Quantifying The Mutational Process

Daniel L. Halligan

PhD

The University of Edinburgh

2004

To Dad, for inspiring me to study biology.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text.

This work has not been submitted for any other degree or professional qualification except as specified.

Daniel Halligan, September 30, 2004.

Acknowledgments

This work would not have been possible without the help and encouragement of a large number of people. I am very grateful to all of the following:

Peter Keightley for all the supervision, encouragement and useful comments on multiple manuscripts and thesis chapters. Brian Charlesworth for additional supervision, invaluable discussions and comments. Andy Peters for a great deal of help with lab work, statistical analysis and for comments on manuscripts. Peter Andolfatto for useful discussions and supervision during my final project. Penny Haddrill for reading all my thesis chapters and for help with lab work and DNA sequencing. Ben Gantenbein for additional help with sequencing. Ian White for statistical advice. Adam Eyre-Walker, Mike Whitlock, Dan Gaffney and Tim Sands for helpful discussions. Jenny Elrick, Fiona Oliver and Helen Cowan for help with lab work. Toby Johnson for the LaTeX 2_ε style file used to typeset this thesis and Julian Christians for useful comments on a manuscript.

Thanks to all my friends at Edinburgh, especially Dave, Tim, Aoife, Stuart, Dan and most of all Penny. Finally thanks to Mom and Dad for all their support and encouragement.

This work was supported by a Biotechnology and Biological Sciences Research Council postgraduate studentship number 00/B1/G/06147.

Publications

The following published papers have arisen from this thesis and they are included in Appendix B.

- Halligan D. L. & Keightley P. D. (2003). How many lethal alleles. *Trends Genet.* **19,** 57–59.
- Peters A. D., Halligan D. L., Whitlock M. C. & Keightley P. D. (2003). Dominance and overdominance of mildly deleterious induced mutations for fitness traits in *Canorhabditis elegans*. *Genetics* **165,** 589–599.
- Halligan D. L., Peters A. D. & Keightley P. D. (2003). Estimating numbers of EMS-induced mutations affecting life history traits in *Caenorhabditis elegans* in crosses between inbred sublines. *Genet. Res.* **82,** 191–205.
- Halligan D. L., Eyre-Walker A., Andolfatto P. & Keightley P. D. (2004). Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. *Genome Res.* **14,** 273–279.

Abstract

Theoretical work has shown that deleterious mutations may explain many evolutionary phenomena, but the role they play depends on their specific properties, including the overall rate at which they occur in the genome and the distribution of their selection (s) and dominance (h) coefficients. However, accurate estimates for these parameters from a range of species are still lacking. In this thesis, two distinct approaches were used to quantify the mutational process.

In Chapter 3, a novel DNA-based method was used to infer levels of evolutionary constraints in the *Drosophila* genome by comparing rates of nucleotide substitution in non-coding and putatively neutrally evolving DNA. Introns were found to have a significantly higher rate of substitution than synonymous sites, and, when introns were used as a neutrally evolving standard, constraint in the 500bp of intergenic DNA upstream and downstream of coding regions was found to be about 44%. Selection against mutations in intergenic DNA should therefore make a substantial contribution to the mutational load in *Drosophila*.

Secondly, a fitness-based approach was used to estimate mutational parameters in lines of *Caenorhabditis elegans* containing large numbers of deleterious homozygous EMS-induced mutations. In Chapter 4, replicated inbred sublines were produced for eight mutant lines, and the performance of the sublines, the mutant lines and the wild-type strain was measured for three fitness-related traits. The number of mutations per line was then estimated for each trait by applying a modified version of the Castle-Wright estimator and a maximum likelihood (ML) method. Both the Castle-Wright and the ML analyses suggest that most of the variation among sublines was due to a small number (~1.5–2.5) of large-effect mutations. Given that each line is expected to have a large number of mutations, this supports the hypothesis that many have very small (but still deleterious) effects.

In Chapter 5, the average dominance coefficient of mildly deleterious mutations ($\bar{h}$) was estimated from a selection of 19 relatively high fitness mutant lines by

comparing the performance of heterozygotes and homozygotes to the wild-type for three fitness-related traits (viability, productivity, and relative fitness). There was very little net heterozygous or homozygous effect of mutations on viability, but for productivity and relative fitness $\overline{h}$ was found to be ~0.1. Combined with the conclusion that most homozygous mutations have very mild effects, this suggests that many newly arising deleterious mutations may have very small heterozygous effects indeed. Furthermore, there was a significant amount of variation in h among lines, and analysis of h for individual lines suggested that several lines were heterotic (the heterozygotes outperformed both homozygotes). This has implications for our understanding of inbreeding depression and levels of genetic variation.

Additionally, the results of this thesis imply that the effects of new mutations are highly variable, and that small effect mutations and mutations in non-coding DNA make large contributions to the genomic deleterious mutation rate (U). Estimates of U may therefore be substantial underestimates if these types of mutation are ignored.

Contents

Abbreviations

List of commonly used abbreviations:

bp	Base Pairs
C	Constraint
dATP	Deoxyadenosine 5'-triphosphate
dCTP	Deoxycytosine 5'-triphosphate
ddH_2O	Double Distilled Sterile H_2O
dGTP	Deoxyguanosine 5'-triphosphate
DNA	Deoxyribose Nucleic Acid
dNTP	Deoxynucleoside 5'-triphosphate
dTTP	Deoxythymidine 5'-triphosphate
EDTA	Ethylenediaminetetraacetic Acid
EMS	Ethylmethane Sulphonate
f_e	Equilibrium G+C Content
g	Grams
h	Dominance Coefficient
kb	Kilobase
l	Litres
M	Molar
mM	Millimolar
MA	Mutation Accumulation
μl	Microlitre
mins	Minutes
ml	Millilitres
mRNA	Messenger RiboNucleic Acid
ML	Maximum Likelihood
MYOB	Modified Youngren's Only Bacto-peptone
n_e	Number of Effective Factors

p	Probability
PCR	Polymerase Chain Reaction
PEG	PolyEthylene Glycol
rpm	Revolutions per Minute
s	Selection Coefficient
SD	Standard Deviation
SE	Standard Error
secs	Seconds
TAE	Tris Acetate EDTA
Taq	*Thermus aquaticus*
Tris HCl	Tris [hydroxymethyl] Aminomethane Hydrochloride
Tris OH	Tris [hydroxymethyl] Aminomethane
U	Genomic Mutation Rate
UTR	Untranslated Region

All other abbreviations are for chemical formulae or are detailed in the main text.

1 Introduction

1.1 Importance of Deleterious Mutations

Mutations are a universal feature of living organisms and provide the source of genetic variability, which is the raw material for evolution. Adaptation of populations to their environment results from advantageous mutations being fixed by natural selection. However, in organisms that are already well adapted, mutations that increase the fitness of individuals are expected to be rare, so the vast majority of new mutations will be harmful (deleterious) or, at best, neutral. Paradoxically, therefore, although mutations are necessary for populations to evolve and increase in fitness over time, the majority of mutations actually reduce fitness in the short term. Several evolutionary phenomena (some of which are outlined below) are thought to have evolved as a result of the constant influx of new deleterious mutations into the gene pool of populations. Moreover, theoretical work has shown that the relative importance of mutations for the evolution of many of these phenomena is highly sensitive to their specific properties. Depending on the rate and effects of spontaneous deleterious mutations, they may have a crucial role in many different facets of evolution or just be an irrelevant nuisance.

Mutations are commonly defined as changing a wild-type locus (A) to the mutant state (a) at a certain rate (μ) per generation (*i.e.* the probability, per generation, that an A gene will mutate to a). In diploid populations, a single mutation creates the possibility of three different genotypes, AA, Aa and aa. The effect of these mutations on the phenotype of an organism is commonly quantified by using the concept of "relative fitness". Fitness is most easily understood by considering populations with separate generations, where it is defined as the the expected number of offspring contributed by an individual to the next generation. More precisely this defines a value known as "absolute fitness". However, under the simplifying assumption that population numbers are regulated then it is possible to introduce the concept of relative fitness which is the relative contribution an individual makes to the next generation

(the fitness of one type of individual is reduced to unity and other individuals are scaled by the same factor). Throughout this thesis, the genotypes AA, Aa and aa are defined as having relative fitnesses of 1, $1 - hs$ and $1 - s$ respectively, where s is the homozygous selection coefficient, and h is the dominance coefficient, relating the heterozygous effect of a mutation to the homozygous effect. On a genome-wide scale, several parameters can be used to characterise the mutational process, the most commonly used are the genomic deleterious mutation rate (U) (per diploid genome per generation), the average homozygous selection coefficient ($\bar{s}$) and the average dominance coefficient ($\bar{h}$). Since mutations may vary widely in their effects, however, it might also be desirable to consider the *distribution* of selection and dominance coefficients. We may also wish to consider the possibility of interactions between mutations at different loci (epistasis). Several theories for which knowledge of some of these parameters are important are outlined below.

1.1.1 The Evolution of Sex and Recombination

It has been known for sometime that asexual reproduction has intrinsic advantages over sexual reproduction. Firstly, there is a cost of male allocation (or allocation of limited resources to male function in hermaphrodites) (Maynard-Smith 1971, 1978). Asexuals have an advantage over sexual species by eliminating males, making them more efficient at producing offspring. For example, if a new asexual clone arises in a sexual population with equal numbers of males and females, the proportion of asexual individuals will double each generation, all else being equal (Maynard-Smith 1971). Secondly, there is a "cost of meiosis" (*i.e.* the reduction in genetic contribution to offspring) and this occurs regardless of whether the sexes are separate or combined (Williams 1971, 1975). In a sexual species, an individual only contributes half of the genes to its offspring, whereas in an asexual species an individual contributes all the genetic material to its offspring. These two costs are not interchangeable, apply to different situations and are probably mutually exclusive for any particular situation (Lively & Lloyd 1990). Additionally, there may be other costs associated with sexual reproduction itself; for example, in many species it takes time and energy to find a mate, the act of sexual reproduction may be slow, there may be costly sexual conflicts involved and there is a possible risk of obtaining sexually transmitted disease and parasitic genetic elements.

However, despite the costs to sexual reproduction, the majority of species engage in some form of sex (Bell 1982) and only a few species appear to have been asexual for a long period of time (Judson & Normark 1996). Many hypotheses have been proposed to explain the prevalence of sex, but they can essentially be divided into environmental or mutational hypotheses. In turn, both of these types of theories can be divided into stochastic (operating in finite populations) and deterministic explanations (which also apply to very large populations). Environmental hypotheses may apply to many species and do not contradict our knowledge of population biology, although it is questionable whether they can be applied to all species because the conditions necessary to offset the more than two-fold advantage of asexual reproduction may be limited (Case & Bender 1981, Moore & Hines 1981; but see West *et al.* 1999). On the other hand, mutational hypotheses apply to all species, because they all suffer from mutations.

The first mutational model proposed is known as the "Fisher-Muller" model and concerns advantageous mutations. In this model it is proposed that recombination hastens the rate of adaptation by combining advantageous mutations that occur in separate individuals. In an asexual populationn on the other hand beneficial mutations can only be combined if there occur within the same lineage (Fisher 1930, Muller 1932, 1964). This is generally thought to be a group selection argument and applies only to advantageous mutations and therefore requires some form of directional selection. Two other mutational models have been proposed that require only purifying selection since they are only concerned with deleterious mutations.

The first of these deleterious mutational hypotheses, known as "Muller's ratchet" is a stochastic mechanism and therefore applies to finite populations (Muller 1964). It makes the assumption that forward mutations (allele A to a) are common but reverse mutations (a to A) are comparatively very rare for any particular gene. In a finite population of asexuals, without recombination, at some point individuals with no mutations will be lost from the population, in Muller's phrase, the ratchet has clicked around one notch. The new optimal class now carries one mutation, but again in time it will be stochastically lost, as the ratchet clicks around again. In a sexual population, however, recombination effectively increases the variance in the number of mutations per individual, and the class of individuals with no mutations can therefore be re-established. This theory is only relevant if the number of individuals with no mutations is small and therefore only applies to small populations, because the

chance of stochastically losing the genotype with zero mutations becomes negligible in large populations (Maynard-Smith 1978, Pamilo *et al.* 1987). Furthermore, it only applies to species that are obligately asexual, as even a small amount of sex can prevent the ratchet from turning (Pamilo *et al.* 1987, Bell 1988, Charlesworth *et al.* 1993).

The second major deleterious mutational hypothesis (the so called "deterministic mutation hypothesis"; Kondrashov 1982) relies on the observation that sex can increase the efficacy of natural selection, which could lead to a lower mutation load in sexuals. The mutation load is defined as the proportion by which population fitness is decreased due to the elimination of recurrent harmful mutations (Crow 1958). In an asexual population, it has been shown that this quantity is equal to $1-e^{-U}$ (where U is the deleterious mutation rate per genome per generation, assuming that deleterious mutations are partially recessive). If the effects of mutations on fitness are independent across loci the population is randomly mating, then the same relationship holds for a sexual population. However, this load could be much lower in a sexual population (depending on the mode of selection; Kimura & Maruyama 1966), giving a possible advantage to sex. Whether or not this process can overcome the inherent cost of sexuality, depends heavily on the deleterious mutation rate (U) and type of selection. Firstly, it requires that there is either truncation selection (where having a few mutations does not affect fitness, but more than a critical number gives zero fitness), or, more generally, that mutations act synergistically (where the each additional deleterious mutation leads to a larger decrease in relative fitness). It also requires that U is of the order of one per diploid genome per generation (Kondrashov 1982). If synergistic epistasis is a general property of deleterious mutations, and $U \gg 1$ then this hypothesis would be sufficient to explain the maintenance of sex. On the other hand, if $U \ll 1$, then it is insufficient as an explanation, regardless of whether mutations act synergistically. It is also possible that a similar mechanism could operate to provide an advantage to sex even without synergistic interactions between mutations (Chasnov 2000). In this model, an advantage to sex results from nearly-recessive mutations that act independently, providing, again, that the mutation rate is of the order of one.

Interestingly, a genomic deleterious mutation rate of this magnitude could also provide an explanation for the evolution of eusociality (Cherry 2002) because siblings produced through helping parents could carry fewer deleterious alleles than offspring produced by mating.

1.1.2 The Evolution of Ploidy level

The origin of sex brings with it the origin of haploid and diploid life cycles (the haploid and diploid phases are tightly linked with that of the origin of mixis (or fertilisation) and meiosis). However, the length of the diploid phase varies amongst different classes of organisms, and in higher eukaryotes, the diploid phase dominates. Why is this, and what are the advantages of diploidy? Although diploidy may have immediate ecological consequences, that could affect the persistence of a lineage (Otto & Goldstein 1992), it is possible that the widespread prevalence of diploidy is due to its genetic consequences. There are three genetic explanations for the prevalence of diploidy. Firstly, it may have evolved because deleterious mutations can be masked in the diploid state if they are recessive (Kondrashov & Crow 1991, Perrot *et al.* 1991, Otto & Goldstein 1992, Jenkins & Kirkpatrick 1995). Secondly, it is possible that the greater number of genes results in more frequent production of adaptive mutations (Paquin & Adams 1983, Orr & Otto 1994). Finally, it has been suggested that diploidy would be favoured if spontaneous deleterious mutations are overdominant (Crow & Kimura 1965).

Diploid individuals may have an advantage over haploids because deleterious recessive mutations are masked. This advantage, although possibly only transient, could allow diploid individuals to invade a population of haploids (Perrot *et al.* 1991). Although the genetic load in diploids could be higher at equilibrium due to the doubled mutation rate (Crow & Kimura 1965), this may not prevent the evolution of diploidy. This is because a return to haploidy (if diploidy became fixed) would be difficult, due to the unmasking of recessive deleterious mutations. The feasibility of this theory depends on how many mutations can be masked in the diploid state, and therefore on the average dominance coefficient and the distribution of dominance coefficients for spontaneous deleterious mutations are crucial parameters. It has also been suggested that diploidy may have a long term advantage over haploidy because the diploid genetic load may actually be lower than the haploid load, even at equilibrium. For this to be true, double the number of mutations would have to be removed in a diploid population, with fewer genetic deaths than in a haploid population at equilibrium. This may be possible, but requires a high input of deleterious mutations (more than one per genome per generation; Kondrashov & Crow 1991), and also depends on the dominance coefficients of deleterious mutations and the mode of selection.

For example, if there is truncation selection then diploidy will be favoured when dominance is less than 0.25 (Kondrashov & Crow 1991).

It is possible that diploidy evolved because of the more frequent production of adaptive mutations (because diploid organisms have twice the number of genes). However, diploids may not actually adapt faster, since if an advantageous mutation is not dominant, diploidy increases fixation time, meaning that haploids may actually have the advantage. Whether or not they do so depends on whether the fixation time or the appearance of new beneficial alleles is limiting (which depends on the population size and the advantageous mutation rate respectively; Orr & Otto 1994, Otto & Whitton 2000). The predictions of this theory have been borne out experimentally, where an evolutionary advantage to haploidy was observed in large yeast populations, but not in small ones (Zeyl *et al.* 2003).

Finally, if a substantial fraction of deleterious mutations are overdominant then diploids would have an immediate advantage over haploids and could therefore invade a population (Crow & Kimura 1965). However, if the diploid phase is initially only short, then overdominance will not be able to maintain the polymorphism, negating the advantages of diploidy (Goldstein 1992).

1.1.3 Maintenance of Genetic Variability

Understanding how genetic variance is maintained in a population is an important challenge for evolutionary biologists (Lewontin 1974). Fisher's fundamental theorem of natural selection (Fisher 1930) implies that additive genetic variance for fitness should be removed by selection. However, this is not observed empirically, instead genetic variance for life-history traits is often found to be quite substantial (Charlesworth 1987, Mukai 1988).

Three possible processes for maintaining genetic variance for fitness have been proposed. In two of these models, alleles are maintained by selection at intermediate frequency either by pure balancing selection (due to overdominance or frequency dependence of genotypic fitnesses), or by special cases of directional selection. Alternatively, genetic variation may be maintained by the balance between mutation and selection (Charlesworth & Hughes 1999). In order to determine the contribution to genetic variation from these two basic types of process, it is necessary to have

estimates for parameters associated with deleterious mutations, including the rate at which they arise (U), the distribution of selective effects, and the average dominance coefficient ($\bar{h}$). For example, in order for there to be a substantial contribution to genetic variation from new deleterious mutations, they must usually be partially recessive and not dominant. On the other hand, if a substantial fraction of new mutations are overdominant, then balancing selection will contribute greatly to the maintenance of genetic variation (Charlesworth & Hughes 1999).

1.1.4 Inbreeding Depression

Inbreeding depression is the decrease in the mean value of a character upon inbreeding. It is observed in most naturally outbreeding species and occurs as a result of increased homozygosity at loci affecting fitness across the genome (Wright 1977). There are two major mechanisms by which increased homozygosity could lead to a reduction in fitness. Firstly, inbreeding depression could represent the effects of loci at which there is overdominance, so that heterozygotes are superior to homozygotes (Charlesworth & Charlesworth 1987). Secondly, inbreeding depression could be caused by recessive or partially recessive deleterious mutations (Wright 1977), maintained in the population by recurrent spontaneous mutation (or balancing selection acting on their net fitness effects). The relative importance of these two mechanisms is still under debate and it is likely that both mechanisms are involved (Charlesworth & Charlesworth 1999).

To tease apart the relative contribution of these two processes, estimates of the genomic deleterious mutation rate and the average degree of dominance are needed. Inbreeding depression due to deleterious mutations will increase with the deleterious mutation rate if mutations are partially recessive. More precisely, the number of mutations that can be hidden in the heterozygous state at mutation-selection balance increases as the average dominance coefficient of mutations decreases, resulting in greater amounts of inbreeding depression. On the other hand, a large proportion of observed inbreeding depression could be the result of balancing selection if even a fairly small fraction of newly arising mutations are overdominant. This is because overdominant mutations will be maintained by selection and can therefore persist for longer time periods than partially recessive mutations.

