. reinhardtii* (CC-124 obtained from the *Chlamydomonas* Resource Center, University of Minnesota) were initiated from a single vegetative (haploid) cell. Lines were grown mixotrophically on 100 × 20-mm Petri dishes containing enriched YA medium. To minimize gliding or dispersal of colonies, 2% (wt/vol) agar was used in the dishes. Strains were propagated from a single well-isolated colonies every week for ~1,730 cell divisions. The bottlenecking procedures used both experiments ensure that mutations accumulate in an effectively neutral fashion (17).

To estimate *Mesoplasma* generation times, entire colonies were transferred to PBS saline buffer, vortexed, serially diluted, and replated. Cell densities calculated from viable cell counts imply an average of 21 (SE = 0.13) divisions every 4 d over the entire experiment. To estimate the *Chlamydomonas* generation time, entire colonies were removed from the agar plate with a glass pipette and suspended in 100 μL of mild detergent (0.5% nonidet P40). Cells were agitated with glass beads for 2 min to homogenize the suspension, and hemocytometer counts yielded an estimate of ~500,000 cells/mm^2 per colony. The diameters of 500 colonies measured after 7 d of growth, implied an average 15 (SE = 0.2) generations per week over all 100 *Chlamydomonas* MA lines.

DNA Extraction, Sequencing, and Assembly. DNA was extracted from 30 random *Mesoplasma* lines using the Wizard extraction kit (Promega), and DNA extraction of four random *Chlamydomonas* lines was completed by lysing the cells in SDS-EB buffer [2% (wt/vol) SDS, 400 mM NaCL, 40 mM EDTA, 100 mM Tris•HCL, pH 8.0] followed by phenol/chloroform extractions to Illumina library standards.

Using 101-bp paired-end Illumina (Hi-Seq platform) sequencing, each *Mesoplasma* line was sequenced to an average 100× coverage depth, and each *Chlamydomonas* line to ~15 times. The average library fragment size (distance between paired ends) was ~175 bp.

The paired-end reads for each *Mesoplasma* line were individually mapped against the *M. florum L1* reference genome (assembly and annotation available from the Broad Institute) using two separate alignment algorithms: BWA (38) and NOVOALIGN (available at www.novocraft.com). The paired-end reads for each *Chlamydomonas* line were mapped against the *C. reinhardtii* (rev4) reference genome (14) using NOVOALIGN. The resulting pile-up files were converted to SAM format using SAMTOOLS (39). Using in-house Perl and Python scripts, the alignment information was further parsed to generate forward and reverse mapping information at each site, resulting in a configuration of eight numbers for each line (A, a, C, c, G, g, T, t), corresponding to the number of reads mapped at each genomic position in the reference sequence. To identify small- and medium-sized deletion/insertion events (>11 bp), we implemented the pattern growth alignment algorithm PINDEL (40).

Data Processing. A consensus approach was used to identify mutations, comparing each individual line (focal line) against the consensus of all of the remaining lines. This approach is ideal with a large number of samples, and is robust against sequencing or alignment errors in the reference genome, with previous applications yielding very low false-positive rates (11, 15, 41). The consensus approach employs three steps in mutation identification: (*i*) At each nucleotide position, the consensus is identified for each line, requiring 80% of the reads in an individual line to indicate the same nucleotide (A|C|T|G), with at least two forward and two reverse reads. (*ii*) The overall consensus (ancestral) base call is identified by requiring 80% of the reads across all remaining lines to indicate the same nucleotide. (*iii*) If a line-specific consensus has a base call differing from the overall consensus, and at least two other lines contain enough reads to be used in the comparison, the site was designated as a mutation for the discordant line. Because of the limited sequencing coverage of the *Chlamydomonas* lines, this final criterion was modified to require the three other lines to collectively have a minimum of two forward and two reverse reads supporting the consensus nucleotide.

For the *Mesoplasma* lines, the consensus approach identified 540 base substitutions when applied to the BWA mapping output and 527 base substitutions when applied to the NOVOALIGN mapping output, with 527 of the base substitutions overlapping between the two algorithms (Table S6). Of the 13 discordant base substitutions, seven were false-positives directly adjacent to an insertion/deletion event and six were a cluster of mutations that could not be uniquely mapped to the reference genome with NOVOALIGN. These 13 base substitutions, a very minor fraction of the overall pool, were discarded. NOVOALIGN, BWA, and PINDEL identified a total of 101 small indels (Table S7), only one of which was not identified by all methods. This same methodology provides a false-positive rate of 0% when applied to an ongoing *Bacillus subtilis* MA project (involving ~100× sequencing coverage, and verification of 69 of 69 putative base substitutions and 29 of 29 small insertion/deletion events using conventional fluorescent sequencing technology applied to PCR-amplified products).

For the *Chlamydomonas* lines, the consensus approach identified 20 base substitutions and 13 small insertion/deletions (Tables S3 and S4). We designed primer sets to PCR amplify 300- to 500-bp regions surrounding all of the base substitutions and five of the indels, and each was confirmed by conventional sequencing, as was the implied ancestral nucleotide.

Mutation Rate Calculations. The base-substitution mutation rate (per nucleotide site per cell division) was calculated for each line as $u_{bs} = m/(nT)$, where m is the number of observed base-substitution mutations, n is the number of sites analyzed in the line, and T is the number of generations that occurred in the line. The SE for an individual line is calculated as SE = $[u_{bs}/(nT)]^{0.5}$ (41, 42), whereas the total SE of the base-substitution mutation rate is given by the SD of the mutation-rate estimates across all lines divided by the square root of the number of lines analyzed.

