There have been many attempts to measure the genome-wide mutation rate for spontaneous mutations, using measurements of traits in inbred lines in which mutations have accumulated. However, these are likely to miss many small-effect mutations that are important for evolutionary processes. Recently, the genome-wide spontaneous mutation rate in inbred lines of Caenorhabditis elegans was estimated, using DNA sequencing. The results imply that the mutation rate is surprisingly high, and that insertion-deletion mutations are unexpectedly common. Phenotypic assays of the same lines detected only a small proportion of mutations that were predicted to have evolutionarily significant fitness effects.

Introduction
The greatest gap in our knowledge of the causes of evolution is the frequency of spontaneous mutations. Although spontaneous mutations are the primary source of all variation, new mutations are so infrequent that measurements of mutation rates at the molecular level are extremely difficult. Conversely, it has proved difficult to infer the distribution of mutational effects for fitness and other quantitative traits because mutations with large effects tend to dominate phenotypic changes. However, mutations with small effects below the experimental detection limit are likely to have the most influence on the evolution of phenomena such as ageing, outbreeding and genetic recombination.

For forty years, mutation accumulation (MA) experiments have been the basis of our knowledge of genome-wide rates of spontaneous mutations and the sizes of their effects. The standard design for a MA experiment is to start with an isogenic line, subdivide the line into several independent MA lines, then maintain these by close inbreeding for several generations (e.g. full sib mating, selfing or as chromosomes sheltered from selection by balancers), during which time the sublines diverge as a result of the random accumulation of new mutations (Figure 1). Under close inbreeding conditions, all but the most strongly deleterious mutations behave as effectively neutral and accumulate randomly at a rate that is equal to the mutation rate.

It is important to distinguish between the estimates of mutation rates derived from changes in quantitative traits from those obtained by detecting changes at the DNA sequence level (Box 1). Many experiments measuring quantitative trait changes have been performed [1]. A trait-based estimate of the mean number of non-lethal mutations affecting a fitness trait that arise per generation \((U_d)\) can be obtained by comparing the rate of change of the mean phenotype with the rate of increase in the variance of the means of the MA lines [2]. This estimate is only equal to the mutation rate if all mutations have equal effects; if there is a wide distribution of effects, \(U_d\) can be greatly underestimated. An experiment that detected changes at the DNA level to estimate the mutation rate per nucleotide site in the eukaryotic nuclear genome was performed for the first time by Denver and colleagues [3].

Estimating the per nucleotide mutation rate in Caenorhabditis elegans
Two studies in Caenorhabditis elegans have attempted to measure the extent by which trait-based estimates of \(U_d\) in MA experiments are underestimates. Davies et al. [4] estimated the frequency of mutations induced in functionally important regions of the genome by the chemical mutagen ethyl methanesulphonate (EMS), and concluded that only \(\sim 4\%\) of these mutations were detectable on the basis of changes in mean and variance of fitness among mutagenised cell lines in the laboratory. Estes et al. [5] studied MA lines of a DNA repair-deficient C. elegans strain, and concluded that the number of mutations detectable from phenotypic changes was as little as \(1\%\) of the number of deleterious mutations, estimated using DNA sequencing. Thus, in C. elegans at least, most of the functionally significant EMS-induced mutations (mainly transition point mutations) and those generated by deficiency in mismatch repair (the detailed nature of these is currently unknown) appear to have ‘cryptic’ effects on fitness traits.

Obtaining direct estimates of the spontaneous mutation rate at the DNA level is more challenging because spontaneous mutations occur at a low rate at the nucleotide level. Previous attempts to estimate these rates in eukaryotes have focused on the appearance of new allozyme variants (‘band-morph mutations’) in large cohorts of human families [6], and in replicated inbred Drosophila melanogaster lines that accumulated spontaneous mutations at random (summarized in Ref. [1]). Estimates of point-mutation rates per base pair from these studies are broadly consistent with estimates based on the evolutionary divergence between closely related species at ‘silent’ sites in the genome [1].