1.1.5 Survival and Conservation of Small Populations

When population size is reduced, the rate of inbreeding increases, leading to greater amounts of inbreeding depression and increasing the risk of extinction (Hedrick & Kalinowski 2000). Secondly, as genetic drift becomes a more important factor, the genetic variability available for adaptive evolution may be lost, and mildly deleterious mutations could drift to fixation and accumulate over time. This will result in a gradual erosion of fitness in small populations, further threatening their persistence (Lande 1994, 1995, Lynch *et al.* 1995*a*,*b*). Mutations that have homozygous effects less than the inverse of the effective population size are effectively neutral, and can accumulate by genetic drift, this in turn is expected to lead to a reduction in the population size. As the population size decreases, more mutations will act neutrally and the rate of accumulation of deleterious mutations will increase. There is therefore a interaction between population size and the accumulation of mutations leading to an extinction process known as mutational meltdown (Lynch *et al.* 1993). The threat that mutation accumulation poses depends not only on the rate of mutation, but also on the distribution of mutation effects. Mutation accumulation could also threaten the persistence of our own species, although in this case, this effect is not due to a low effective population size, but rather due to the possible relaxation of natural selection (Muller 1950, Kondrashov 1995, Crow 1997).

1.1.6 Evolution of Ageing

Ageing is the gradual deterioration in an organism's condition leading to increased risk of death and decreased fertility. From an evolutionary perspective, ageing limits the reproductive potential of an individual and should therefore be opposed by natural selection (Kirkwood & Rose 1991, Partridge & Barton 1993). However, many organisms throughout the animal kingdom age, even in natural populations (Loison *et al.* 1999). Furthermore, ageing occurs at very different rates in different kinds of animals, suggesting that the rate of ageing itself might actually have evolved. There are two current theories to explain the evolution of ageing. Firstly, the "mutation-accumulation" theory suggests that ageing is the result of lower selection in later life (Medawar 1946, 1952, Hamilton 1966, Charlesworth 1994). It is possible that this lowered selection could result in an accumulation of age-specific deleterious mutations in a population over time (Medawar 1952, Charlesworth & Hughes 1996).

The "antagonistic pleiotropy" theory, however, suggests that ageing is the result of positive selection for genes that are advantageous in early life but deleterious later in life (Williams 1957, Kirkwood & Rose 1991, Partridge & Barton 1993, Charlesworth 1994). Although there is increasing evidence for the antagonistic pleiotropy theory of ageing (Partridge & Gems 2002), the two processes are not mutually exclusive and both may contribute to ageing.

1.1.7 Sympatric Speciation

Deleterious mutations can also provide a mechanism for sympatric speciation, if they have habitat-specific effects in a spatially heterogeneous environment. This requires that some deleterious mutations are deleterious in only one (marginal) habitat but neutral, or nearly so in the main habitat. These mutations can accumulate If their effect on total reproduction is weak (because the contribution of the marginal habitat to total reproduction is small). Such mutations could lead to indirect selection for habitat preference, and thus could result in the subdivision of populations into different habitats that then become genetically and reproductively isolated from one another (Kawecki 1997). The rate of deleterious mutation and the fraction of deleterious mutations with habitat-specific effects are therefore both crucial parameters for this theory.

1.2 Measuring Parameters Associated with Deleterious Mutations

There are two main approaches that have been used to estimate some of the key parameters associated with deleterious mutations and they can be classified as fitness-based and DNA-based (Kondrashov 1998, Bataillon 2000). A brief outline to each of these methods and a summary of previous work is given below.

1.2.1 Fitness-based Methods

Despite the potential importance of deleterious mutations, there are still very few estimates of the genomic rate of deleterious mutation, or of the number of deleterious alleles segregating in wild populations. One of the problems in obtaining such

estimates is that the majority of deleterious mutations may have very small or effectively undetectable effects on fitness (Keightley & Eyre-Walker 1999). It is much more straightforward to estimate the frequency of recessive mutations with very large homozygous effects. In particular, the number of recessive lethals per individuals in wild populations may be estimated objectively and unambiguously, and there are several published estimates of this number.

The majority of estimates of the number of lethal alleles per individual in wild populations come from various *Drosophila* species. The general method for detecting lethal genes in *Drosophila* was suggested by Muller (1928), and involves the use of balancer chromosomes to test for lethal alleles in a specific chromosome (see Figure 1.1). It is possible that lethal mutations revealed using this scheme are in fact the result of one or more smaller effect mutations, but there evidence to suggest that many are the result of a single point mutation (Lewontin 1974). It is possible to estimate the mean number of lethals per genome per individual, taking into account the proportion of the genome accounted for by the particular chromosome tested and assuming that lethals within a chromosome follow a poisson distribution. Several experiments using this approach have been performed in *Drosophila*, and much of the data have been summarised and converted to estimates of the mean number of recessive lethal alleles per individual by Lewontin (1974) (see also Simmons & Crow 1977). In all 18 such experiments (Lewontin 1974), all but one estimate of the number of recessive lethals per individual fell in the range 0.5 to 3.

It is less straightforward to measure parameters associated with all deleterious mutations, however, given that it is likely that lethal mutations only account for a small fraction of the total spectrum, estimates of these parameters would be desirable. The classical approach to estimating these mutational parameters relies on keeping inbred lines or individual chromosomes protected from natural selection, so that mutations accumulate by drift over several generations. This is often achieved by keeping the population size to a minimum, so that drift dominates selection allowing all but the most severely deleterious mutations accumulate neutrally. The mean and variance for fitness at different time points during mutation accumulation (MA) can then be compared to controls to infer the average number and average homozygous effect of mutations. Accumulated mutations can also be used to estimate the average dominance coefficient, by comparing the performance of homozygotes and heterozygotes to the wild-type.

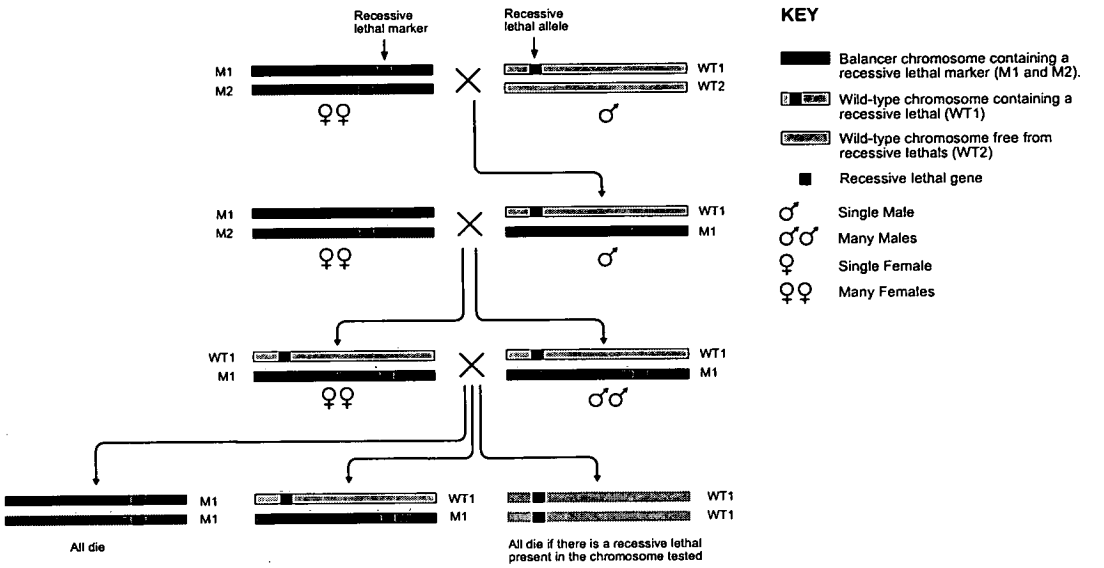


Figure 1.1: Balancer chromosome crossing scheme for the detection of recessive lethal alleles on a particular chromosome of interest in *Drosophila* (adapted from Lewontin 1974). This scheme allows the detection of recessive lethal alleles in one chromosome from a wild-type individual by crossing the individual to a balanced marker stock population. A single wild-type male (carrying two homologous wild-type chromosomes) is crossed to many balanced marker stock females. The balanced marker stock have two different dominant marker genes (M1 and M2) on homologous chromosomes. One of these chromosomes (M1) also contains recombination suppressing inversions that keep the wild-type chromosome intact. A single male is selected from the F1 offspring on the basis of having the M1 heterozygous phenotype, thereby choosing one wild-type chromosome to study, and backcrossed to the marker stock. The backcrossed offspring are intercrossed, producing many offspring, which are scored. If a recessive lethal is present on the wild-type chromosome tested, as in this case, then only heterozygous individuals will be produced from the final cross, providing a simple and objective scheme to test for the presence of a recessive lethal on a random wild chromosome

1.2.1.1 Rate of Mutation and Homozygous Effects

This mutation accumulation method was pioneered by Terumi Mukai and collaborators working on *Drosophila* in the 1960s and 1970s (Mukai 1964, Mukai *et al.* 1972, Ohnishi 1977*a*; see reviews in Simmons & Crow 1977, Keightley & Eyre-Walker 1999 and Lynch *et al.* 1999). Mukai used a *Cy/Pm* balancer chromosome as a control, which largely protects the heterozygous wild-type (+) second chromosome from selection, allowing the accumulation of mutations on a single chromosome. He then carried out competitive viability assays to measure the effects of the accumulated mutations, and found that fitness declined linearly over time (a decrease of about 1–2% per generation). The mutation accumulation lines were replicated so that the change in genetic variance between lines per generation could be estimated. This was found to increase, presumably due to the fixation of different numbers of mutations in each line. Mukai used a method of moments (originally formulated by Bateman 1959) to estimate the rate and effects of mutations, based on the decrease in mean fitness and increase in genetic variance. Mukai estimated U to be on the order of 0.4 per genome per generation (assuming equal effects of mutations) (see Keightley & Eyre-Walker 1999) . There are now other methods that have been developed to infer mutational parameters using either maximum likelihood (ML) or minimum distance (MD) approaches. These methods have the advantage that they can use the distribution of mutant line means and do not just use the variance amongst them. This often allows better and more robust estimates with smaller standard errors (Keightley 1998). Additionally, these methods can allow for variation in the effects of mutations by fitting a parametric distribution of mutation effects. The gamma distribution has been widely used for this purpose because changing a single parameter value leads to distributions with a wide range of properties. If effects are assumed to be equal, however, only minimum and maximum estimates of the rate and effects of mutations can be obtained respectively.

These methods have been applied to the original data from Mukai and Ohnishi's experiments (Keightley 1994, 1996, García-Dorado 1997, Fry 2001) as well as to new MA experiments carried out using *Drosophila* and other model organisms (Fernández & López-Fanjul 1996, Keightley & Caballero 1997, Fry *et al.* 1999, Vassilieva *et al.* 2000, Zeyl & DeVisser 2001, Shaw *et al.* 2002). In some cases the distribution of mutation effects has also been inferred (Fry *et al.* 1999, Vassilieva *et al.* 2000, Shaw

et al. 2002). Some of these analyses suggest that the mutation rate in *Drosophila* may be much lower than first suggested by Mukai (possibly on the order of 0.01) (Fry *et al.* 1999, García-Dorado *et al.* 1999), although there is still much debate over the interpretation of MA results (see Keightley & Eyre-Walker 1999 and Lynch *et al.* 1999 for reviews). There is also a large disparity in the distribution of effects that have been inferred from such experiments, some finding evidence for a large variation in effects (Keightley 1994, Fry *et al.* 1999) and some finding much less (Keightley & Caballero 1997, Vassilieva *et al.* 2000).

Despite taking into account variation in the effects of mutations, these estimates of U may still be large underestimates. In distribution-based approaches such as those discussed above, U and the distribution of effects are confounding variables, making it difficult to estimate one without knowledge of the other (Keightley 1998). Unfortunately, estimates of the distribution of effects of mutations are dependent on higher order moments, which require large quantities of data for accuracy. Since MA experiments are often limited in the amount of data that can be collected, estimates of the distribution of effects are often noisy and imprecise, making estimates of U questionable. Since MA experiments are also sensitive to environmental variation, they may miss mutations with extremely small but deleterious effects all together, and the presence of a large number of mutations in this class is often impossible to rule out.

There is direct evidence to suggest that the majority of mutations have small effects on fitness (<5%), the majority of which would be missed in mutation accumulation experiments (see Lynch *et al.* 1999 for a summary). Additionally, indirect evidence for a large class of small effect mutations comes from an analysis of EMS-induced mutations by Davies *et al.* (1999). The distribution of effects of EMS-induced mutations can be inferred by comparing an estimate of the number of induced mutations to an estimate of the number of mutations detectable from fitness assays. This approach was used by Davies *et al.* and further analysis was carried out by Keightley *et al.* (2000). An estimate of the number of mutations induced can be obtained from the expected rate of EMS-mutatagenesis, from experiments to measure forward mutation rates (Bejsovec & Anderson 1988) and suppresser-induced reversion mutation rates (Waterston 1981, Hodgkin 1985, Kondo *et al.* 1990). This can be converted to a conservative estimate of the number of deleterious mutations induced, by using information on the size of the genome, the percentage of the genome that is

protein coding, and the level of constraint within protein coding sequences. Davies *et al.* estimated that they had induced approximately 45 mutations per homozygous mutant *C. elegans* line studied but only 3.60 (±1.31) were detectable (Keightley *et al.* 2000) from estimates of relative fitness. It was inferred, therefore, that the majority of mutations had undetectably small (but still deleterious) effects on fitness.

1.2.1.2 Heterozygous Effects

U and $\bar{s}$ are well studied parameters (although current results are inconclusive), but $\bar{h}$, on the other hand, has received comparatively much less attention. There are several experiments that have been carried out in an attempt to estimate $\bar{h}$, and once again Terumi Mukai is responsible for the first direct estimates. Mukai used the same MA lines described above but selected only chromosomes with ≥50–60% normal viability ("quasinormals") in order to study the effects of mildly deleterious mutations. The viabilities of these chromosomes were assayed in the homozygous and heterozygous states, alongside controls that were homozygous for wild-type or wild-type-like second-chromosomes. Similar experiments were subsequently carried out by Ohnishi (1977*b*) (see reviews in Simmons & Crow 1977, Houle *et al.* 1997 and García-Dorado & Caballero 2000).

The results obtained from Mukai's experiments depended on the type of heterozygote formed (coupling or repulsion), and the method used to calculate dominance (whether a proportional difference in means, or a regression approach was used). In general, however, estimates of $\bar{h}$ from Mukai's experiments suggest that mutations are partially recessive on average ($0 < \bar{h} < 0.5$), but there is considerable variability among the estimates, with some from coupling heterozygotes even suggesting that mutations could be overdominant on average ($\bar{h} < 0$). Similarly, more recent estimates using a range of organisms and experimental designs, suggest that deleterious mutations are partially recessive on average, but again the results are variable (Eanes *et al.* 1985, Hughes 1995, Johnston & Schoen 1995, Vassilieva *et al.* 2000) (see Section 5.1 for a more detailed discussion).

1.2.2 The DNA-based Method

The DNA-based method for estimating parameters associated with deleterious mutations uses comparisons between homologous DNA sequences of extant species. It is based on the concept that the rate of evolution in a neutrally evolving sequence is equal to the neutral mutation rate (μ_t) (Kimura 1983). The method also assumes that mutations are either neutral or deleterious, so that in constrained regions, substitutions occur at a rate $f\mu_t$, where f is proportion of mutations that are neutral ($1 - f$ are deleterious, and these will be removed by natural selection). The rate of evolution of a sequence (μ_t) can be estimated if the number of generations since the last common ancestor of the two species is known, and this can be extrapolated to the whole genome to estimate the genomic mutation rate U_t if the number of nucleotides in the genome is known. To estimate the deleterious mutation rate, U, the fraction of mutations that are deleterious throughout the genome ($1 - f$) must be calculated ($U = (1-f)U_t$). This figure can be obtained by comparing the divergence in a random selection of constrained regions (*e.g.* exons) to that in non-functional neutral regions (*e.g.* pseudogenes). This gives the proportion of mutations that are deleterious in the constrained regions, which can be scaled up to the whole genome by multiplying by the fraction of the genome that is constrained (Kondrashov & Crow 1993).

Eyre-Walker & Keightley (1999) used a modified version of this method to estimate the deleterious mutation rate using 46 homologous genes from humans and chimpanzees, assuming that synonymous sites were evolving neutrally. Extrapolating to the whole genome gave $U = 0.8$. However, this method has a number of caveats. Firstly, it ignores advantageous mutations, which would bias the estimate of U downwards. Secondly, it gives no information about the size or distribution of mutation effects (other than that the mutations that are removed by selection have effect sizes approximately greater than the reciprocal of the effective population size). Another potential problem is that the possibility of constraint in non-coding DNA was ignored and the deleterious mutation rate could be a large underestimate if this is substantial.

There are good reasons for believing that constraint in non-coding DNA might be substantial. For example, promoters are located in non-coding DNA, and have been shown to be conserved (Blackman & Meselson 1986, Fisher & Maniatis 1986, Dickinson 1991, Cavener 1992, Oeltjen *et al.* 1997, Gottgens *et al.* 2000,

Hardison 2000) and 5' and 3' untranslated regions maybe under constraint due to their functional role in pre-messenger RNA structure (Stephan & Kirby 1993, Kirby *et al.* 1995, Leicht *et al.* 1995). Furthermore, many previous studies on individual gene clusters (den~ Dunnen *et al.* 1989, Liew *et al.* 1990, Epp *et al.* 1993, Hardison & Miller 1993, Koop & Hood 1994, Koop *et al.* 1996, Jan *et al.* 1997, Oeltjen *et al.* 1997) and on genome-wide scales (Duret *et al.* 1993, Webb *et al.* 2002, Jareborg *et al.* 1999, Shabalina *et al.* 2001, Bergman & Kreitman 2001, Keightley & Gaffney 2003) have found evidence for constraint in non-coding DNA (but the relative importance of intronic *vs.* intergenic *vs.* exonic sequences remains unclear; for a more detailed analysis of previous work see Section 3.1). Genome-wide estimates of constraint in non-coding DNA could theoretically be used to estimate U (at least in non-coding DNA) if an estimate of the per nucleotide mutation rate could be obtained.

1.3 Aims Of This Study

In this study, the mutational process has been quantified using both DNA-based and fitness-based approaches. Three different studies have been carried out, all of which relate to the effects and rate of accumulation of deleterious mutations.

The fraction of the genome that is functionally important is vital for estimating the genomic deleterious mutation rate. However, although non-coding DNA makes up the vast majority of the genome in multicellular eukaryotes, the fraction that is functionally important is still unclear and the relative importance of intergenic *vs.* intronic DNA is debatable. In Chapter 3, a DNA-based approach has been used to quantify the level of functional constraint in non-coding DNA. Differences in the rate of fixation of mutations between a neutral sequence and a potentially functional sequence can be assumed to be due to negative selection, so a lower rate of substitution implies functional constraint in that region. Two closely related species were compared, for which it was possible to align almost all non-coding DNA reasonably well. The rate of evolution in putatively neutral sequences (introns or synonymous sites) was used to estimate the mutation rate. The fraction of missing substitutions in a nearby functional sequence was then used to quantify constraint (Kimura 1983). Since the putatively neutral sequences were adjacent to the functional sequences, the mutation rate in each should be similar. *Drosophila* provide a close to ideal choice of organism for studying constraint in non-coding DNA, because the

genome sequence for *D. melanogaster* is available and these species are not affected by the hyper-mutability of CpG dinucleotides, which show a 10- to 15-fold higher rate of substitution in mammals (Shen *et al.* 1994). In order to obtain data to carry out a sufficient number of comparisons, additional upstream DNA from *Drosophila simulans* was sequenced.