ACKNOWLEDGMENTS. We thank U. Zekonyte, S. Surzycki, and E. Williams for assistance with experiments. This work was supported by National Institutes of Health Grant R01 GM036827 (to M.L. and W.K. Thomas); National Science Foundation Grant EF-0827411 (to M.L.); and Multidisciplinary University Research Initiative Award W911NF-09-1-0444 from the US Army Research Office (to M.L., P. Foster, H. Tang, and S. Finkel).

1. Lynch M (2009) Rate, molecular spectrum, and consequences of spontaneous mutations in man. *Proc Natl Acad Sci USA* 107(3):961–968.
2. Hershberg R, Petrov DA (2010) Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet* 6(9):e1001115.
3. Hildebrand F, Meyer A, Eyre-Walker A (2010) Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet* 6(9):e1001107.
4. Lynch M (2010) Evolution of the mutation rate. *Trends Genet* 26(8):345–352.
5. Mira A, Ochman H, Moran NA (2001) Deletional bias and the evolution of bacterial genomes. *Trends Genet* 17(10):589–596.
6. Kuo CH, Ochman H (2009) Deletional bias across the three domains of life. *Genome Biol Evol* 1:145–152.
7. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer Associates, Sunderland).

8. Drake JW (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA* 88(16):7160–7164.
9. Lynch M, et al. (1999) Spontaneous deleterious mutation. *Evolution* 53:645–663.
10. Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: Causes and consequences. *Nat Rev Genet* 8(8):619–631.
11. Lynch M, et al. (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci USA* 105(27):9272–9277.
12. Keightley PD, et al. (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res* 19(7):1195–1201.
13. Denver DR, et al. (2012) Variation in base-substitution mutation in experimental and natural lineages of *Caenorhabditis* nematodes. *Genome Biol Evol* 4(4):513–522.
14. Merchant SS, et al. (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318(5848):245–250.
15. Ossowski S, et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327(5961):92–94.
16. Sung W, et al. (2012) Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. *Proc Natl Acad Sci USA,* 10.1073/pnas.1210663109.
17. Hoffman PD, Leonard JM, Lindberg GE, Bollmann SR, Hays JB (2004) Rapid accumulation of mutations during seed-to-seed propagation of mismatch-repair-defective *Arabidopsis*. *Genes Dev* 18(21):2676–2685.
18. Kibota TT, Lynch M (1996) Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* 381(6584):694–696.
19. Moran NA, McLaughlin HJ, Sorek R (2009) The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. *Science* 323(5912):379–382.
20. Delaney NF, et al. (2012) Ultrafast evolution and loss of CRISPRs following a host shift in a novel wildlife pathogen, *Mycoplasma gallisepticum*. *PLoS Genet* 8(2):e1002511.
21. Grogan DW, Carver GT, Drake JW (2001) Genetic fidelity under harsh conditions: Analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. *Proc Natl Acad Sci USA* 98(14):7928–7933.
22. Mackwan RR, Carver GT, Drake JW, Grogan DW (2007) An unusual pattern of spontaneous mutations recovered in the halophilic archaeon *Haloferax volcanii*. *Genetics* 176(1):697–702.
23. Lynch M (2011) The lower bound to the evolution of mutation rates. *Genome Biol Evol* 3:1107–1118.
24. Kimura M (1967) On the evolutionary adjustment of spontaneous mutation rates. *Genet Res* 9:23–34.
25. Kondrashov AS (1995) Modifiers of mutation-selection balance: General approach and the evolution of mutation rates. *Genet Res* 66:53–70.
26. Dawson KJ (1999) The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. *Theor Popul Biol* 55 (1):1–22.
27. Lynch M (2008) The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* 180(2):933–943.
28. Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet Res* 11(3):247–269.
29. Lewontin RC (1974) *The Genetic Basis of Evolutionary Change* (Columbia Univ Press, New York).
30. Charlesworth B (2012) The effects of deleterious mutations on evolution at linked sites. *Genetics* 190(1):5–22.
31. Joyner-Matos J, Bean LC, Richardson HL, Sammeli T, Baer CF (2011) No evidence of elevated germline mutation accumulation under oxidative stress in *Caenorhabditis elegans*. *Genetics* 189(4):1439–1447.
32. Matsuba C, Ostrow DG, Salomon MP, Tolani A, Baer CF (2012) Temperature, stress and spontaneous mutation in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. *Biol Lett*, in press.
33. Wang AD, Agrawal AF (2012) DNA repair pathway choice is influenced by the health of *Drosophila melanogaster*. *Genetics*, in press.
34. Raghavan R, Kelkar YD, Ochman H (2012) A selective force favoring increased G+C content in bacterial genes. *Proc Natl Acad Sci USA* 109(36):14504–14507.
35. Karasov T, Messer PW, Petrov DA (2010) Evidence that adaptation in *Drosophila* is not limited by mutation at single sites. *PLoS Genet* 6(6):e1000924.
36. Kingman JFC (1982) The coalescent. *Stoch Proc Appl* 13:235–248.
37. Andersson DI, Hughes D (1996) Muller's ratchet decreases fitness of a DNA-based microbe. *Proc Natl Acad Sci USA* 93(2):906–907.
38. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.
39. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
40. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z (2009) Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25(21):2865–2871.
41. Denver DR, et al. (2009) A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc Natl Acad Sci USA* 106(38):16310–16314.
42. Denver DR, Morris K, Lynch M, Thomas WK (2004) High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* 430 (7000):679–682.