Denver et al. [3] employed direct DNA sequencing of more than four million bases of DNA to study spontaneous mutations in \(> 50\) C. elegans MA lines of the N2 strain.

Peter D. Keightley and Brian Charlesworth
University of Edinburgh, Institute of Evolutionary Biology, School of Biological Sciences, West Mains Road, Edinburgh, UK EH9 3JT

Available online 15 December 2004

www.sciencedirect.com
that had undergone an average of 339 generations of MA. Previous analyses of fitness in these lines had detected only \( w_0.02 \) mutations per diploid genome per generation \([7]\). Even after 339 generations, the expected cumulative frequency of mutations at the DNA level is so low that sequencing with an accuracy of \( 10^{6} \) per base pair or more is required to obtain reliable estimates, both from the point of view of not missing mutations and declaring non-existent mutations. It is impressive that Denver et al. could achieve sufficient accuracy in their study by using an initial single sequencing pass on one strand, followed by verification of the sequences of clear-cut putative variants on the opposite strand and by reamplifying and resequencing a subset of the mutations that were detected. Denver et al. reported a total of 30 mutations in their study, of which more than half (17) were insertions or deletions (indels) and the remainder were single nucleotide mutations (SNMs). These numbers translate to mutation rates per site per generation of \( 1.2 \times 10^{-8} \) and \( 0.9 \times 10^{-8} \), respectively.

**Implications of the mutation rate estimates**

How do these estimates compare with less direct estimates of mutation rates per nucleotide site for other eukaryotes?

---

**Box 1. Estimating mutation rates in MA lines**

In a mutation accumulation (MA) experiment, mutations are allowed to accumulate in inbred lines in the virtual absence of natural selection (see Figure 1). Because each line is expected to accumulate different mutations, the lines will diverge both phenotypically and at the DNA level. Estimates of the mean fitness of the MA lines at generation \( t \) and of a mutation-free control population (i.e. the ancestral line that has, for example, been maintained in a cryopreserved state) can be used to infer the change of mean fitness per generation that is due to mutation accumulation, \( \Delta M \). A second parameter that can be estimated is the increase in genetic variance in fitness among lines per generation, \( V_{M} \). Under a model of equal mutational effects \([s]\), \( \Delta M = U_d s \), where \( U_d \) is the genomic rate for mutations affecting fitness per generation, and \( V_{M} = U_d s^2 \) \([1,2]\). \( U_d \) can therefore be estimated from \( U_d = \Delta M^2 / V_{M} \). However, variation among mutational effects leads this method to underestimate \( U_d \); the greater the variability the more serious the underestimation. A second method to obtain a molecular-based estimate of the genome-wide deleterious mutation rate \( (U_m) \) in a MA experiment is to count mutations at the DNA level directly by sequencing or by some other mutation-detection method, and to infer the fraction of the mutations detected that are deleterious. A minimum estimate of the fraction of deleterious mutations is the fraction of nucleotides in the genome that, if changed, would lead to an amino acid substitution in a protein-coding gene; it is known that the vast majority of amino acid substitutions are deleterious. There will be a further contribution to \( U_m \) from selectively constrained sites in non-coding DNA.