In Chapters 4 and 5, lines of *Caenorhabditis elegans* containing large numbers of homozygous EMS-induced mutations were used to study the fitness effects of mutations. These lines, generated by Davies *et al.* (1999), provide an excellent system for studying mutations for several reasons. Firstly, *C. elegans* is an ideal study organism since it has a short generation time (three to four days at 20°C), the adults are easily maintained and can be cryopreserved, allowing lines to be kept indefinitely. It is also a diploid eukaryote for which we have the complete genome sequence, which can reproduce either by selfing or by crossing. Secondly, EMS-mutations provide a very powerful way to study mutational effects because large numbers of mutations can be studied without the need for long periods of mutation accumulation, and it has been argued that the effects of induced mutations should be similar to those of spontaneous mutations (Davies *et al.* 1999).

The aim of the work described in Chapter 4 was to verify the conclusions made by Davies *et al.* (1999). Davies *et al.* inferred that the majority of homozygous deleterious mutations induced in 56 lines of *Caenorhabditis elegans* had very small, undetectable, but still deleterious effects on fitness. This is an important conclusion, because if it is also true of spontaneous mutations, then it could mean that many estimates of U are substantial underestimates. The results also imply that the distribution of mutation effects is highly leptokurtic and L-shaped. However, the estimate of a small number of detectable mutations made by Davies *et al.* was based on the distribution of fitness-related trait values for lines with and without mutations. They used a maximum-likelihood approach to estimate the number and distribution of mutation effects, but it is hard to estimate these parameters simultaneously under this experimental design (Keightley 1998). Furthermore, estimates of the distribution of effects of mutations are dependent on higher order moments, which require large quantities of data for accuracy.

In this study, inbred sublines were created from a random selection of the EMS-induced mutant lines produced by Davies *et al.* (1999). Sublines, which should

contain a random selection of half of the mutations present in each mutant line, ought to make it possible to obtain more precise estimates of the number of detectable mutations per line and make more accurate inferences about the distribution of effects. This is because a large-effect mutation present in a line should segregate amongst its sublines, and the pattern of segregation of mutations should give information about the distribution of mutation effects without having to rely on information from higher order moments. A modification of the Castle-Wright estimator (Castle 1921, Wright 1968) and a maximum likelihood (ML) method were used to estimate the average number of mutations per line. The ML approach can deal with data for which the distribution of residual data points is expected to be significantly different from the expectations of a normal distribution. The method also allows for two classes of mutation effect.

The aim of the work described in Chapter 5 was to provide a new estimate of the average dominance coefficient for mildly deleterious induced mutations in *Caenorhabditis elegans*. In order to study only the effects of mildly deleterious mutations, a selection of EMS-induced lines that performed well in previous assays were selected. The average dominance coefficient was then estimated by comparing the performance of worms with sets of homozygous mutations and heterozygous mutations to the wild-type strain. Both a proportional difference in means approach and a regression approach were used to estimate $\bar{h}$. The variation in values of h for individual lines has also been examined.

2 Materials & Methods

The standard experimental techniques and solutions used throughout the course of this work are detailed in this chapter. Section 2.1 refers to Chapter 3 and Section 2.2 refers to Chapters 4 and 5.

2.1 DNA Sequencing Techniques

This section provides details of techniques used to purify, amplify and sequence genomic *Drosophila simulans* DNA.

2.1.1 Isolation of DNA from a Single *Drosophila*

Genomic template DNA was obtained from a single male *Drosophila simulans* using the protocol below and a PUREGENE® DNA Isolation Kit (Gentra Systems, Research Triangle Park, NC), which contains *Cell Lysis Solution*, *Protein Precipitation Solution* and *DNA Hydration Solution*. A single male fly was used as a source of DNA in all cases to reduce sequencing problems associated with heterogeneity in template DNA.

1. Cell lysis: one fly was added to a chilled 1.5ml centrifuge tube containing 100μl *Cell Lysis Solution* on ice. The fly was then homogenised thoroughly using a microcentrifuge tube pestle and the sample placed back on ice before incubating at 65°C for 15 minutes.

2. Protein precipitation: the sample was cooled to room temperature and 33μl *Protein Precipitation Solution* was added to the cell lysate. Then the *Protein Precipitation Solution* was gently mixed with the cell lysate and the sample placed back on ice for 5 mins. The sample was then centrifuged at 13,000 rpm for five minutes so that the precipitated proteins formed a tight pellet.

3. DNA precipitation: the the supernatant containing the DNA was poured into a 1.5ml centrifuge tube containing 100μl 100% Isopropanol (2-propanol) (leaving behind the precipitated protein pellet). The sample was then mixed by inverting gently 50 times before centrifuging at 13,000 rpm for five minutes. The supernatant was then poured off and the tube was drained onto clean absorbent paper. 250μl 70% ethanol (v/v aqueous solution) was then added and the tube inverted several times to wash the DNA pellet. The sample was then centrifuged once more at 13,000 rpm for 1 minute and the ethanol poured off. The tube was drained onto clean absorbent paper again by inverting and then allowed to air dry for 15 mins.

4. DNA hydration: 50μl of *DNA Hydration Solution* was added to the dried tube, and the DNA was allowed to rehydrate overnight at room temperature. Rehydrated DNA was stored at 2–8°C.

2.1.2 Standard Polymerase Chain Reaction (PCR)

PCR reactions were carried out in a PTC-100® programmable thermal cycler (MJ Research Inc., Reno, NV). All PCR reagents were obtained from Qiagen (Sussex, UK). Primers were designed using Primer Premier 5 (Premier Biosoft International, Palo Alto, CA), and were supplied by Sigma-Genosys (The Woodlands, TX 77380).

1. Reaction mix: for a single 50μl reaction, the following were mixed in 0.5ml tubes on ice:

35.8μl	ddH_2O
5μl	10× Qiagen PCR buffer
0.5μl	10mM dNTP
2.5μl	25mM $MgCl_2$
0.2μl	Taq DNA polymerase
2.5μl	20mM 5' Primer
2.5μl	20mM 3' Primer
1μl	Template DNA
50μl	

The volume of a single reaction was often halved to conserve PCR reagents and template DNA. Increasing the $MgCl_2$ concentration in a reaction decreases the

binding specificity of primers and so this was varied (by adding 1–3μl 25mM $MgCl_2$) in order to get successful but specific amplification of the target DNA. To check for specific amplification, 3μl of the PCR product was run on an agarose gel (see Section 2.1.3).

2. PCR program:

   96°C for 2 mins
   94°C for 30 secs
   53°C for 45 secs } 35 cycles
   72°C for 90 secs
   72°C for 5 mins

   The annealing temperature was initially set at 53°C, although this was raised if PCR amplification was not specific enough, or lowered if no product was obtained.

2.1.3 Agarose Gel Electrophoresis

Successful amplification of the target DNA was checked by running samples on an agarose gel. Samples to be loaded onto the gel were mixed with an equal volume of 1.5× loading dye and were loaded alongside a size marker (1kb DNA ladder, Promega, Madison WI). Horizontal gel electrophoresis was carried out in 1× TAE buffer in "Bio-Rad" gel tanks (Hercules, CA). Gels were run at 100V. After electrophoresis, results were visualised and photographed under UV light using a polaroid camera (Genetic Research Instrumentation Ltd., Essex, UK)

2.1.4 PEG Precipitation

Polyethylene glycol (PEG) can be used to remove unincorporated primers and nucleotides after PCR. For 50μl PCR reactions, 7μl 5M NaCl and then 19μl 33% PEG was added to each tube containing PCR product. The tubes were then vortexed and allowed to stand for five minutes before they were centrifuged in a plate rotor at 4,400 rpm for 30 minutes. The PEG was then knocked off, leaving the DNA pellet behind, any remaining PEG was removed by spinning the tubes upside down without their lids at 500 rpm for two minutes. The DNA pellet was then washed by adding

200μl of cold 70% ethanol. The ethanol was then knocked off and the tubes spun upside down without lids for one minute at 500 rpm. The DNA pellets were allowed to airdry before resuspending in 7μl ddH_2O.

2.1.5 Enzyme Digestion

Enzyme digestion can also be used to remove unincorporated nucleotides and primer sequences. Five microlitres of PCR product (5–20ng) was taken and mixed with 1μl of exonuclease 1 and 2μl of shrimp alkaline phosphatase. This mixture was incubated at 37°C for 30 minutes and then 80°C for 15 minutes. Two microlitres of the product was used in the subsequent sequencing reaction.

2.1.6 Sequencing Reaction

BigDye® Version 3 (Applied Biosystems, Foster City, CA 94404) was used for all sequencing reactions.

1. Reaction mix: the following reaction mix was prepared in sterile 0.5ml microtubes on ice.

   2μl BigDye® terminator ready reaction mix
   2μl PCR product (7.5–22.2ng)
   1μl Primer (0.8 pmol/μl)

   Both a forward and reverse primer reaction was prepared if necessary.

2. Sequencing program:

   96°C for 30 secs
   55°C for 15 secs } 25 cycles
   60°C for 4 mins

2.1.7 Asymmetric PCR and Sequencing

The protocol used was based on that described by Miller *et al.* (2003)

1. For each pair of primers, two primer mixes (Mix 1A and 1B) were prepared. The primer stock solutions were mixed so that for Mix 1A, the 5' primer was

at 10μM and the 3' primer was at 1μM, and for Mix 1B, the 3' primer was at 10μM and the 5' primer was at 1μM.

2. The following mixture (Mix 2) was prepared.

1μl	10× PCR buffer ($MgCl_2$ free)
1.4μl	25mM $MgCl_2$
1μl	1mM dNTP
0.04μl	Taq DNA polymerase
4.56μl	ddH_2O
8μl	

3. PCR reaction: the following were mixed in 0.5ml microtubes on ice

Forward Reaction		Reverse Reaction	
1μl	Mix 1A	1μl	Mix 1B
8μl	Mix 2	8μl	Mix 2
1μl	Template DNA	1μl	Template DNA

4. PCR program:

95°C for 2 mins
92°C for 10 secs
58°C for 20 secs } 35 cycles
68°C for 30 secs
68°C for 10 mins

Successful amplification was checked by running 3μl on an agarose gel (see Section 2.1.3).

5. Asymmetric sequencing reaction: the following were mixed in a 0.5ml microtube on ice:

2.5μl	PCR product
1μl	5× Asymmetric sequencing buffer
2μl	BigDye® terminator ready reaction mix
6.5μl	ddH_2O
12μl	

6. Asymmetric sequencing program:

96°C for 2 mins	
94°C for 15 secs	} 35 cycles
50°C for 1 second	
60°C for 4 mins	

2.1.8 Solutions

- PEG: to make 40ml 33% PEG solution:

13.2g	Polyethylene glycol 8000
0.08132g	$MgCl_2$
40ml	ddH_2O

Filter sterilised into 10ml aliquots.

- 10× Tris acetate EDTA (TAE) buffer

48.4g	Tris OH
10.9g	Acetic acid
2.92g	EDTA

Distilled water was added to 1 litre.

- 6× Loading dye

0.09%	Bromophenol blue
0.09%	Xylene cyanol ff
60%	Glycerol
60mM	EDTA

- 5× Asymmetric sequencing buffer

400mM	Tris HCl
10mM	$MgCl_2$

Adjusted to pH9.

- Agarose gel

90ml	1× TAE buffer
1g	Ultrapure agarose

This mixture was heated in a microwave for 1–2 mins until the agarose was fully dissolved, and then allowed to cool to approximately 60°C. Two microlitres of 10mg/ml ethidium bromide was added, and the molten gel was poured into a gel holder with combs and allowed to set.

- dNTP mixture

10mM	dATP
10mM	dCTP
10mM	dGTP
10mM	dTTP

2.2 Worm Culture Techniques

This section describes techniques used to culture *Caenorhabditis elegans* and is applicable to Chapters 4 and 5.

2.2.1 Strains and Culture Conditions

All worms lines used for experiments described in this thesis were derived from the Bristol-N2 wild-type strain of *C. elegans*, which were originally obtained from the *Caenorhabditis* Genetics Center (St. Paul, MN). The mutant worm lines used for experiments described in this thesis were generated (by Davies *et al.* 1999), using ethylmethane sulphonate (EMS), according to the protocol of Anderson (1995) (see Figure 2.1). The number of mutations induced at the DNA level for specific doses of EMS have previously been calibrated from experiments to measure forward mutation rates (Bejsovec & Anderson 1988) and suppressor-induced reversion rates (Hodgkin 1985, Kondo *et al.* 1990, Waterston 1981). Davies *et al.* (1999) exposed N2 worms to 50mM EMS for 4 hours. This dosage of EMS is expected to result in large numbers of $G/C \rightarrow A/T$ transitions (approximately 220 per haploid genome), plus a small number of other point mutations and small deletions. The estimate of 220 point mutations was converted to a conservative estimate of ~45 mutations per genome that would be deleterious under natural conditions, by incorporating information on the size of

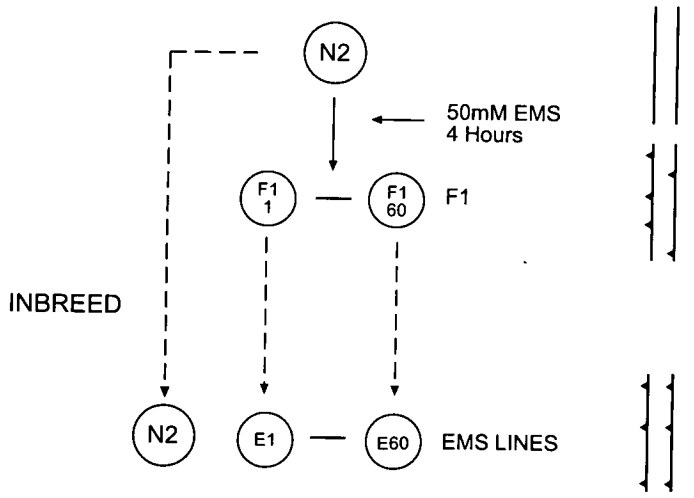


Figure 2.1: Illustration of the procedure used by Davies *et al.* (1999) to generate inbred EMS-mutant lines. The wild-type strain was exposed to EMS and a selection of (60) F1 offspring were inbred to homozygosity. After this time, each inbred line is expected to be homozygous for half the mutations present in the heterozygous state in the original F1 individual. The vertical lines on the right show the distribution of mutations present on sample homologous chromosomes for the wild-type line (N2), an F1 individual and a single inbred line (triangles represent individual mutations). N2 worms are assumed to have no mutations, whereas F1 individuals contain heterozygous induced mutations on both chromosomes. After inbreeding the EMS lines are homozygous for a sample of the mutations present in the F1.

the *C. elegans* genome, the percentage of this genome that is protein coding, and the level of evolutionary constraint within protein coding sequences. After mutagensis, 60 independent EMS-lines (labelled E1–E60) were bred towards homozygosity by transfering one hermaphrodite to a new plate each generation. This method is expected to minimise the effects of natural selection, so that the number of mutations fixed should be equal to the number of mutations induced per haploid genome. During this procedure four worms lines were lost, leaving 56 independent EMS-induced homozygous mutant lines, which were then frozen at −85°C. For a full description of the methods see Davies *et al.* (1999).

Unless otherwise stated live cultures of *C. elegans* were maintained using standard techniques (Sulston & Hodgkin 1988), described below. Worms were fed with *Escherichia coli* strain OP50. Suspensions of *E. coli* were produced by seeding 10ml of LB medium (see Section 2.2.6) with a single colony of *E. coli* (using

aseptic technique), and then incubating the medium at 37°C overnight. The bacterial suspension was aliquoted into sterile 2ml tubes and stored at 4°C to restrict further growth. This suspension was then used to feed *C. elegans* by creating seeded agar plates. 30μl of *E.coli* suspension was pipetted onto individual 3.5cm MYOB agar plates and the plates were incubated at room temperature overnight to allow the solution to soak into the agar and a bacterial lawn to grow. Worms were transferred onto seeded agar plates using a flattened platinum wire pick, sterilised in a blue bunsen burner flame and incubated at 20°C until they needed to be transferred to fresh agar plates (when either the food source was scarce or the plate became overcrowded). At 20°C worms needed to be transferred to fresh plates approximately every five to seven days.

2.2.2 Alkaline Hypocholrite Cleaning

Alkaline hypochlorite (see Subsection 2.2.6) kills larval and adult worms and the vast majority of fungal and bacterial contaminants but *C. elegans* eggs are resistant. This technique can therefore be used to obtain uncontaminated, viable eggs from a contaminated worm line. 7μl of fresh alkaline hypochlorite was pipetted onto a fresh seeded agar plate near the edge of the bacterial lawn. A single gravid hermaphrodite from a contaminated plate was then transferred into the solution. The alkaline hypochlorite solution should absorb into the agar before the eggs hatch. Once hatched, young larval worms were removed and transferred onto fresh plates as soon as possible to reduce the chances of carry over contamination.

2.2.3 Strep Cleaning

Streptomycin inhibits the growth of many bacterial and fungal contaminants but does not prevent the growth of *E. coli* OP50. Streptomycin was added to agar plates when strains of *C. elegans* were removed from storage at −85°C, since thawed tubes often contain contaminants. One millilitre of streptomycin sulphate stock solution was added to one litre of cooled molten MYOB agar (see Section 2.2.6) before it started to solidify. The agar was then gently mixed before pouring as usual. At this concentration, the growth of many contaminants is inhibited, but the growth of *E.coli* OP50 is not.

2.2.4 Cryopreservation of Worm Lines

Worms lines can be stored indefinitely at −85°C (Sulston & Hodgkin 1988). In order to obtain a culture of worms suitable for freezing, five adult worms were placed onto each of two medium sized (5.5cm) agar plates. These plates were incubated at 20°C for approximately five to six days, until they contained predominantly starving L1 and L2 stage larvae. The worms on each plate were collected by pipetting 1ml of M9 buffer over the surface of the agar and adding the resultant worm suspension to 1ml of freezing solution in a sterile 2ml tube. Each tube was then mixed and the contents aliquoted into four sterile 0.5ml cryogenic tubes. The tubes were then labelled and immediately placed at −85°C in styrofome boxes. One day later, the first and last vials were thawed, poured onto seeded medium sized agar plates and incubated at 20°C for several days, to check whether they contained viable worms.

2.2.5 Generation of Males

C. elegans males have one X chromosome (XO) as opposed to hermaphrodites which have two (XX). The normal rate of male production due to non-disjunction of the X chromosomes is ~1/1000 (Hodgkin 1988), and this rate is increased when worms are incubated at higher temperatures. In order to generate males, a few young hermaphrodites were taken and placed on medium sized agar plates at 26°C. After one generation the offspring were examined for males. Males can be detected on the basis of thier different morphology, they are more tapered and do not contain eggs. If any males were present, they were transferred onto fresh agar plates with non heat-shocked hermaphrodites from the same line in the ratio four to six males to one hermaphrodite. These plates were then incubated at 20°C for several days until crossed offspring had been produced. The male offspring can be used to set up further crosses (in the same ratio as before) to maintain or increase the number of male worms.

2.2.6 Solutions

- MYOB agar

  7.4g Pre-mix (see below)
  20g Agar

100g Pre-mix:

27g	NaCl
7.4g	Tris HCl
3.2g	Tris OH
62.2g	Bacto-tryptone
0.108g	Cholesterol

Distilled water was added to 1 litre. The agar was then autoclaved and allowed to cool to ~55°C before pouring. After pouring, plates were covered and stored at ~4°C, either before or after seeding with *E. coli*.

- LB (Luria-Bertani) medium for *E. coli*

10g	Bacto-tryptone
5g	Bacto-yeast extract
10g	NaCl

ddH_2O was added to 1 litre then the solution was autoclaved.

- Alkaline hypochlorite

2ml	Fresh Clorax bleach or equivalent (4–6% NaOCL)
5ml	1M NaOH

A fresh alkaline hypochlorite solution was made for each use.

- M9 buffer

6g	Na_2HPO_4
3g	KH_2PO_4
5g	NaCl

Distilled water was added to 1 litre, then the buffer was autoclaved. Once cooled, 1ml of 1M $MgSO_4$ was added using aseptic technique.

- Freezing solution

5.85g	NaCl
6.8g	KH_2PO_4
300g	Glycerol
5.6ml	1M NaOH

Distilled water was added to 1 litre, the solution was autoclaved, and then 3ml of 0.1M $MgSO_4$ was added using aseptic technique.

- Streptomycin sulphate stock solution

50mg	Streptomycin sulphate
1ml	ddH_2O

3 Quantifying Functional Constraint in *Drosophila* Non-Coding DNA

The work described in this Chapter forms part of a published paper (Halligan *et al.* 2004).

3.1 Introduction

Non-coding DNA constitutes the majority of multicellular eukaryotic genomes from around 70% in *C. elegans* (The *C. elegans* Sequencing Consortium 1998) up to approximately 95% in *Homo sapiens* (Lander *et al.* 2001). Although non-coding DNA is known to be functionally important in many respects, for example, in the regulation of gene expression, replication, chromosome packaging and recombination, the fraction which shows functional constraint at the level of the DNA sequence is still unclear. Furthermore, since such a large fraction of the multicellular eukaryotic genome is non-coding, quantifying the functional importance of non-coding DNA is vital for determining the genome wide mutation load due to deleterious mutations, and this is important in many aspects of evolutionary theory including the maintenance and evolution of sexual reproduction, the evolution of diploidy and the evolution of senescence (see Charlesworth & Charlesworth 1998). It has also been hypothesised that since mammals possess only two– to three–fold more proteins than *C. elegans* (but potentially more functional non-coding DNA), the greater complexity of mammals could be due to the extra functionality of intronic and intergenic DNA (Shabalina *et al.* 2001).

If a gene is defined as a locus of cotranscibed exons then non-coding DNA can be divided into two different classes: intronic (between the exons within a gene) and intergenic (between genes). Intergenic DNA can then be further subdivided into 5' UTRs (untranslated regions), 3' UTRs, and the section in between (see Figure 3.1). There are some *a priori* reasons for believing there may be functional constraint

within some of these non-coding DNA regions. For example, promoters are known to be situated upstream of the first exon, and enhancers have been found in a number of genes and gene clusters, and have been shown to be conserved (Blackman & Meselson 1986, Fisher & Maniatis 1986, Dickinson 1991, Cavener 1992, Oeltjen *et al.* 1997, Gottgens *et al.* 2000, Hardison 2000). One study has also shown that matrix-scaffold attachment regions are conserved in non-coding DNA (Glazko *et al.* 2003). Untranslated regions may be under functional constraint themselves, for example, Duret *et al.* (1993) found that the 3' UTR sequence in several classes of vertebrate is highly conserved, and some studies have also shown that pre-messenger RNA (pre-mRNA) structure may be under selection (Stephan & Kirby 1993, Kirby *et al.* 1995, Leicht *et al.* 1995). Introns are known to contain 5' and 3' splice sites, which are necessary for correct splicing of the mRNA and, therefore, would be expected to be conserved (Sharp 1994, Majewski & Ott 2002). Finally, Kondrashov & Shabalina (2002) found that some of the non-coding DNA conserved between human and murine genomes corresponds to known functional consensuses.

Analyses of non-coding DNA for individual gene clusters and particular parts of the genome give some idea of the constraints operating outside of coding regions (den˜ Dunnen *et al.* 1989, Liew *et al.* 1990, Epp *et al.* 1993, Hardison & Miller 1993, Koop & Hood 1994, Koop *et al.* 1996, Jan *et al.* 1997, Oeltjen *et al.* 1997), although the results from these analyses seem to give somewhat different estimates of the level of constraint in non-coding regions flanking loci (see Koop 1995). Some analyses have found evidence for non-coding regions being substantially more divergent than nearby coding regions (den˜ Dunnen *et al.* 1989, Hardison & Miller 1993, Ludwig & Kreitman 1995), whereas some have found them to be highly conserved (Liew *et al.* 1990, Epp *et al.* 1993, Koop & Hood 1994, Oeltjen *et al.* 1997). However, it is hard to draw any general quantitative conclusions from these analyses about the average level of functional constraint operating on non-coding DNA as a whole, since they are restricted to particular genes, gene clusters and sections of the genome.

There are several analyses, however, that have attempted to study general patterns of constraint by sampling data from the whole genome, and a selection of these studies are discussed below. There is now good evidence to suggest that at least some non-coding DNA is conserved; for example, a comparison of the human and mouse genomes has revealed that approximately 5% of DNA is conserved (although variation in mutation rates may bias this estimate). This is suggestive of at least

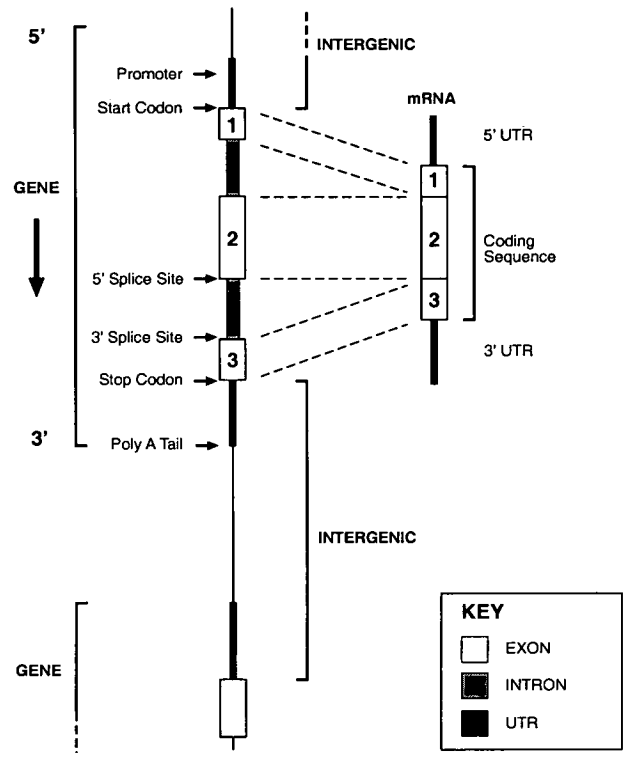


Figure 3.1: Diagram showing the basic intron/exon structure of a gene. Non-coding DNA can be divided into intronic or intergenic and intergenic DNA can be subdivided into 5' UTRs, 3' UTRs and the section in between. A gene is transcribed in the 5' to 3' direction, and the introns spliced out to form a mature messenger RNA (mRNA) molecule. The mRNA is subsequently translated into a protein.

some constraint in non-coding regions, given that only 2% of the mouse/human genome corresponds to coding sequences (Waterston *et al.* 2002). This conclusion is supported by several other recent studies which have identified large numbers of conserved non-coding DNA blocks in mammals (Shabalina *et al.* 2001, Dermitzakis *et al.* 2002, 2003, Thomas *et al.* 2003) and *Caenorhabditis* (Shabalina & Kondrashov 1999). However, the relative functional constraint operating in intronic *vs.* intergenic sequences generally has not been fully established, and it still remains unclear how functional constraint varies with distance from exonic boundaries.

There is some evidence to suggest that constraint is weak in introns and similar to that at four-fold sites and pseudogenes (Li & Graur 1991, Li 1997). One study has even suggested that the majority of intronic sites could be evolving faster than four-fold synonymous sites in rodent genes (Keightley & Gaffney 2003). On the other hand, there is evidence to suggest that constraint in introns might be substantial. In an analysis using 75 mouse and human gene pairs, Jareborg *et al.* (1999) looked for colinear-conserved blocks. It was found that 23% of the length of the introns was covered by conserved blocks (>60% identical). Assuming that the conservation observed is due to functional selection, then, on average, 19% of intron sequences have some functional role. Bergman & Kreitman (2001) found that, on average, approximately 22–26% of the intronic sequences studied were conserved in *Drosophila* (22–26% are within blocks of >70% identity), a similar level to that observed in intergenic DNA. Similarly, Shabalina & Kondrashov (1999), found that constraint in intergenic sequences was comparable to that in intronic sequences, with approximately 17% of nucleotides being conserved in *Caenorhabditis* species. It has also been demonstrated that intronic sequences can be involved in the regulation of gene expression and, where this has been observed, the sequences are conserved (Oeltjen *et al.* 1997).

The majority of genome-wide studies to assess levels of conservation or functional constraint in intergenic sequences agree, in that they suggest that at least some intergenic DNA is conserved. Some studies have even suggested that the level of constraint in intergenic sequences may be similar to that in coding sequences, if not stronger (Jareborg *et al.* 1999). Many genome-wide analyses of constraint in intergenic DNA have concentrated on the flanking regions of known genes, where functionally important regions are expected to be located. These have all found evidence for high levels of conservation near to genes (Duret *et al.* 1993, Jareborg

et al. 1999, Shabalina *et al.* 2001, Webb *et al.* 2002, Keightley & Gaffney 2003), a conclusion which is confirmed by the low levels of sequence diversity near genes (Li & Sadler 1991). However, it is still unclear whether 3' or 5' sequences are more conserved and whether these patterns are consistent across species. Some studies have found that 5' sequences are more conserved on average (Li & Graur 1991, Li & Sadler 1991), others have found the opposite (contrary to what might be expected) (Duret *et al.* 1993, Webb *et al.* 2002, Jareborg *et al.* 1999), whilst another study found the pattern of conservation to be fairly symmetrical at both ends of coding sequences (Keightley & Gaffney 2003).

There are a number of different methods that can be used to infer the level of functional constraint in non-coding DNA. All are based on the fact that negative selection constrains changes in functionally important regions, so important sections of DNA will show lower rates of evolution, or will be conserved between divergent species. Rough estimates of the level of conservation in non-coding DNA can be obtained from DNA hybridisation studies; for example, only 2% of human DNA is thought to encode proteins, but Britten (1986) found that approximately 13% of non-repetitive DNA hybridises between human and rat. This suggests that at least some non-coding DNA is under functional constraint. Most approaches to studying constraint in non-coding DNA, however, rely on comparing the actual DNA sequences of closely or distantly related species.

One approach to analysing the level of functional constraint in non-coding DNA is to look for conserved blocks between distantly related species; this method was used by Shabalina & Kondrashov (1999) and subsequently by Shabalina *et al.* (2001) and Glazko *et al.* (2003). If it is assumed that sites are either freely evolving or highly constrained, (which may be a reasonable approximation, see Shabalina & Kondrashov 1999), then sites will either be saturated or fully conserved in comparisons between distantly related species (although the analysis of Shabalina *et al.*, compared *Mus musculus* and *Homo sapiens*, which are not divergent enough to have reached saturation at neutral sites yet). Under these assumptions, any sections that are alignable (forming part of a conserved block, called "HITS") represent functionally important regions, and the fraction of conserved sites can be calculated to give an estimate of the level of functional constraint for a given section of non-coding DNA. A similar approach was used by Thomas *et al.* (2003) and Dermitzakis *et al.* (2003), although clearly if the assumption that sites are either totally conserved or totally neutral is

incorrect, and there are in fact many sites under moderate levels of constraint, then the estimated level of constraint will vary depending on how closely related the two compared species are. A further complication arises in comparisons between distantly related species if there is variation in the mutation rate around the genome, since this can generate variation in conservation that is not due to differences in the level of functional constraint (Clark 2001). There is also evidence that variation in mutation rates could be considerable, for example in a study by Bachtrog *et al.* (2000), significant variation in mutation rates was found between microsatellite repeat motifs in *Drosophila melanogaster.*

An alternative approach to assessing functional constraint in non-coding DNA is to look at more closely related species, *e.g. Mus musculus* and *Rattus norvegicus* (divergence of ~0.18 per synonymous site), for which it is possible to align almost all non-coding DNA reasonably well. The rate of evolution in a nearby putatively neutral sequence can then be compared to the rate of evolution in the non-coding DNA in question. A lower rate of evolution suggests that the sequence is under negative selection, and the fraction of missing substitutions can be used to quantify constraint in the region (Kimura 1983). This approach has the advantage that it allows the level of constraint to be quantified by measuring differences in the rates of evolution and does not assume that sites are either totally conserved or totally free from selective constraint, an assumption that could be incorrect. Furthermore, since the putatively neutral sequences are adjacent to the functional segments, it is reasonable to assume that the mutation rates are equal (providing that variation in mutation rate occurs over a large scale). However, studies in mammalian species are complicated by the hyper-mutability of CpG dinucleotides, which show on average a 10- to 15-fold higher than average rate of substitution (Shen *et al.* 1994). This creates problems for the approach, because the frequency of CpG dinucleotides varies between coding and non-coding DNA, and between different categories of non-coding DNA (Chen & Li 2001, Hellmann *et al.* 2003, Subramanian & Kumar 2003). Rates of mutation in different sections of DNA may therefore be affected differently.

It is clear, therefore, that it has not yet been fully established which sections of non-coding DNA show the highest degree of sequence conservation due to functional constraint. Furthermore, few analyses have tried to quantify how the level of functional constraint varies around the genome and, in particular, how the level of constraint varies with distance from exonic boundaries in non-coding DNA.

In this chapter, constraint in sections of upstream, downstream and intronic DNA has been quantified using sequence comparisons between *Drosophila melanogaster* and *Drosophila simulans*. Additional upstream DNA from *Drosophila simulans* was sequenced in order to increase the numbers of comparisons. In these species, CpG dinucleotides are not hyper-mutable, making estimates of mutation rates less complicated. In addition, the synonymous divergence between *D. simulans* and *D. melanogaster* is 0.100 (SE 0.008) (Keightley & Eyre-Walker 2000), making alignments of neutrally evolving DNA reasonably reliable. MCALIGN, a maximum likelihood program to align non-coding DNA using a specific model of insertion-deletion (indel) evolution, was used to align non-coding DNA in an attempt to reduce the possibility of systematic biases in sequence alignments. This model has been previously parameterised for *Drosophila* species (Keightley & Johnson 2004).

Synonymous sites were initially used as a putative neutral standard for estimating constraint in adjacent non-coding DNA. This assumption of neutrality is not unrealistic, since several lines of evidence point to a relaxation in selection at synonymous sites in the two lineages. Firstly, there is evidence for a surge in the rate of preferred to unpreferred synonymous substitutions (Akashi 1995, 1996, Takano 1998, Begun 2001, McVean & Vieira 2001), possibly due to demographic changes reducing the efficacy of selection. Secondly, a population genetics analysis of the pattern of synonymous divergence suggests that there is currently no selection acting on codon usage in *D. melanogaster*, while the sister species, *D. simulans*, experiences only half the selection pressure for codon usage of their common ancestor (McVean & Vieira 2001). Additionally, weak selection of the magnitude thought to be acting on synonymous codon usage in *Drosophila* (Akashi 1995, 1996) is predicted to have only a small effect on substitution rates (Eyre-Walker & Bulmer 1995).

3.2 Materials and Methods

3.2.1 Compilation of Data

A dataset of homologous *Drosophila melanogaster* and *Drosophila simulans* genes was compiled by manually searching Genbank for *D. simulans* genes that had at least 200bp of coding sequence and at least one intron or at least 60bp of either upstream or downstream intergenic DNA. Coding sequences were blasted against the

mosquito (*Anopheles gambiae*) genome sequence to check that they were genuine coding sequences. Pseudogenes would not be expected to be conserved, given that the two dipteran insects are thought to have separated ~250 million years ago (Yeates & Wiegmann 1999), but the majority of protein coding sequences should still have identifiable homologues between the two species (Zdobnov *et al.* 2002). In some cases it was possible to extract introns but not 5' or 3' intergenic sequences from a gene, and in some cases only intergenic and not intronic data could be extracted. These sets of genes were compiled separately, so there are 2 datasets, hereafter referred to as the intronic dataset and the intergenic dataset, each containing different (but overlapping) sets of genes. Intergenic non-coding DNA was classified as either 5' or 3'. If the distinction was difficult because genes were too close together, assignment into either category was arbitrary. Introns were analysed either as complete or as partial sequences, after removal of putative splice control sequences. The splice control regions were defined as base pairs 1–6 at the 5' end and 1–16 at the 3' end, although the exact limits of the control sequences are not necessarily consistent across genes (Sharp 1994). See the Appendix (Tables A.1 and A.2) for lists of loci used.

Additional intergenic DNA sequences from *D. simulans* were obtained by sequencing the 5' flanking regions of genes for which the orthologous coding sequences were already available for both *D. simulans* and *D. melanogaster* on Genbank. Genes which only had a short length of available coding sequence in *D. simulans* were removed from the dataset (an arbitrary cut off of 200bp was used). Upstream DNA for genes that already had a reasonable amount available on Genbank was not sequenced (a cut off of 200bp of upstream sequence was used). Primers for sequencing were designed (using Primer Premier 5.00, Premier Biosoft International, Palo Alto, CA) to be approximately 650 to 700bp apart (based on the *D. melanogaster* sequence). Upstream primers were designed using *D. melanogaster* DNA; where possible upstream coding sequence was used but in the majority of cases they were designed from non-coding sequence. Downstream primers were always designed using the *D. simulans* coding sequence. For a list of primer sequences used, see Appendix, Table A.3. Genomic DNA for PCR reactions was prepared from a single male *D. simulans* fly (see Section 2.1.1) from a partially inbred line (Dsim A08, collected in Aswan and inbred for eight generation).

A combination of standard PCR (Section 2.1.2) and asymmetric PCR (Section 2.1.7) was used to amplify the appropriate section of DNA. If the primers failed

to amplify the appropriate section of DNA, the upstream primer was redesigned, as it would be expected that the downstream primer (designed from *D. simulans*) would match perfectly. Both primers were redesigned up to two more times but if the targeted section of DNA still failed to amplify, the gene was removed from the data set. In 18 out of 63 cases it was not possible to get sufficient amplification of the appropriate section of DNA. Purified PCR products were sequenced on both strands using an ABI prism BigDye® terminator cycle sequencing kit (Applied Biosystems, Foster City, CA) and run on an Applied Biosystems 3730 DNA Analyzer (Applied Biosystems, Foster City, CA). Sequences from each strand for each gene were then assembled using Sequencher 3.0 software (Gene Codes, Ann Arbor, MI) and alignments were checked by eye. The Genbank accession numbers for the 45 novel sequences obtained are AY459538–82.

3.2.2 Potential Bias in Sampling

There are two potential sources of bias in the methods used to obtain the sample of *D. simulans* loci. Firstly, it was not possible to sequence upstream DNA in 18 of the genes selected; since the upstream primer sequences were designed from the *D. melanogaster* sequence it is possible that the primers did not amplify the appropriate section and length of DNA because the *D. simulans* sequence differed. It is therefore possible that this subset of genes could have the least amount of upstream constraint, biasing the estimate of constraint upwards. Secondly, it is conceivable that there is a bias towards more highly expressed genes in the sample of coding regions that have been sequenced in *D. simulans* and it is possible that this bias may influence the estimates of constraint. Estimates of codon usage bias were used to test whether the non-random sample of *D. simulans* genes show any evidence of being more highly expressed than would be expected for a random sample, since it has been established that codon usage bias is highly correlated with expression level in *Drosophila* (Duret & Mouchiroud 1999). Estimates of the effective number of codons (Wright 1990) were calculated for each gene in the non-random sample (for both the intronic and intergenic datasets) and these were compared to the distribution for a random sample of 400 *D. melanogaster* genes. The *D. melanogaster* genes were selected by randomly choosing 400 genes from a list of all putative *D. melanogaster* genes that showed a significant match when blasted against the *Anopheles gambiae* genome sequence (see

Gilbert 2002 for a list of all matching loci). The two distributions of effective numbers of codons were tested to see if they were significantly different, by creating 10,000 bootstrapped datasets for both species. Estimates of the mean, variance, skewness and kurtosis for the bootstrapped datasets were then compared between species.