---

Figure 1. The design of a mutation accumulation (MA) experiment for *Caenorhabditis elegans*. The progeny of an inbred progenitor at generation 0 are used to found a series of independent MA sublines that are subsequently maintained on plates by the transfer of a randomly picked single hermaphrodite progeny in each generation. Mutations that arise in the lines become fixed at random, because the transfer of a single worm each generation implies that selection is ineffective. The mutant individuals, and their progeny, are represented in colour in subsequent generations. At generation \( t \), estimates of fitnesses of the sample of worms from the individual sublines can be used to obtain a phenotype-based estimate of the genome-wide deleterious mutation rate, \( U_d \) (Box 1). Denver et al. \([3]\) sequenced the DNA of samples from generation \( t \) to estimate the per nucleotide mutation rate, which can be used to produce a molecular-based estimate of \( U_m \) (Box 1).
The *C. elegans* SNM-rate estimate is approximately three times lower than an estimate for hominids based on the mean nucleotide divergence at pseudogenes between humans and chimpanzees (i.e. $\sim 2.5 \times 10^{-9}$ per generation [8]). In *Drosophila*, molecular divergence at synonymous sites suggests a mutation rate of $\sim 2 \times 10^{-9}$ [1], which is 4.5 times less than the *C. elegans* estimate. As mentioned previously, the estimates for hominids and *Drosophila* are in broad agreement with the estimates obtained from the accumulation of allozyme band morph variants. Unfortunately, there are no known close relatives of *C. elegans* that could enable indirect estimation of the mutation rate via molecular evolutionary divergence.

The high proportion of indel mutations observed by Denver et al. is also surprising because indel polymorphisms are about a quarter as frequent as single nucleotide polymorphisms in non-coding sequences in *C. elegans* populations [9]. One possible explanation is that some SNMs were missed by Denver and colleagues’ direct sequencing strategy (e.g. because indels are easier to identify in sequence traces than SNMs). A second possibility is that indels are more strongly selected against than SNMs; therefore, indels segregate in populations at lower frequencies than single nucleotide polymorphisms. This would particularly apply to species with large effective population sizes in which natural selection is most effective (Box 2). The population size of *C. elegans* is clearly large, but less is known about the effective population size ($N_e$) of breeding individuals. Low polymorphism rates among worldwide samples of *C. elegans* [10,11] seem to imply low $N_e$ values. Using the mean silent-site diversity for three nuclear loci reported by Graustein et al. [10] and Denver and colleagues’ estimate of the mutation rate for single nucleotides, the effective population size of *C. elegans*, is estimated to be only 15 600 if the diversity is equated to its expected equilibrium value [12]. This suggests that there is currently little scope for effective selection on weakly selected mutations.

By multiplying the per nucleotide mutation rate estimate by the number of nucleotides in the genome that lead to amino acid changes if mutated (which are almost always deleterious in natural conditions), Denver et al. estimated that the deleterious mutation rate per haploid genome ($U_m$) in protein coding genes is $\sim 0.5$, and that there must be a further contribution from mutations in functionally constrained non-coding DNA sites. The molecular mutation rate estimate is therefore nearly two orders of magnitude higher than the phenotypic estimates. The study of Denver et al., together with the two previous comparisons of phenotypic and molecular results, therefore, strongly suggests that the vast majority of new mutations that are deleterious have extremely small effects in laboratory conditions. How does this relate to data on genomic mutation rates and the distribution of mutational effects for fitness in other species? By far the greatest amount of work on measuring spontaneous mutational parameters in eukaryotes has been carried out in *Drosophila*, by either MA on chromosomes sheltered against balancers [2,13–16] or in lines maintained by full sib mating in which all but the most strongly deleterious mutations accumulate at random [17,18]. Estimates for $U_d$ based on phenotypic measures vary greatly between experiments. The highest estimates for $U_d$ approach one. If the worm results are extrapolated to flies, this implies that the ‘true’ value for $U_d$ in flies for all deleterious mutations, irrespective of their phenotypic effects, lies somewhere between ten and 100. These values appear implausibly high, however, given the information we have on band morph mutation rates in *Drosophila* and rates of molecular divergence at ‘silent’ sites between species.

Can the fly and worm results be reconciled? One possibility is that the great variation among phenotypic estimates for $U_d$ in flies is due to variation in the mutation rate, brought about by variability among genotypes in the rate of movement of transposable elements (TEs) [16,19]. Given that TE insertions in flies tend to have average fitness effects of the order of a few percent [16], their activity could account for the high $U_d$ values in some of the fly experiments, whereas the mutation rate caused by less drastic changes in the DNA could in fact be lower than in *C. elegans*. TE activity is essentially absent from the N2 strain of *C. elegans*.