3.2.3 Alignment of DNA

Coding sequences are easily aligned across their entire length, even between fairly divergent species, because of the evolutionary pressures to conserve coding capacity. Given the low synonymous divergence between *D. simulans* and *D. melanogaster* of 0.100 (SE 0.008) (Keightley & Eyre-Walker 2000), reasonably unambiguous alignments of coding sequences could be obtained using CLUSTALW. All coding sequence alignments were subsequently checked by eye and corrected manually, using the translated sequence alignment to verify any potential ambiguities. However, non-coding DNA is more problematical to align because it may include indels (insertions / deletions), particularly large indels, and secondly because unlike coding DNA, the amino acid sequence cannot be used to aid alignment.

Non-coding DNA was aligned using MCALIGN (Keightley & Johnson 2004), a program which finds the most probable alignment according to a specific model of indel evolution. MCALIGN has already been parameterised for use with *Drosophila* (Keightley & Johnson 2004) and therefore the default parameters could be used for alignments. Unlike many other alignment algorithms (see McClure *et al.* 1994), MCALIGN does not use a linear gap penalty, an assumption that is not supported by empirical data (Gu & Li 1995). Instead, MCALIGN uses a specific model of indel evolution for the species being studied, and should, therefore, produce unbiased alignments (without a specific model of indel evolution, alignments may be biased; Thorne *et al.* 1991). The parameters of the model were derived by Keightley & Johnson (2004) based on the relative rates of indels to single nucleotide substitutions, and the distribution of indel lengths, in two closely related species of *Drosophila* (*D. simulans* and *D. sechellia*) for which alignments of non-coding DNA are unambiguous.

3.2.4 Distance Method for Calculating Constraint

Following distance-based methods for calculating constraint in coding DNA (Eyre-Walker & Keightley 1999), the method employed here uses rates of substitution at putatively neutral sites in a gene to predict the expected numbers of substitutions in an adjacent non-coding DNA segment, such as an intron or flanking sequence, assuming equal rates of mutation in the two sequences. It also assumes that both species behave in the same manner, this assumption is partially tested in Section 3.3.1. In this study both four-fold degenerate synonymous sites and introns (with splice control regions removed) have been used as the neutral standard. This method takes into account differences in base composition between the neutral standard and the adjacent non-coding sequence. It does not assume that the neutral standard is at equilibrium, but that it is evolving towards it neutrally. The expected numbers of substitutions (E) are compared to the observed numbers (O) to calculate constraint (C). For example, if $E = O$, the constraint in the non-coding segment is zero; if $O = 0$, constraint takes the value of 1. The method is only applicable to closely related species for which multiple hits can be safely ignored. It also assumes that both species are evolving similarly such that they have equal mutation rates.

In a pairwise comparison it is not possible to determine the direction of a particular substitution (*i.e.* whether a C↔T difference is due to a C→T or a T→C substitution), this would require a parsimony approach. However, it is possible to partition the total number of substitutions into the proportion that occur in any particular direction if As and Ts are grouped together and Gs and Cs are grouped together and the equilibrium base composition is known.

Let the rate of AT→GC substitution be u and the rate GC→AT substitution be v. The aim of this method is to derive equations to estimate u and v by measuring the total number of AT↔GC of mutations (this category of differences involves the following pairwise differences A↔G, A↔C, T↔G, and T↔C) and using estimates of the equilibrium G+C content and the current G+C content. Let f_e be the equilibrium G+C content of the sequence; this is the G+C content that the sequence would eventually reach. At equilibrium the number of GC→AT substitutions that occur per unit time ($f_e v$) equals the number in the opposite direction ($(1-f_e)u$), so $f_e v = (1-f_e)u$.

Rearranging, it is possible to get

$$u = f_e(v + u) \tag{3.1}$$

and

$$v = (1 - f_e)(v + u) \tag{3.2}$$

When a sequence is not at equilibrium, the number of GC→AT changes per unit time is equal to the rate of change of GC→AT (v) times the current G+C content (assuming that the time period is sufficiently short such that the G+C content does not change dramatically). Let f_a be the current G+C content, then the number of AT→GC ($X_{AT\rightarrow GC}$) changes is $u(1 - f_a)$, substituting the result from equation 3.1 into this equation gives:

$$X_{AT\rightarrow GC} = f_e(v + u)(1 - f_a) \tag{3.3}$$

Similarly, the number of GC→AT changes ($X_{GC\rightarrow AT}$) is vf_a, substituting in the result equation 3.2 gives:

$$X_{GC\rightarrow AT} = (1 - f_e)(v + u)f_a \tag{3.4}$$

An equilibrium is reached when the number of mutations in one direction equals the number of mutations in the opposite direction, *i.e.* when $f_e(v+u)(1-f_a) = (1-f_e)(v+u)f_a$, or when $f_e = f_a$.

The total number of mutations in either direction ($X_{AT\leftrightarrow GC}$) that occur per unit time when not at equilibrium is the sum of equations 3.3 and 3.4

$$X_{AT\leftrightarrow GC} = f_e(v + u)(1 - f_a) + (1 - f_e)(v + u)f_a \tag{3.5}$$

and rearranging for $(v + u)$ we get

$$(v + u) = \frac{X_{AT\leftrightarrow GC}}{f_e(1 - f_a) + (1 - f_e)f_a} \tag{3.6}$$

Substituting this into equations 3.1 and 3.2 we get

$$u = \frac{X_{AT\leftrightarrow GC} f_e}{f_e(1 - f_a) + (1 - f_e)f_a} \tag{3.7}$$

and

$$v = \frac{X_{AT\leftrightarrow GC}(1 - f_e)}{f_e(1 - f_a) + (1 - f_e)f_a} \tag{3.8}$$

This method makes it possible to obtain estimates of the rate of GC→AT and the rate of AT→GC substitution without having to resort to a parsimony approach (for which the necessary data is currently lacking). However, it cannot be inferred whether an observed G↔C difference is due to a G→C mutation or C→G mutation. Similarly, polarity cannot be assigned to any observed A↔T differences. Therefore only 4 different rates (i = 1...4) can be calculated; two pairwise rates (A↔T and G↔C) and two directional rates (AT→GC and GC→AT).

The predicted (expected) number of substitutions in a segment of non-coding DNA is the sum of all rates multiplied by the number of sites at which those mutations can occur, *i.e.*

$$E = \sum_{i=1}^{4} K_i M_i \tag{3.9}$$

where M_i is the number of sites in the non-coding segment corresponding to the rate (K_i) of type i. The observed number of differences in the segment, O, is the number of nucleotide differences in the non-coding segment. Constraint for a segment is given by $C = 1 - O/E$; or for several segments, it is

$$C = 1 - \sum \frac{O_j}{E_j} \tag{3.10}$$

where the summation is carried out over (j) segments. Standard errors of O, E, and C are calculated by bootstrapping the data, by gene (Eyre-Walker & Keightley 1999).

3.2.5 Simulations to Verify Constraint Calculation

Simulations were conducted to check that the distance method used to calculate constraint does not show a systematic bias in the estimation of constraint when a mutation bias is simulated (*i.e.* if the equilibrium G+C content is not 0.5), or when the current G+C content of the non-coding segment is different from the current G+C content of the putatively neutral DNA (in this case, four-fold synonymous sites of coding DNA). To do this, sections of coding DNA, consisting of 30,000 random four-fold degenerate codons, and sections of adjacent non-coding DNA, consisting of 30,000 random bases, were simulated. Both sequences were then duplicated, and mutated independently. When mutating a sequence the total number of mutations per sequence was taken from a poisson distribution and these mutations were randomly distributed over the sequence. For each site that was mutated the given base was changed to one of the other 3 bases with certain probabilities. These probabilities were altered to simulate a mutation bias. For coding sequences, only four-fold sites were mutated. Indels were ignored in the simulations, making alignments of DNA unnecessary, and the divergence at neutral sites was set to be 0.05, allowing multiple hits to be ignored. The G+C content of the non-coding DNA was set to be 0.4 in all simulations prior to mutation, a mutation bias was also modelled such the the equilibrium G+C content would also be 0.4. The G+C content of the four-fold sites was varied between 0.4 and 0.8 in intervals of 0.2 and the same mutation bias was modelled such that four-fold sites would also have an equilibrium G+C content of 0.4. The level of functional constraint (one minus the fraction of mutations that are accepted) in the non-coding DNA was varied between 0.2 and 0.8, also in intervals of 0.2.

Constraint was then estimated for the simulated non-coding DNA by comparing the observed number of substitutions to the expected number of substitutions, calculated from the neutrally evolving four-fold synonymous sites in the simulated coding DNA, using the distance method outlined above. Ten simulations per parameter combination were carried out, and the mean and standard deviations of the observed number of substitutions, the expected number of substitutions and the estimates of constraint for the ten replicates were tabulated.

3.3 Results

3.3.1 Data Summary

Two separate datasets were compiled for analysis in this experiment. The first consisted of intergenic sequences with adjacent coding sequences (referred to as the intergenic dataset). Additional novel upstream sequences for *D. simulans* were obtained by sequencing the flanking regions of genes for which the coding sequences were available on genbank. In total 45 novel upstream sequences were obtained. The average length sequenced was 652bp (although some of this length overlapped with previously sequenced DNA). An average of 647bp of upstream DNA over all intergenic sequences was usable for a total of 80 loci and an average of 331bp of downstream DNA was available for a total of 42 loci (see Table 3.1). The intronic dataset consisted of a total of 91 loci, each with an average of 1020bp of coding sequence and an average of 177bp intronic sequence per locus.

G+C content was calculated for all the different types of sequences available (coding sequences, four-fold sites of coding sequences, upstream sequences, novel upstream sequences, downstream sequences and introns). Most non-coding DNA had a G+C content of ~0.4 which was substantially lower than that for four-fold sites (~0.7). Divergence between *D. melanogaster* and *D. simulans* sequences was also calculated (without correcting for multiple hits) for all types of sequence (Table 3.1). Divergence was highest for introns, followed by four-fold sites, intergenic sequences and then coding sequences (all sites). This is suggestive of a lack of any functional constraint in intronic sequences, and some degree of functional constraint in intergenic sequences. However due to the difference in G+C content between types of sequence it is difficult to infer the level of constraint from estimates of divergence alone.

In the distance method for calculating constraint detailed in Section 3.2.4 we assume that both species are behaving in the same manner. More precisely, it is assumed that the rates u and v are the same in both species. This assumption can be tested by examining the six types of possible pairwise differences between the species for the sequences analysed. If the rates are the same we would expect that for each of the six types of pairwise difference, the frequency of the two constituent bases would be the same in both species. For example, if we consider the pairwise difference AT, we would expect to have the same frequency of A/T differences (A in

Table 3.1: Estimates of G+C content and divergence (not correcting for multiple hits) for coding sequences (CDS) four-fold sites, 5' sequences, novel 5' sequences (*i.e.* those sequenced as part of this thesis), 3' sequences and introns.

Sequences	Loci	bp / locus	G+C content	Divergence (SE)
CDS (all sites)	91	1020	0.550	0.0342 (0.00201)
CDS (four-fold sites)	91	104	0.663	0.0861 (0.00392)
5'	80	647	0.404	0.0687 (0.00375)
novel 5'	45	568	0.401	0.0693 (0.00504)
3'	42	331	0.370	0.0647 (0.00822)
Introns	91	177	0.370	0.104 (0.00541)

The coding sequences analysed are taken from the intronic dataset. Introns for any given locus were joined and anlaysed as one sequence.

D. melanogaster and T in *D. simulans*) as T/A differences (T in *D. melanogaster* and A in *D. simulans*). This hypothesis was checked for each type of pairwise difference, by calculating the total number of each type of possible difference across all sequences and checking that for a given pairwise difference, the number of differences in each direction did not significantly differ from the expectation of equal numbers under a χ^2 test. For example it was checked that the total number of A/T differences and the total number of T/A differences did not significantly differ from the expectation (of equal numbers) across all the intronic sequences in the dataset. This test was carried out separately for the intronic sequences, upstream sequences, downstream sequences and the four-fold sites of coding sequences that were used in the analysis.

For the intronic sequences, the numbers of each type of difference within each given pairwise difference did not differ significantly from the expectation for any of the pairwise differences (after correcting for multiple comparisons, for an average of 209 differences for each type of pairwise difference). This was also true of the downstream sequences (which had an average of 139 differences for each type of pairwise difference). For upstream sequences there was an average of 570 differences for each type of pairwise difference across all the sequences and two of the pairwise difference showed a significant departure from the expectation. These were AG differences (there was an excess of As in *D. melanogaster*, $p < 0.0005$) and CT differences (there was an excess of Ts in *D. simulans*, $p < 0.05$). For the four-fold sites (from the coding sequences in the intronic dataset) there was an average of 137

differences for each type of pairwise difference and three of the pairwise differences showed a significant departure from the expectation even after correcting for multiple tests. There was a significant excess of As in *D. melanogaster* for AC differences ($p < 0.05$), a significant excess of As in *D. melanogster* for AG differences ($p < 0.05$) and a significant excess of Ts in *D. simulans* ($p < 0.01$) for CT differences. All the changes that are significant reflect mutations between AT and GC in one of the species, and will therefore give rise to differences in the GC content between the two species (although the effect will be slight due to the relatively low divergence between the species). This slight bias in the number differences within certain types of pairwise differences for some of the classes of sequence studied could lead to a possible bias in our estimates of constraint although it is hard to predict in which direction the results will be biased without knowing the polarity of the mutations involved (and this would require an outgroup which is unavailable for the majority of sequences studied at present).

3.3.2 Distributions of Codon Usage Bias

The sample of genes for which coding and non-coding sequence is available in *D. simulans* (91 genes for the intronic dataset and 85 for the intergenic dataset) may be biased in terms of expression level, since more highly expressed genes are more likely to have been sequenced. It has already been established, that expression level in *Drosophila* is highly correlated with codon-usage bias (Duret & Mouchiroud 1999). This bias could therefore be tested by comparing the distribution of estimates of effective number of codons for the sample of *D. simulans* genes to that of a random sample of *D. melanogaster* genes. The distributions of effective number of codons appear to be broadly similar (see Figure 3.2) (although due to the low number of loci available the distributions for *D. simulans* are somewhat noisy). Differences in the distribution of codon usage bias between the nonrandom sample of *D. simulans* genes (intron dataset) and a random sample of 400 *D. melanogaster* genes were tested by bootstrapping estimates of the effective number of codons. No significant difference in the mean effective number of codons ($p = 0.126$), and no significant difference in the skewness of the distributions ($p = 0.485$) was found. However, the codon usage bias estimates for the sample of *D. simulans* genes are significantly more variable ($p < 0.001$), and the distribution is significantly more platykurtic (excess

kurtosis and therefore a flattened shape) (p = 0.018), than the random sample of textitD. melanogaster genes. Similar results are found when the *D. simulans* intergenic dataset was analysed, although the difference in kurtosis between the distributions was marginally non-significant (p = 0.056).

These results imply that, although the dataset may contain more genes with very low and very high codon usage bias than would be expected by chance, there is probably no bias towards more highly expressed genes in the sample.

3.3.3 Results of Simulations to Verify Constraint Calculation

In order to verify the functionality of the distance method to calculate constraint whilst taking into account differences in G+C content, a number of simulations were carried out (see Table 3.2). If the method is unbiased, the estimated level of constraint should equal the simulated level of constraint, whether or not there is a mutation bias, and/or a difference in the G+C content of the neutral and non-coding sections of DNA. In all cases the estimated values of constraint are very similar to the simulated values, suggesting that the method accurately corrects for differences in G+C content. However, there is a consistent bias for estimates of constraint to be lower than the simulated values, although the difference is very small in all cases. This is due to the fact that multiple hits were ignored, since they would cause a greater reduction in the expected number of substitutions than the observed number of substitutions (if there is any functional constraint).

3.3.4 Constraint in Intronic DNA sequences

The level of constraint in 91 intron sequences of *D. melanogaster* and *D. simulans* was calculated using four-fold synonymous sites from neighbouring exons as a putatively neutral standard. The distance-based approach described in Section 3.2.4 was used to estimate the expected number of substitutions in a stretch of intronic DNA, and this was compared to the observed number. Constraint was then calculated as the fraction of missing substitutions (see Methods, Section 3.2). It was assumed that the equilibrium G+C content (f_e) was the same as the current G+C content of the intronic sequences in the dataset (0.370), which is markedly different from the G+C content of the putatively neutral four-fold sites (0.663) used to calculate the expected number of

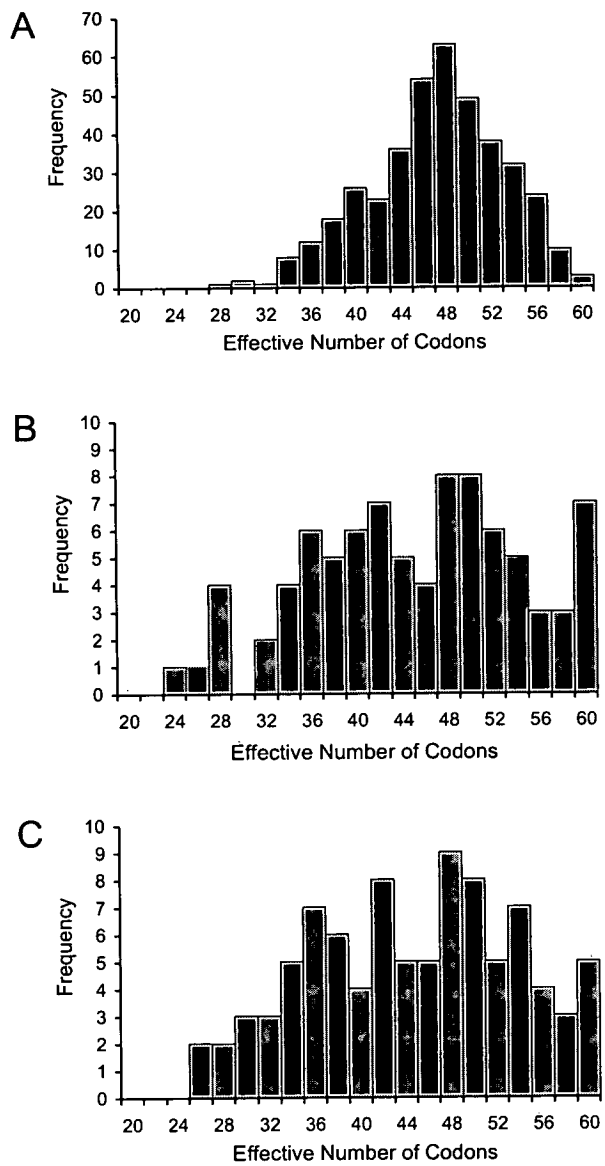


Figure 3.2: Distribution of effective number of codons for a random sample of 400 *D. melanogaster* genes (A), compared to the distributions of effective number of codons for the sample of *D. simulans* genes used to analyse constraint in intergenic (B) and intronic (C) DNA.

Table 3.2: Simulation results to verify the functionality of the pairwise distance method to calculate constraint.

Simulated Values		Estimated Values		
C	G+C (four-fold)	Obs (SD)	Exp (SD)	*C* (SD)
0.2	0.4	1160 (11.3)	1440 (7.42)	0.198 (0.0105)
0.4	0.4	880 (16.5)	1450 (8.08)	0.392 (0.0125)
0.6	0.4	591 (16.9)	1450 (6.56)	0.594 (0.0115)
0.8	0.4	299 (14.1)	1450 (6.30)	0.794 (0.00977)
0.2	0.6	1170 (15.2)	1450 (14.1)	0.194 (0.0155)
0.4	0.6	873 (24.8)	1440 (11.8)	0.395 (0.0179)
0.6	0.6	597 (20.7)	1450 (16.0)	0.588 (0.0163)
0.8	0.6	296 (11.1)	1450 (8.21)	0.795 (0.00829)
0.2	0.8	1170 (16.9)	1440 (29.1)	0.191 (0.0229)
0.4	0.8	883 (13.9)	1440 (30.2)	0.386 (0.0157)
0.6	0.8	586 (12.1)	1440 (24.3)	0.594 (0.00686)
0.8	0.8	290 (20.0)	1440 (24.4)	0.799 (0.0161)

Means (and standard deviations) of the calculated number of observed (obs) and expected (exp) substitutions alongside estimates of constraint (*C*), are shown (standard deviations are shown in brackets). Ten replicates were carried out for each parameter combination. Different values of constraint in the non-coding section (from 0.2 to 0.8) and different G+C contents of the neutrally evolving four-fold sites (0.4, 0.6 and 0.8) were simulated. The equilibrium G+C content, and current G+C content of the non-coding sites was set at 0.4 in all cases.

Table 3.3: Observed (obs) and expected (exp) numbers of substitutions and estimates of constraint (C) in intronic DNA sequences with and without splice sequences omitted.

Data Set	Loci	bp / locus	Obs (SE)	Exp (SE)	C (SE)
Complete	91	230	16.2 (1.68)	14.8 (1.54)	−0.0973 (0.0574)
Splice sites omitted	91	191	13.8 (1.60)	11.8 (1.42)	−0.172 (0.0710)

Estimates were calculated using four-fold synonymous sites as a neutral standard. Standard errors (in brackets) were calculated by bootstrapping estimates by gene.

substitutions. It has already been established, however, that the distance method can accurately take into account differences in current G+C content, if both sequences are evolving towards the same f_e.

Constraint was calculated for whole intron sequences, and for intron sequences with putative splice control regions removed (6 bp at 5' end and 16 bp at the 3' end) using four-fold sites as the neutral standard (see Table 3.3). Constraint is non-significantly different from zero when complete introns are analysed ($p = 0.0963$) and significantly negative ($p = 0.00836$) when splice sequences are omitted. Introns therefore appear to be evolving faster than four-fold degenerate sites, suggesting that four-fold sites are not completely neutral. These results also suggest that splice sequences are reasonably highly conserved, since the difference between rates of evolution in four-fold sites and introns becomes more pronounced when they are removed.

To further investigate the level of constraint in putative splice control regions, constraint was calculated for the putative splice control regions defined above using both four-fold sites and introns (with splice sites removed) as the neutral standard (see Tables 3.4 and 3.5). Constraint is absolute for the 5' GT splice site and the 3' AG splice site, when either introns or four-fold sites are used as the putative neutral standard. Constraint is also high for bases pairs 3–6 from the 5' end of introns, but is very weak, if not absent, for base pairs 3–16 from the 3' end (constraint is significantly negative ($p = 0.0254$) when four-fold sites are used as the neutral standard, but non-significantly different from zero when introns are used). This is somewhat surprising, since it has been suggested that these base pairs correspond to a reasonably well conserved polypyrimidine tract (Sharp 1994). Where constraint is not absolute, using

Table 3.4: Observed (obs) and expected (exp) numbers of substitutions and estimates of constraint (C) in nucleotides close to the 5' and 3' ends of introns, using four-fold synonymous sites as a neutral standard.

Type	Base pairs	Loci	Obs (SE)	Exp (SE)	C (SE)
5'	1–2	91	0.00 (0.00)	0.293 (0.0210)	1.00 (0.00)
	3–4	91	0.143 (0.0365)	0.2771 (0.0206)	0.484 (0.126)
	5–6	91	0.0540 (0.0241)	0.295 (0.0215)	0.818 (0.0776)
3'	1–2	91	0.00 (0.00)	0.292 (0.0217)	1.00 (0.00)
	3–16	91	2.26 (0.210)	1.92 (0.141)	−0.181 (0.0980)

Standard errors (in brackets) were estimated by bootstrapping by gene.

Table 3.5: Observed (obs) and expected (exp) numbers of substitutions and estimates of constraint (C) in nucleotides close to the 5' and 3' ends of introns, using introns with splice control regions removed as a neutral standard.

Type	Base Pairs	Loci	Obs (SE)	Exp (SE)	C (SE)
5'	1–2	91	0.00 (0.00)	0.366 (0.0248)	1.00 (0.00)
	3–4	91	0.143 (0.0367)	0.350 (0.0231)	0.591 (0.105)
	5–6	91	0.0549 (0.0240)	0.365 (0.0259)	0.850 (0.0649)
3'	1–2	91	0.00 (0.00)	0.366 (0.0251)	1.00 (0.00)
	3–16	91	2.25 (0.199)	2.38 (0.161)	0.0531 (0.0745)

Standard errors (in brackets) were estimated by bootstrapping by gene.

introns as the putative neutral standard gives higher estimates of constraint, due to the fact that they appear to be evolving faster than four-fold sites.

3.3.5 Constraint in Intergenic DNA sequences

Constraint in intergenic DNA sequences of *D. melanogaster* and *D. simulans* was calculated using the same distance-based two-lineage approach as above. Both four-fold synonymous sites and intronic sites (without splice control regions) were used as putative neutral standards when calculating constraint. It was assumed in all cases that the equilibrium GC content was the same as that for intronic sites (0.370).

In contrast to the findings for intronic DNA, there appears to be a substantial amount of constraint in upstream and downstream sequences, and this implies the

Table 3.6: Observed (obs) and expected (exp) numbers of substitutions along with estimates for constraint (C) in intergenic DNA sequences using four-fold synonymous sites as a neutral standard.

Type	Base Pairs	Loci	bp / locus	Obs (SE)	Exp (SE)	C (SE)
5'	1–100	80	98.7	5.55 (0.363)	7.97 (0.462)	0.301 (0.0559)
	101–200	75	98.3	5.45 (0.426)	7.64 (0.448)	0.285 (0.0577)
	201–300	71	98.4	6.34 (0.510)	7.44 (0.477)	0.145 (0.0787)
	301–400	68	98.1	7.05 (0.641)	6.97 (0.407)	−0.0148 (0.100)
	401–500	65	94.7	6.72 (0.531)	6.53 (0.412)	−0.0322 (0.0981)
	501–700	56	139	8.28 (0.934)	8.70 (0.776)	0.0426 (0.118)
3'	1–100	42	94.6	5.80 (0.698)	7.12 (0.576)	0.184 (0.0894)
	101–200	31	87.7	5.37 (0.774)	6.25 (0.786)	0.131 (0.146)
	201–300	21	91.8	5.86 (0.794)	6.41 (0.977)	0.0697 (0.163)
	301–500	18	150	7.12 (1.29)	11.3 (1.96)	0.347 (0.183)

Standard errors (in brackets) were estimated by bootstrapping by gene.

action of negative selection (see Tables 3.6 and 3.7). Since intronic sites seem to be evolving faster than four-fold sites, estimates of constraint using intronic sequences as the neutral standard are higher than those obtained when four-fold sites are used. Nevertheless, there is significant positive constraint, on average, up to 500bp upstream ($p = 0.0034$) and downstream ($p = 0.0473$) of the coding sequence when four-fold synonymous sites are used as the neutral standard. This is also true when intronic sequences are used as the neutral standard, and the result is more significant ($p < 0.0001$ for both upstream and downstream sequences). Constraint even appears to be strong when up to 1kb of upstream sequence is included ($C = 0.377$ using introns as the neutral standard). It should be noted, however, that in the case of downstream sequences, the p-value is somewhat unreliable when using intronic sites as the neutral standard, since it is based on the bootstrapping of only a small number of constraint estimates.

Constraint in intergenic DNA appears to be strongest close to the exon boundary, for both 5' and 3' sequences, suggesting that functionally important regions of intergenic DNA tend to lie close to exons. It is possible that this pattern of constraint could be caused by selection acting on the DNA sequence of UTRs (which may be involved in post-transcriptional processes), or due to the presence of regulatory sequences.

Table 3.7: Observed (obs) and expected (exp) numbers of substitutions along with estimates for constraint (*C*) in intergenic DNA sequences using intronic sequences (with splice control regions removed) as a neutral standard.

Type	Base Pairs	Loci	bp / locus	Obs (SE)	Exp (SE)	*C* (SE)
5'	1–100	40	98.8	5.55 (0.488)	10.3 (0.815)	0.458 (0.0643)
	101–200	37	96.6	4.95 (0.540)	9.99 (0.919)	0.500 (0.0739)
	201–300	34	100	6.86 (0.802)	10.3 (0.971)	0.325 (0.102)
	301–400	34	100	7.01 (0.998)	9.72 (0.854)	0.271 (0.135)
	401–500	34	93.4	6.44 (0.723)	9.33 (0.911)	0.305 (0.0987)
	501–700	28	132	8.62 (1.38)	13.7 (1.50)	0.367 (0.0920)
3'	1–100	26	96.1	5.08 (0.842)	9.23 (1.01)	0.443 (0.110)
	101–200	20	83.6	3.56 (0.754)	7.51 (0.908)	0.521 (0.113)
	201–300	13	91.5	4.86 (0.830)	7.35 (1.14)	0.327 (0.141)
	301–500	11	150	4.93 (1.20)	13.4 (2.63)	0.621 (0.110)

Standard errors (in brackets) were estimated by bootstrapping by gene.

3.4 Discussion

In this chapter the level of functional constraint for non-coding intronic and intergenic DNA sequences in *D. melanogaster* and *D. simulans* was calculated. Constraint was calculated by comparing expected numbers of substitutions, obtained from putatively neutral sequences, to the observed number of substitutions, allowing for differences in base composition. Constraint was defined as the fraction of missing substitutions (Kimura 1983). Initially, four-fold synonymous sites in nearby exons were used as the neutral standard, but subsequently intronic sites with splice sites removed were used, since they appeared to be evolving faster, on average.

Genes to be studied were selected non-randomly by searching Genbank for *D. simulans* genes that already had reasonable amounts of coding sequence available. It is possible that the sample of genes selected may be biased towards more highly expressed genes. In order to test this, the distributions of codon-usage bias (which is correlated with expression level in *Drosophila*; Duret & Mouchiroud 1999) were plotted for the non-random sample of *D. simulans* genes and for a random sample of 400 *D. melanogaster* genes (see Figure 3.2). There does not appear to be a consistent bias in codon-usage bias between the two datasets. It is therefore unlikely that the results are substantially affected by the sampling method.

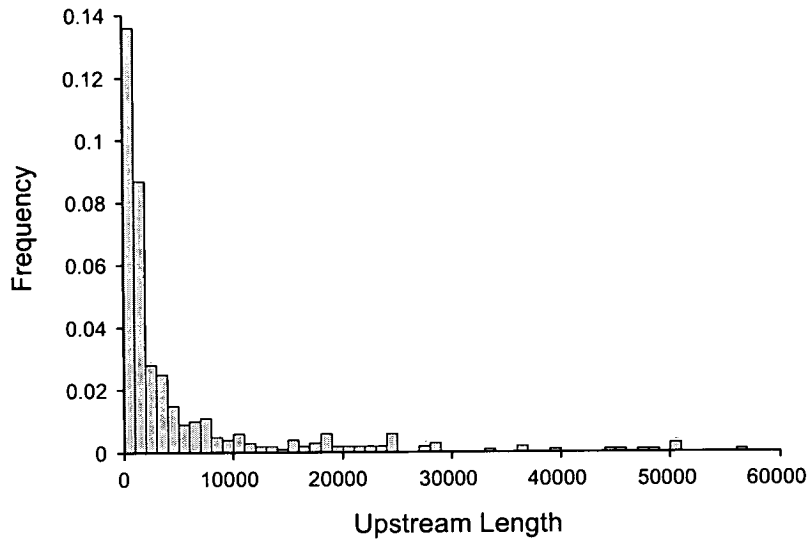


Figure 3.3: Distribution of intergenic lengths for a random sample of 400 *Drosophila melanogaster* genes

Exons and genes appear to be highly clustered in *Drosophila*, meaning that the distributions of intron and intergenic lengths are highly L-shaped and leptokurtic (the distributions have a high peak and long tails) (see Figures 3.3 and 3.4). Many intronic and intergenic sequences are therefore very short; in the random sample of 400 *D. melanogaster* genes, 34% of upstream lengths are shorter than 1000bp and 56% of introns are shorter than 100bp. If introns or intergenic sequences have common regulatory functions, it should therefore be possible to detect them in the sample of intronic and intergenic sequences studied.

There are also *a priori* reasons for believing that intergenic and intronic sequences should show evidence of functional constraint, for example, they are known to contain regulatory regions which are thought to be under selection (Stephan & Kirby 1993, Kirby *et al.* 1995, Leicht *et al.* 1995). There have been several studies that indirectly indicate that constraint in non-coding DNA might be quite substantial in various species (Britten 1986, Li & Graur 1991, Oeltjen *et al.* 1997, Jareborg *et al.* 1999, Shabalina & Kondrashov 1999, Bergman & Kreitman 2001, Waterston *et al.* 2002, Thomas *et al.* 2003, Dermitzakis *et al.* 2003). The present study agrees with many of these recent studies in that it suggests that there is substantial constraint in intergenic sequences (C = 0.158 (SE 0.0545) for 500bp upstream and 0.200 (SE 0.107) for 500bp downstream using four-fold sites as a neutral standard). There is

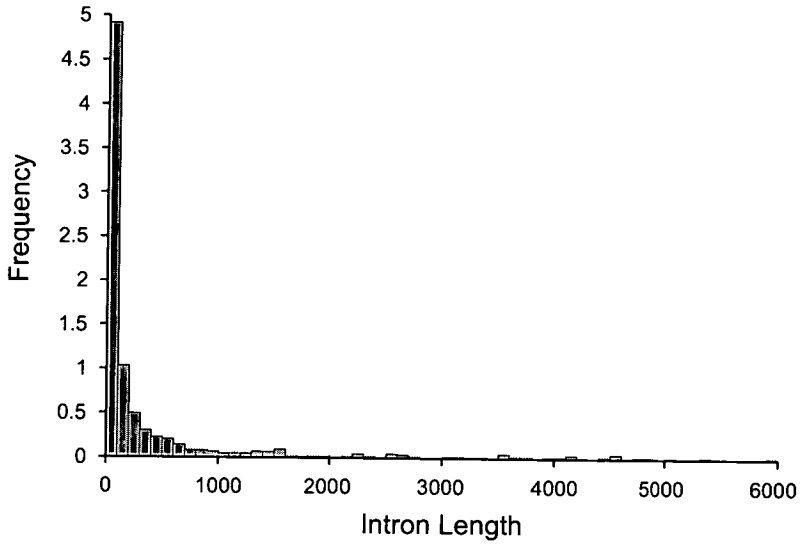


Figure 3.4: Distribution of intron lengths for a random sample of 400 *Drosophila melanogaster* genes

evidence for constraint upstream and downstream of the coding sequence, and for constraint being stronger near to coding sequences. However, the power to detect constraint is currently limited by the lack of available *D. simulans* sequence data, especially for 3' sequences. Completion of the *D. simulans* and *D. yakuba* genome sequences will greatly help to clarify whether or not the patterns observed in this study and others are genuine. If this extra sequence data were available and the annotations of genes were reliable, it would also be interesting to determine what fraction of sequence constraint in intergenic regions is due to constraint in UTR sequences.

The robustness of constraint estimates in intergenic DNA was examined by varying the assumed equilibrium G+C content (f_e). For upstream sequences over 1000bp long, using four-fold sites as a neutral standard, the estimate of constraint is significantly higher when f_e is set above 0.56, and significantly lower when f_e is set below 0.14. There is therefore quite a wide range of f_e values for which estimates of constraint do not differ significantly. When introns are used as the putative neutral standard there is no significant change in the estimated value of constraint over the entire range of f_e values, further suggesting that the conclusion that there is substantial constraint in intergenic non-coding DNA is robust.

Introns are known to contain conserved splice sites (Sharp 1994), and in accordance with this, constraint in splice control regions within introns was shown

to be strong (see Tables 3.4 and 3.5), the first two base pairs at the 3' and 5' end of introns were absolutely constrained in all cases. However, estimates of constraint in introns, outside of splice control regions, suggest that introns are subject to little or no purifying selection. In fact the majority of intronic sites appear to be evolving faster than four-fold synonymous sites (as measured by constraint), which themselves are thought to be only under weak selection. This result disagrees somewhat with previous analyses, for example the results presented by both Bergman & Kreitman (2001) and Shabalina & Kondrashov (1999) suggested that the level of constraint within introns was similar to that in intergenic sequences (which in this study were found to be under significant levels of functional constraint). These results presented here also appear to be fairly robust, in terms of the assumed value of f_e. Values of f_e below 0.53 result in higher rates of evolution at intronic sites than four-fold sites, and for values below 0.51 and above 0.21, the difference in rates (as measured by constraint) is non-significant at the $p = 0.05$ level.

Estimates of the level of functional constraint in some previous studies (Jareborg *et al.* 1999, Shabalina & Kondrashov 1999, Bergman & Kreitman 2001) were calculated using the fraction of conserved nucleotides in alignable blocks of DNA between distantly related species. This method, however, has two potential problems. Firstly, variation in mutation rate could give a false level of constraint in neutrally evolving DNA (Clark 2001) and, secondly, alignment of non-coding DNA between divergent species is difficult and could be biased if not based on a specific model of indel evolution (Thorne *et al.* 1991). Interestingly, the results from this study agree with those from another recent study using similar methodology to estimate constraint in the non-coding DNA of rodents (Keightley & Gaffney 2003). Keightley & Gaffney (2003) found that intronic sites evolve faster than four-fold synonymous sites, in contrast to upstream and downstream sequences which appeared to be highly conserved.

The data in Table 3.4 suggest that the number of constrained nucleotides per intron is about 4.1. If there are 41,000 introns in the *Drosophila* genome (Adams *et al.* 2000), the predicted number of constrained intronic nucleotides is only 0.17Mb. The level of constraint at amino acid sites in *Drosophila* genes has been estimated at about 84% (Eyre-Walker *et al.* 2002), which implies that the total number of constrained amino acid sites in the *Drosophila* genome is about 16Mb (~14,000 protein-coding genes, on an average 591 codons long (Adams *et al.* 2000), with mutations in about

three-quarters of sites leading to amino acid changes). The number of constrained nucleotides in introns is therefore relatively small compared to the protein coding segment of the genome. However, the number of constrained nucleotides in intergenic DNA could potentially be of the same order as that in coding DNA. For example, using the average constraint values estimated for upstream sequences relative to intronic sequences to calculate the number of constrained nucleotides in intergenic DNA, gives 14,000 genes × 1,000bp × 0.377 = 5.3Mb. This is a conservative estimate, because it ignores constraint downstream of genes. The value could therefore be much larger if there is substantial constraint downstream of genes and if there are functional constraints deep within intergenic DNA. If there are many non-coding sites under relatively weak selection, the genetic load caused by the drift and fixation of small-effect mutations could be substantial.

4 Estimating Numbers of EMS-induced Mutations Affecting Life-History Traits in *Caenorhabditis elegans* Using Inbred Sublines

The work described in this Chapter has been recently published (Halligan *et al.* 2003).