The high frequency of indel mutations in *C. elegans* observed by Denver et al. would be expected to cause frequent lethal mutations, because most indels in a coding sequence would disrupt gene function. RNA interference (RNAi) experiments suggest that knockouts of at least 7% of *C. elegans* genes result in lethality, and RNAi detects a high proportion of known lethal mutations [20]. Combining this information with the proportion of the *C. elegans* genome that is coding sequence (27%) and the mutation rate for indels estimated by Denver et al., suggests that the lethal mutation rate per diploid genome from this source should be $\sim 4.5\%$. This figure is unexpectedly high because estimates of the spontaneous lethal mutation rate per diploid genome from chromosome balancer experiments in *C. elegans* are only $\sim 1\%$ [21,22]. If the lethal mutation rate is so high, it is puzzling that the rate of mutation for non-lethal deleterious alleles is so low [7] because the *Drosophila* experiments suggest that lethal mutations are a minority among deleterious mutations.

**Concluding remarks**

Finally, we note that there are several reasons why it is unlikely that the high mutation rate estimates at the DNA level for *C. elegans* imply that the N2 strain becomes genetically unstable as a consequence of deleterious mutation accumulation, as suggested by Rosenberg and Hastings [23] in their commentary that accompanied the

---

**Box 2. The effectiveness of natural selection against deleterious mutations depends on population size**

A measure of the selective disadvantage of a new deleterious mutation is its selection coefficient ($s$). However, the effectiveness of natural selection against such mutations depends on $N_es$, the product of the effective size of the population ($N_e$) in which the mutation segregates and $s$. The effective population size is related to the number of parents that contribute to the next generation, and can be substantially lower than the actual population size. As $N_es$ approaches zero, the fixation probability of a deleterious mutation approaches that of a neutral mutation (i.e. $1/2N_e$ in a diploid); mutations for which $N_es<1$ are often termed ‘effectively neutral’.
article by Denver et al. First, a genetically unstable state induced by MA would imply an accelerating rate of fitness decline, which was not observed [7]. Second, the mutator state would have to arise in a high proportion of the MA lines. Finally, the mean rates of phenotypic change for fitness traits in several independent sets of nematode MA lines, including a different strain of C. elegans and two other species, are consistent with the rates seen in earlier C. elegans MA experiments (C. Baer, unpublished).

In conclusion, the pioneering results of Denver et al. on the direct determination of mutation rates in the worm raise many questions as to how to reconcile different types of evidence concerning mutation rates and mutational effects on fitness in C. elegans, and how the mutation rate in C. elegans relates to that in other organisms with comparable genome sizes, such as Drosophila. Similar experiments on the fly might help to clarify some of these issues.

Acknowledgements
We thank Asher Cutter and David Houle for helpful comments.

References
13 Mukai, T. et al. (1972) Mutation rate and dominance of genes affecting viability in Drosophila melanogaster. Genetics 72, 335–355

Getting to know your neighbours; a new mechanism for cell intercalation

Kelly K. Nikolaidou1 and Kathy Barrett1,2,3

1Ludwig Institute for Cancer Research, 91 Riding House Street, London, UK W1W 7BS
2Department of Anatomy and Developmental Biology, University College London, Gower Street, London, UK WC1E 6BT
3Department of Biochemistry and Molecular Biology, University College London, Gower Street, London, UK WC1E 6BT

As an embryo gastrulates or makes neural tissue it shortens across the dorso-ventral axis and extends dramatically along the perpendicular, antero-posterior axis, resulting in the embryo doubling in length. This process is known as convergent extension and it is so powerful in remodelling tissue that it is used time and again during development. New research in Drosophila melanogaster and other model organisms is shedding fresh light on how it happens.