4.1 Introduction

The majority of newly arising spontaneous mutations are believed to be deleterious, and several important evolutionary phenomena have been hypothesised to be consequences of recurrent deleterious mutation. These include inbreeding depression (Charlesworth & Charlesworth 1987), the evolution of sex and recombination (Charlesworth 1990, Kondrashov 1988), evolution of mating systems (Charlesworth *et al.* 1990), ecological specialisation (Kawecki *et al.* 1997), genetic variability for quantitative traits (Bulmer 1989), senescence (Charlesworth 1994) and extinction of small populations (Lande 1994, Lynch *et al.* 1995*a*). It has been suggested that mutation accumulation may even threaten the persistence of our own species (Crow 1997, Kondrashov 1995, Muller 1950). Whether or not mutations play a role in these phenomena critically depends on parameters associated with mutations (Caballero & Keightley 1994, Turelli 1984), including the genomic mutation rate (U) and the distribution of selection (s) and dominance (h) coefficients of new mutations.

With theory increasingly showing the potential importance of the properties of mutations, there has been a resurgence of interest in attempting to estimate U, and mean s. Although inference of the distribution of mutation effects has received less attention (Lynch *et al.* 1999), the distribution of effects is important for several reasons. Firstly, there is good reason to expect that mutation effects vary substantially, since genomes contain sites that vary greatly in functional significance. Secondly, evaluation of some evolutionary theories, such as the time to mutational

meltdown, requires knowledge of the distribution of effects (Butcher 1995, Lande 1994, 1995, Lynch *et al.* 1995*b*). Thirdly, it is possible that estimates of U and mean s obtained from mutation accumulation experiments could be substantially biased if the distribution of mutation effects is not co-estimated. Finally, small effect mutations are as important as large effect mutations because the mutation load exerted by a mutation is independent of the strength of selection (under multiplicative selection without drift) and small effect mutations are more likely than strongly selected mutations to produce regional patterning of variability along a chromosome in response to local variation in recombination rate (Nordborg *et al.* 1996).

There is some evidence from natural populations that suggests that the majority of deleterious mutations in a population are of small effect. For example, two studies by J. H. Willis (Willis 1999*a*,*b*) estimated the fraction of segregating mutations in a natural population of *Mimulus gluttatus* that were of large effect by measuring the decrease in inbreeding depression after a period of inbreeding. Since small effect genes are effectively neutral in a highly inbreeding population they will not be effectively purged from a population. On the other hand, large effect mutations will be purged and will lead to a decrease in the level of inbreeding depression. Through this methodology Willis (1999*b*) inferred that lethals and steriles make a minor contribution to the inbreeding depression (steriles only account for about 26% of lifetime reproductive success, Willis 1999*a*). Further, evidence for wide variation in effects of mutations comes from an analysis of the effects of EMS mutagenesis in *C. elegans* (Davies *et al.* 1999, see also Keightley *et al.* 2000). The distribution of effects of EMS-induced mutations was evaluated by comparing an *a priori* estimate of the number of induced mutations at the molecular level to an estimate of the number of mutations detectable from fitness assays. The molecular estimate was obtained from the expected rates of EMS-induced point mutations based on experiments to measure forward mutation rates (Bejsovec & Anderson 1988) and suppressor-induced reversion mutation rates (Hodgkin 1985, Kondo *et al.* 1990, Waterston 1981). This was converted to a conservative estimate of the number of deleterious mutations induced in the genome by incorporating information on the size of the *C. elegans* genome, the percentage of this genome that is protein coding, and the level of evolutionary constraint within protein coding sequences. Davies *et al.* estimated that they had induced approximately 45 deleterious point mutations per homozygous mutant line. However, when a maximum likelihood approach was used to estimate

the number of deleterious mutations per line from the results of fitness assays it was found that only 3.60 (±1.31) were detectable (for the trait relative fitness, assuming a gamma distribution of effects, Keightley *et al.* 2000). It is likely, therefore, that there is a large class of mutations with undetectably small, but still deleterious, effects.

In the present experiment, inbred sublines have been created from a random selection of the EMS-induced mutant lines produced by Davies *et al.* (see also Section 2.2.1) in an attempt to refine estimates of the number of mutations per line and to establish whether or not the fitness differences between the wild-type line and the mutant lines are due to a small number of large-effect mutations, or a large number of smaller-effect mutations. Sublines were produced by crossing the selected mutant lines to a wild-type control, and inbreeding the offspring to homozygosity under conditions of minimal selection. For unlinked mutations, under this design, it is expected that a random selection of half of the mutations present in each mutant line will be present in any subline. By measuring the fitness of each mutant line, the wild-type control and the individual sublines, it should be possible to estimate the number of mutations present in each mutant line. This approach may more accurately estimate the number of mutations per line than an approach that just measures the fitness of the wild-type and homozygous mutant lines because a large-effect mutation present in one line should segregate amongst its sublines and be easily identifiable. The pattern of segregation of mutations among sublines should also give information about the distribution of mutation effects without having to rely on information from higher order moments. A modification of the Castle-Wright estimator (Castle 1921, Wright 1968) and a maximum likelihood (ML) method have been used to estimate the average number of mutations per line. The ML approach can deal with data for which the distribution of residual data points is expected to be significantly different from the expectations of a normal distribution. The method also allows for two classes of mutation effect although it was not possible to fit a continuous distribution of mutation effects due to the computing time required. The results presented here are consistent with the conclusions of Davies *et al.*, although there was not enough power to verify the existence of a large class of very small effect mutations.

4.2 Materials and Methods

4.2.1 Generation of Sublines and Life-History Trait Assays

Of the 56 EMS-induced mutant lines produced by Davies *et al.* (1999) ten were randomly chosen for use in this experiment (E1, E2, E3, E4, E5, E6, E7, E8, E9 and E11—collectively termed p-lines, for "progenitor" lines). Along with one control line (N2) these were thawed from storage at −80°C (see Section 2.2.4). N2 males were generated in order to carry out crosses to the p-lines (see Section 2.2.5). Male worms of the N2 strain were then randomly selected and crossed to hermaphrodites of the ten p-lines to produce offspring that were heterozygous for the mutations in each p-line. It was checked that the ratio of male to hermaphrodite offspring did not significantly differ from the expected 1:1 using a χ^2 test with one degree of freedom. Two of ten p-lines (E1 and E7) produced too few offspring, or insufficient males, and so could not be included in the experiment.

For each of the eight remaining p-lines, ten F1 hermaphrodite offspring were chosen at random and moved to new plates. Each resulting subline was then inbred for a minimum of ten generations by transferring one larval hermaphrodite, chosen at random, to a new plate every generation. After ten generations 99.9% of loci that were heterozygous in the F1 should be homozygous for either the wild-type or the mutant allele. This minimises selection by bottlenecking the population to one individual each generation, and generates offspring that are homozygous for ~1/2 of the mutations in the original mutant line, with wild-type (N2) alleles at the rest of their loci (see Figure 4.1). One backup plate was set up each generation in case the primary plate failed. If both of these plates failed, offspring from the previous generation's plates (kept at 16°C, in order to slow their growth) were used. This procedure yielded ten sublines per p-line, labelled E2.1–E2.10; E3.1–E3.10 *etc.* Only one subline (E4.10) was lost during the inbreeding process due to the primary, backup and previous generation's plates failure to produce a viable worm, suggesting that the worms were subject to very little natural selection.

Daily productivity and longevity were measured contemporaneously for the control line (N2), the eight p-lines and their respective sublines, over three assays. In each assay, each of three people (counters) assayed one worm for each p-line and subline and eight worms for the control (N2) line per assay, giving a total of nine

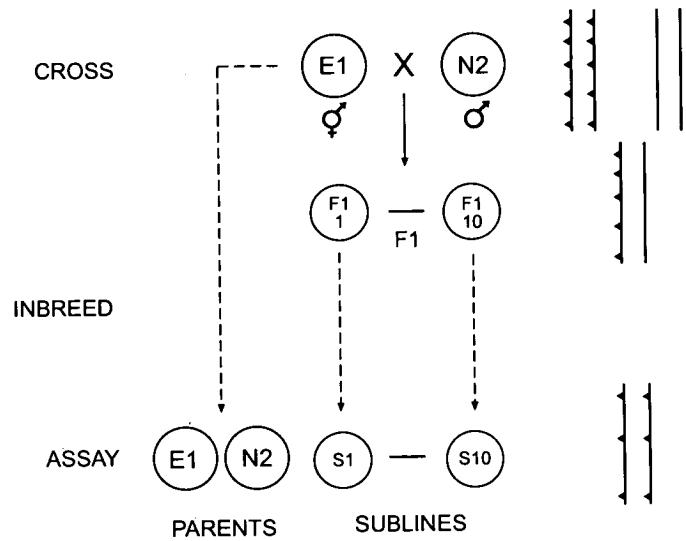


Figure 4.1: Illustration of the procedure to produce sublines from a single homozygous mutant line. The mutant line (E1) is initially crossed to the mutation free wild-type line (N2), then a random selection of the F1 offspring are inbred for >10 generations. After inbreeding each subline is expected to be homozygous for half the mutations present in the progenitor mutant line. The vertical lines on the right show the distribution of mutations present on sample homologous chromosomes for the progenitor lines (E1 and N2), an F1 individual and a single inbred subline (triangles represent individual mutations). N2 worms are assumed to have no mutations, and the mutant progenitor (E1) is assumed to be homozygous for all induced mutations, whereas F1 individuals are heterozygous. After inbreeding the sublines are homozygous for a random sample of the mutations present in the F1.

replicates for each p-line and subline, and 72 replicates for the control line. Within each assay, each counter's plates were randomised with respect to their position in the incubator and the order in which they were counted. Prior to each assay, replicates were maintained separately for three generations in an attempt to remove any possible maternal effects. If any replicates failed in one assay as a result of unnatural death due to human error or worms crawling off the plate, extra replicates were added to the same counter's quota in the following assay.

Daily productivity was recorded by counting the number of offspring surviving to the L3 larval stage daily for the first five days of productivity. Longevity was scored by recording the day on which the parental worm failed to respond to a light touch from a platinum pick and showed any loss of turgor or visible sign of decay. Four fitness correlates were obtained from the productivity data: early productivity (days one to two), late productivity (days three to five), total productivity (days one to five) and relative fitness (w). Relative fitness is a measure related to intrinsic population growth rate that gives more weight to early days of productivity (according to an exponential distribution) and is suitable for an age-structured population. To calculate w, the intrinsic growth rate (r) of the controls was computed by solving

$$\sum_x e^{-r_c x} l_c(x) m_c(x) = 1 \tag{4.1}$$

where $l_c(x)$ and $m_c(x)$ are the least-square means of the proportion of worms surviving to day x and fecundity at day x respectively, for the controls within an assay. Relative fitness was then calculated separately for each individual from

$$w_{ijk} = \sum_x e^{-r_{ci} x} l_{ijk}(x) m_{ijk}(x) \tag{4.2}$$

where r_{ci} is the average intrinsic growth rate for the control lines within an assay i, and $l_{ijk}(x)$ and $m_{ijk}(x)$ are the proportions of worms surviving to day x and fecundities at day x respectively, for assay i, worm j of line k (Charlesworth 1994, p. 120).

4.2.2 Castle-Wright Estimator of Number of Mutations

The Castle-Wright estimator can be used to calculate the effective number of factors (n_e) contributing to the difference in a trait between two divergently selected inbred

lines using information on the phenotypic means and variances of the two progenitor lines and their line-cross derivatives (Castle 1921, Wright 1968) with modifications by (Cockerham 1986, Lande 1981). This quantity is equivalent to the actual number of genes contributing to the fitness difference between the two lines if it is assumed that all mutations are additive, unidirectional in effect, unlinked and have equal effects. The method can be modified to estimate the number of effective factors contributing to the fitness difference between N2 and a given p-line and with this modification, the Castle-Wright estimator is as follows:

$$\hat{n}_e = \frac{(\hat{\mu}_{N2} - \hat{\mu}_i)^2 - \hat{\sigma}^2_{\hat{\mu}_{N2}} - \hat{\sigma}^2_{\hat{\mu}_i}}{4\hat{\sigma}^2_{s_i}} \tag{4.3}$$

where $\hat{\mu}_{N2}$ and $\hat{\sigma}^2_{\hat{\mu}_i}$ are the observed mean and sampling variance of the trait value for N2, and $\hat{\mu}_i$ and $\hat{\sigma}^2_{\hat{\mu}_i}$ are the observed mean and sampling variance of mutant p-line i. $\hat{\sigma}^2_{s_i}$ is the segregational variance among the inbred sublines for p-line i (Lynch & Walsh 1998). The above means, sampling variances of the means and the segregational variances amongst each p-line's sublines were estimated using the MIXED procedure of SAS 6.12 (SAS Institute Inc. 1997, Littell *et al.* 1996) for each trait. Factors included in the model were assay (1–3), counter (1–3), line (1–8), line-type (N2, p-line or subline) and subline (1–10, nested within line × line-type). Counter, assay and subline(line × line-type) were treated as random effects; all other effects were treated as fixed.

The standard error of $\hat{n}_e$ for the Castle-Wright estimator can be approximated using the delta method (Lande 1981). Modifying this formula to use a variance estimate from sublines instead of an F1, the following is obtained:

$$\mathrm{Var}(\hat{n}_e) \simeq 4\hat{n}_e^2 \left(\frac{4(\hat{\sigma}^2_{\hat{\mu}_{N2}} + \hat{\sigma}^2_{\hat{\mu}_i})}{(\hat{\mu}_{N2} - \hat{\mu}_i)^2} + \frac{\mathrm{Var}(\hat{\sigma}^2_{s_i})}{\hat{\sigma}^4_{s_i}} \right) \tag{4.4}$$

This estimate ignores the correction factor proposal for the numerator of the Castle-Wright estimator, ($\hat{\sigma}^2_{\hat{\mu}_{N2}} + \hat{\sigma}^2_{\hat{\mu}_i}$), as it has been suggested that this would unduly complicate the variance (Cockerham 1986).

4.2.3 Likelihood Approach for Estimating Mutational Parameters

Using a maximum likelihood (ML) method to estimate the number of loci contributing to the fitness difference between the N2 line and a given p-line has the advantage that it uses information about the distribution of fitness values amongst sublines. Similar ML approaches have been used to estimate mutational parameters in previous experiments (Keightley 1994, Keightley & Bataillon 2000, Keightley *et al.* 2000, Vassilieva *et al.* 2000; the method used here is based on Keightley & Bataillon 2000). In general, these approaches assume that mutations have additive effects on fitness, which follow a given distribution, and that once these effects are removed, the residual data points are normally distributed with the same environmental variance and mean. As an extension to this method the assumption of normally distributed residuals has been relaxed by assuming instead that the residuals are distributed normally when transformed by an unknown (but estimated) power (κ), following Box & Cox (1964).

Following Box & Cox (1964), we assume that for some unknown κ, observations (y) transformed by the function

$$y^{(\kappa)} = \begin{cases} \frac{y^{\kappa}-1}{\kappa} & \text{if } (\kappa \neq 0); \\ \log y & \text{if } (\kappa = 0). \end{cases} \tag{4.5}$$

satisfy the full normal theory assumptions, assuming $y > 0$. This function is continuous at $\kappa = 0$ and is therefore preferable to simply using y^{κ} as the transformation (Box & Cox 1964).

Replicates for the N2 line were assumed to have a mean μ, a variance V_E, and to follow a normal distribution when transformed by an unknown power (κ). The p-line and subline replicates were also assumed to have the same underlying environmental variance (V_E), and the number of mutations in each of the p-lines was assumed to be a Poisson random variable with mean λ. Each mutation was assumed to be unlinked from others, have a negative effect on the trait, and fall into one of two discrete classes of effect size (s_1 and s_2), where the proportion of class 1 mutations (R) is also a parameter of the model. As a special case, it can be assumed that the proportion of mutations in class 1 is 1; this is termed the one-class model. The model allows any number of fixed effects with any number of levels; for the experimental data both counter and assay were modelled as fixed effects, each having three possible levels.

The levels within a fixed effect were all assumed to have the same variance but different means (scaled relative to the largest level in each fixed effect). For more than one fixed effect, the total of the relevant difference between levels for each fixed effect is calculated separately for each replicate (k) of each p-line (i) and this total is labelled a_{ik} for the following equations. Since all levels are scaled relative to the largest for each fixed effect, a_{ik} can only be negative, meaning all residuals will be positive when a_{ik} is removed, satisfying the requirement that $y > 0$ for the Box-Cox transformation.

Let X_{ik} equal the phenotypic value of p-line i replicate k, then according to the assumptions above

$$X_{ik} = \mu + x_{1i}s_1 + x_{2i}s_2 + a_{ik} + e_{ik}^{(\kappa)} \tag{4.6}$$

where x_{1i} is the number of mutations in class 1 for p-line i and x_{2i} is the number of mutations in class 2 for p-line i. s_1 and s_2 are the effects of class 1 and class 2 mutations respectively. $(x_1 + x_2)$ is a Poisson deviate with mean λ and x_1 is a binomial deviate from a total of $x_1 + x_2$ possible mutations with a probability of success of R. $e_{ik}^{(\kappa)}$ is a transformed Gaussian deviate, with mean 0, variance V_E.

Similarly, if Y_{iln} is the phenotypic value of subline l, replicate n from the p-line i, then

$$Y_{iln} = \mu + y_{1il}s_1 + y_{2il}s_2 + a_{iln} + e_{iln}^{(\kappa)} \tag{4.7}$$

where y_1 and x_1 are binomial deviates with a total of x_1 and x_2 possible events respectively and probabilities of success of 0.5.

In the calculation of the likelihood for each line, the likelihood of obtaining the data for that line at every point in the probability space needs to be summed across all the possible points. In the model presented here, there can be anywhere from 0 to an infinite number of mutations present in each p-line. Of these (j) mutations, any number m ($0 \leq m \leq j$) could be in class 1; the remainder ($j - m$) belong to class 2. Some number p ($0 \leq p \leq m$) of class 1 mutations and some number q ($0 \leq q \leq (j - m)$) of class 2 mutations are present in each of the 10 sublines of a given p-line. For each possible combination of subline class 1 and class 2 mutations, the likelihood of obtaining the subline data for the ten sublines belonging to each p-line needs to be calculated.

The likelihood associated with the data from a single line (i) will therefore be:

$$L_{(\text{Line}_i)} = \sum_{j=0}^{\infty}\left(p(j|\lambda) \times \sum_{m=0}^{j}\left(\text{bi}(m|j) \times \prod_{k=1}^{\text{p-reps}} f(X_{ik} - ms_1 - (j-m)s_2 - a_{ik})\right.\right.$$
$$\left.\left. \times \prod_{l=1}^{\text{sublines}} \sum_{p=0}^{m}\left(\text{bi}(p|m) \times \sum_{q=0}^{(j-m)}\left(\text{bi}(q|(j-m)) \times \prod_{n=1}^{\text{s-reps}} f(Y_{iln} - ps_1 - qs_2 - a_{iln})\right)\right)\right)\right) \tag{4.8}$$

where $p(j|\lambda)$ denotes the (Poisson) probability that the p-line i contains j mutations given the mean λ and $\text{bi}(m|j)$ denotes the (binomial) probability that p-line i contains m class 1 mutations given that line i contains a total of j mutations, and the probability of each mutation being class 1 is R. $\text{bi}(p|m)$ is the (binomial) probability that subline i, l has p class 1 mutations (given that p-line i has m), and $\text{bi}(q|(j-m))$ is the (binomial) probability that subline i, l has q class 2 mutations (given that p-line i has $j-m$)

f is a transformed Gaussian probability density function, shown below (adapted from Box & Cox 1964).

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2_{(\kappa)}}}.\exp\left(-\frac{1}{2}\left(\frac{(y^{(\kappa)} - \mu_{(\kappa)})^2}{\sigma^2_{(\kappa)}}\right)\right).y^{(\kappa-1)} \tag{4.9}$$

where y and $y^{(\kappa)}$ are the untransformed and transformed observations as described above. There are three parameters, $\mu_{(\kappa)}$ and $\sigma^2_{(\kappa)}$ are the mean and variance of the transformed variable respectively, and κ is the power of the transformation.

The overall log-likelihood is then obtained by adding the sum of the log-likelihoods across all p-lines to the log-likelihood for N2 data. The log-likelihood for the N2 data was summed over all N2 replicates, where the likelihood for each N2 replicate is $L(Z_i) = f(Z_i - a_i)$, where Z_i is the observation for N2 replicate i and a_i denotes the total effect of any fixed effects modelled. In order to calculate approximate standard errors, a quadratic function was fitted to a profile likelihood graph of the parameter of interest. The standard error for the parameter of interest is then approximately equal to the square root of the inverse of the negative second derivative of the fitted quadratic function (Weir 1996, pp. 60–61).

To verify the calculations and functionality of the maximum likelihood program, simulations were carried out. The same mutational model was used in the simulations

and the maximum likelihood program. It was checked that the known values for the parameters used to simulate data were estimated correctly using the maximum likelihood approach. Due to the computer intensive nature of the approach only a limited number of simulations could be carried out, with limited numbers of data points per simulation. To further reduce the time taken to run the simulations, a low number of mutations and a low number of sublines relative to p-lines were modelled.

4.2.4 Likelihood Maximisation

Finding the true maximum likelihood can be difficult when the likelihood space has many dimensions and interdependent variables, since there can be multiple peaks. Therefore, it is necessary to thoroughly search the likelihood space in order to be sure that any maximum found is the true global maximum. In the one-class model there are five parameters (μ, V_E, λ, s and κ) plus any parameters associated with fixed effects; in the two-class model there are two additional parameters (s_2 and R). Starting values for μ, V_E and any fixed effects were estimated from the N2 data. In order to obtain starting values for the remaining parameters a grid search was carried out, without maximisation, where the likelihood was evaluated for a combination of set values for each parameter over a broad range.

A linear search strategy was then employed, using the most likely values obtained during the grid search, in which a series of fixed values for λ were selected about its starting value, since this is the parameter of interest. The likelihood was maximised with respect to all other parameters, using the simplex algorithm (Nelder & Mead 1965). The simplex was then restarted using the values for μ, V_E, λ, s and κ that gave the highest likelihood during the linear search, and the likelihood was maximised with respect to all parameters. The simplex algorithm was restarted after each maximisation until there was no further increase in the likelihood.

4.2.5 E5.2 and E5 Extra Line Crosses

From the primary experiment it was clear that line E5.2 had a significantly lower relative fitness than either of its progenitor lines (E5 and N2). Under the assumptions that all mutations are deleterious, freely recombining and show no epistasis, this result is unexpected. Possible explanations are: 1. Line E5 carries mixtures of mutations

with both positive and negative effects on relative fitness, in which case it would be possible for sublines to have fitnesses outside the range of their progenitor lines. 2. Mutations in line E5 interact epistatically, such that they only cause the dramatic reduction in fitness visible in line E5.2 when segregated in a line cross. 3. A new spontaneous mutation occurred during the generations of selfing that produced subline E5.2.

In order to test the hypothesis that the reduction in relative fitness in subline E5.2 was due to a new spontaneous mutation, lines E5 and E5.2 were subjected to further line crosses. If a new spontaneous large-effect mutation occurred during the generations of selfing, then this mutation should segregate in sublines generated from a cross between E5.2 and N2 but there would be no evidence of it in sublines produced from a cross between E5 and N2. Alternatively, if mutations present in line E5 cancelled out each others effects on w, through epistasis or by having both positive and negative effects on w, then further sublines (generated from lines E5 and N2) would be expected to perform outside the range of the two progenitor lines.

Twenty sublines were generated from both lines E5 and E5.2 using the same experimental design as for the main experiment except that two new (independently frozen) replicates of the ancestral wild-type line (labelled N2A and N2B) were thawed. E5 and E5.2 were each crossed to the males of N2A and N2B and ten offspring from each cross were selected randomly and selfed under minimal selection conditions for seven generations. This produced 44 different lines that were then assayed for total productivity: N2A, N2B, E5.2, E5, and 20 sublines for each of E5.2 and E5. Six replicates were set up for each of these lines, giving a total of 264 data points. All replicates were kept separated for three generations prior to the assay to reduce the possibility of any maternal effects.

The results of the productivity assay were used to obtain four fitness correlates (early, late and total productivity and relative fitness), which were analysed as before using the MIXED procedure of SAS 6.12 (SAS Institute Inc. 1997, Littell *et al.* 1996). Lines E5 and E5.2 were analysed separately, and the factors included in each model were line (N2 or E5 / E5.2), line-type (wild type, p-line or subline), subline (1–20, nested within line × line-type) and N2type (A or B). N2type and subline(line × line-type) were treated as random effects; all other effects were treated as fixed. The number of mutations segregating in line E5.2 and E5 was also estimated by applying

Table 4.1: LS-means from proc MIXED (SAS Institute Inc. 1997) for lines of type N2, p-line or subline.

Trait	N2 mean (SE)	p-line mean (SE)	subline mean (SE)
w	1.00 (0.0293)	0.611 (0.0739)	0.792 (0.0244)
Early (worms)	211 (15.9)	140 (21.0)	171 (15.7)
Total (worms)	258 (8.30)	208 (16.7)	231 (7.58)
Late (worms)	46.7 (10.3)	70.7 (12.1)	60.1 (9.61)
Longevity (days)	11.9 (0.367)	11.6 (0.523)	11.9 (0.367)

Standard errors are shown in brackets.

the Castle-Wright estimator and the maximum likelihood approach discussed above.

4.3 Results

4.3.1 Segregating Mutant Phenotypes and Castle-Wright Estimates

A total of 830 data points were obtained from the experiment for five fitness correlates, and a total of 193,157 offspring were counted to obtain the productivity data. EMS mutagenesis has the strongest effects on early productivity; this is reflected in a large effect on relative fitness (Table 4.1, Fig. 4.2). Mutational effects on late productivity and longevity, however, are relatively weak on average (Table 4.1, Fig. 4.2). This pattern was also noted by Keightley *et al.* (2000), who hypothesised that it was due to mutations lengthening mean development time, resulting in a decrease in early reproductive output. Deleterious mutations may therefore either increase or decrease late productivity, by delaying development or by reducing total productivity. Longevity in particular appears to be a small "mutational target", with large amounts of environmental variation. This has also been noted in previous literature, several experiments finding little evidence for strong directional effects of mutations on longevity (Keightley & Caballero 1997, Keightley *et al.* 2000, Pletcher *et al.* 1999, Vassilieva & Lynch 1999). Neither longevity nor late productivity fit a model with only negative-acting mutations, so these traits were excluded from any of the maximum likelihood analyses.

The EMS-induced mutant lines tested all had lower point estimates for *w* than

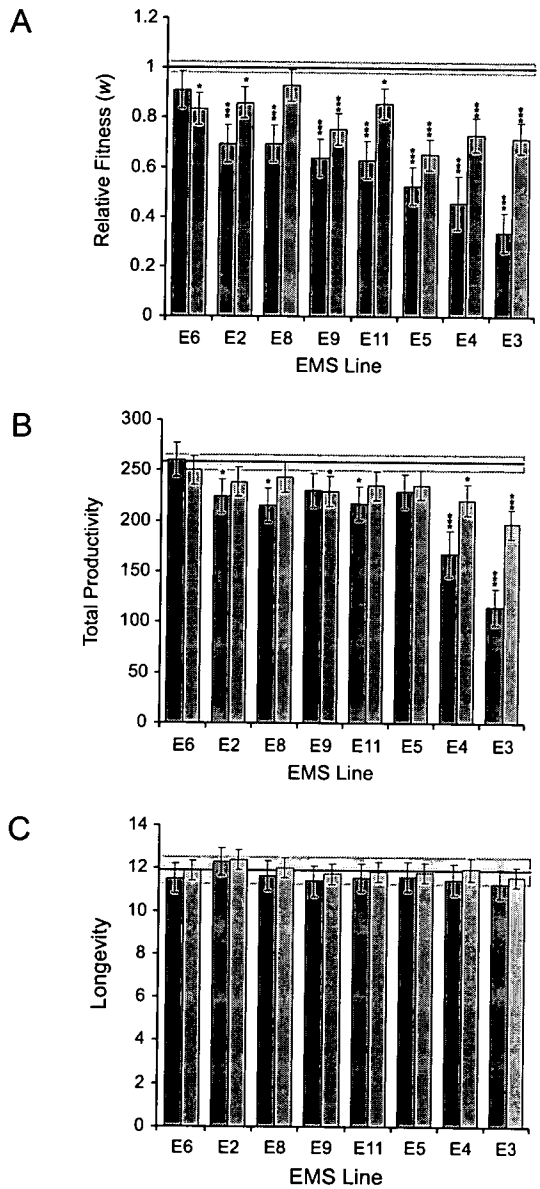


Figure 4.2: Means for relative fitness (w) (A), total productivity (B) and longevity (C) for the wild-type (N2), p-lines and sublines by line. The mean for N2 is shown as a horizontal bar ± standard error (grey box), the means of the p-lines and sublines are darkly shaded and lightly shaded bars respectively (± standard error). Asterisks above the means of the p-lines and sublines correspond to the significance of the difference between the given genotype and the wild type. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.0001$.

N2 (seven out of eight were significantly lower; p-values < 0.0001; Fig. 4.2A), seven out of eight had lower point estimates for total productivity (five were significant; p-values < 0.05; Fig. 4.2B), and for longevity none were significant (all p-values > 0.5; Fig. 4.2C). For w the mean values of the ten sublines fell between those of their respective p-lines and the N2 for all but one of the lines studied (Fig. 4.2A). This was also true for all but two lines for total productivity (Fig. 4.2B) and all but three lines for longevity (Fig. 4.2C). Most individual sublines also had point estimates for w between their respective p-line and N2 (Fig. 4.3), with two major exceptions: (i) line E5.2 had a significantly lower early productivity, total productivity and w ($p < 0.0001$) than both line E5 and N2 from which it was derived (Fig. 4.3D) (ii) three sublines generated from line E6 (E6.2, E6.3 and E6.5, $p < 0.05$ for all) had significantly lower relative fitness than either E6 or N2. It is shown later that the dramatically lower productivity observed in line E5.2 is likely to be the result of a single large-effect spontaneous mutation that occurred during the 10 generations of inbreeding needed to produce sublines. The data for this subline were therefore excluded from the following analyses. On the other hand it is unlikely that the lower fitness observed in sublines E6.2, E6.3 and E6.5 is due to new spontaneous mutations during the period of inbreeding (this would require three independent mutations). It is perhaps more likely that this pattern is due to mutations present in the progenitor line E6 which have both positive and negative effects on relative fitness. If these mutations were segregated amongst the sublines then it would be possible for the sublines to perform worse than both parental lines (if they inherited proportionally more deleterious than advantageous mutations). It is also possible that the mutations present in line E6 interact epistatically, such that the deleterious effects of some mutations are cancelled out in the progenitor line.

Several of the data points for line E4 were also excluded on account of many of the worms dying during the assay of what were considered to be unnatural causes (if the worm died after day two, the data were still used in early productivity and if they died after day five, the data were used for all the productivity traits but not longevity). The majority of these deaths were a result of the worms desiccating after crawling onto the plastic edge of the agar plate. Significantly ($p < 0.0001$) more worms from Line E4 and its sublines (17 worms) died in this manner in comparison to deaths from the rest of the experiment put together (only two worms). It is conceivable that line E4 contains a behavioural mutation, which causes them to be more likely to die in this

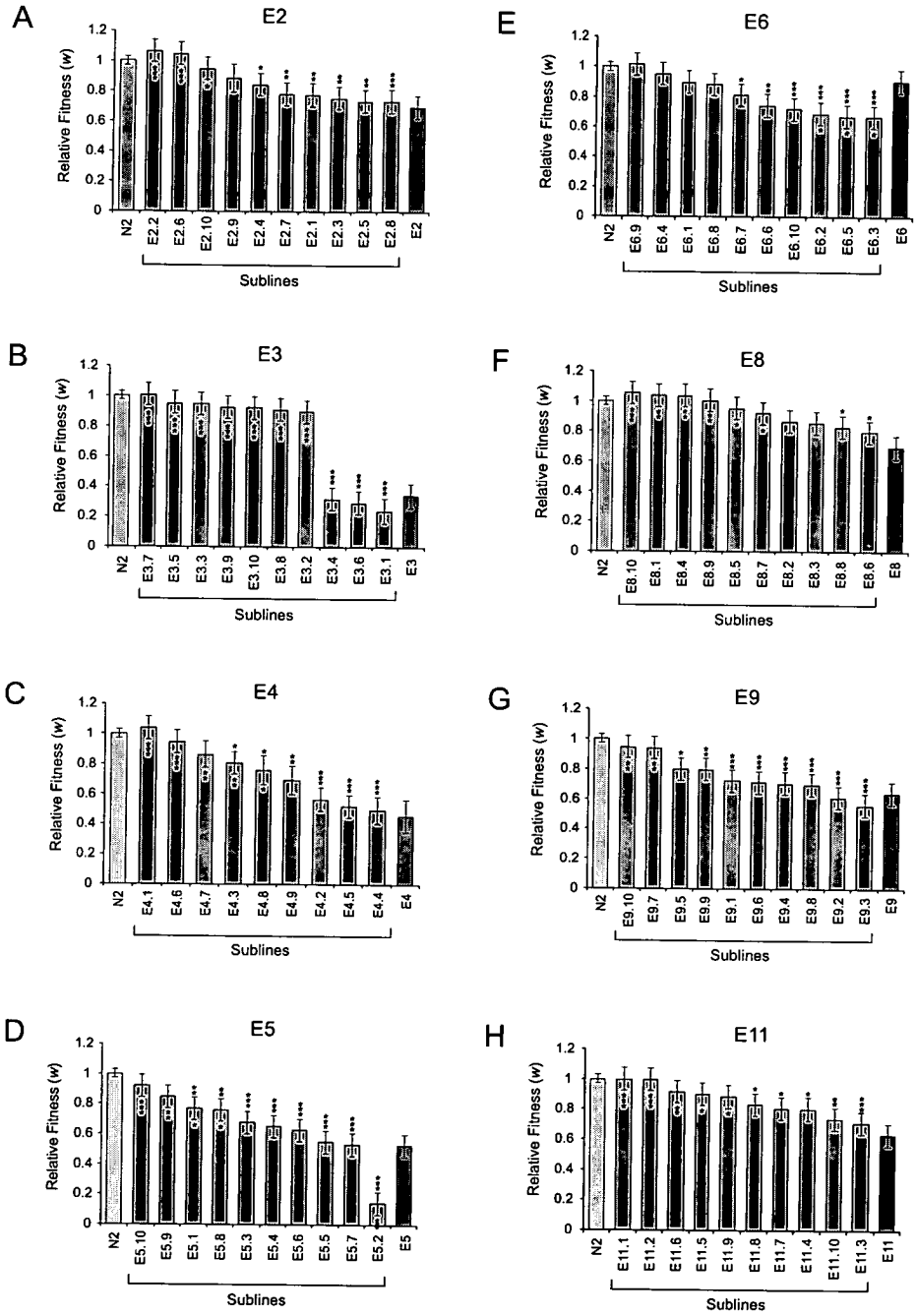


Figure 4.3: Means for relative fitness by line, comparing the means for the two progenitor lines (p-line and N2) with all the sublines generated for that line (± standard error). Asterisks above the error bars correspond to the significance of the difference between the given subline and the wild type. Asterisks below the error bars correspond to the significance of the difference between the given subline and the p-line. $*p < 0.05$; $**p < 0.01$; $***p < 0.0001$.

manner.

To estimate the variability among sublines, a mixed-model analysis was performed (SAS Institute Inc. 1997, Littell *et al.* 1996). The effects of counter and assay on all three traits are non-significant, but there is significant variation among sublines within lines for most traits (Table 4.2, Fig. 4.3). This suggests that much of the variation among sublines within a line is due to a few mutations of large effect, or that there is substantial variation in mutational effects or epistasis among mutations. The large variation in relative fitness among sublines for several p-lines can be seen in Fig. 4.3. For example, the sublines of line E3 (Fig. 4.3B) appear to have a bimodal distribution of relative fitness values, implying that there is one large-effect mutation segregating amongst them. Contrasts between p-line E3 and E3 sublines show that three of the sublines (E3.1, E3.4 and E3.6) are significantly different from N2 ($p \leq 0.0005$) but not E3 whereas the other seven sublines are significantly different from E3 ($p \leq 0.0001$) but not N2 (Fig. 4.3B). This pattern is most striking in line E3, although most sublines for the other p-lines show significant differences from one progenitor, but not the other. Very few sublines were non-significantly different from either progenitor (seven out of 78 excluding subline E5.2), these are sublines E2.9, E6.1, E6.4, E6.8, E6.9, E8.2 and E8.3. Similarly very few sublines were significantly different from both progenitors (seven out of 78 excluding subline E5.2); these are E4.8, E5.1, E5.8, E6.2, E6.3, E6.5 and E4.3. However, in three of these cases the subline performed worse than either parent. This limited evidence is suggestive of one or two major effect mutations (rather than many similarly sized small effect mutations) for most of the lines tested.

The Castle-Wright estimator was used to estimate the effective number of segregating factors within each mutant p-line using the variance amongst sublines from that line and the difference between the means of the two progenitor lines assuming additivity and equal negative effects of mutations. Information from the mixed model was used to obtain estimates of the means of the progenitor lines, the sampling variance of those means and the between-subline variance by line with standard errors, for use in the Castle-Wright estimator. Estimates of the effective number of factors were then averaged over all eight p-lines to give estimates for the effective number of factors, for any given trait (Table 4.3).

Estimates of the effective number of factors using the Castle-Wright estimator

Table 4.2: ANOVA Table for mixed-model GLMs of relative fitness (w), early productivity, total productivity, late productivity and longevity.

Trait	Effect	Variance	df_{num}	df_{den}	F	Z
w	Line		7	76.8	1.82	
	Line-type		1	78.7	6.72*	
	Subline	0.0311				5.07***
	Assay	6.52×10^{-5}				0.28
	Counter	9.32×10^{-5}				0.3
	Residual	0.0587				19.1***
Early Productivity	Line		7	77.1	1.48	
	Line-type		1	77	5.36*	
	Subline	1240				5.21***
	Assay	683				0.323
	Counter	0.928				0.912
	Residual	2110				19.4***
Total Productivity	Line		7	76.3	1.65	
	Line-type		1	78	2.37	
	Subline	1570				5.18***
	Assay	86.7				0.90
	Counter	0.00				-
	Residual	2570				19.1***
Late Productivity	Line		7	77.5	5.46***	
	Line-type		1	80.4	2.21	
	Subline	252				3.73***
	Assay	262				0.98
	Counter	11.8				0.69
	Residual	1430				19.1***
Longevity	Line		7	75.7	0.55	
	Line-type		1	80.6	0.62	
	Subline	0.132				0.67
	Assay	0.360				0.92
	Counter	0.00				-
	Residual	8.76				18.8***

Random effects were estimated by REML and significance was tested with Z scores (right-aligned) rather than F statistics (left-aligned). *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 4.3: Results for Castle-Wright approach for estimating gene number, for the 5 traits studied on a line-by-line basis.

Line	*w*		Early		Total		Late		Longevity	
	n_e (SE)	*s*	n_e (SE)	*s*	n_e (SE)	*s*	n_e (SE)	*s*	n_e (SE)	*s*
E2	1.53 (2.78)	0.248	1.41 (3.57)	0.214	0.31 (0.832)	0.234	0.116 (0.461)	−1.28	−0.228 (2.20)	-
E3	0.769 (0.848)	0.749	0.653 (0.828)	0.753	0.536 (0.593)	0.754	−0.247 (2.97)	-	−16.5 (2390)	-
E4	1.50 (2.09)	0.440	1.23 (2.02)	0.387	1.21 (2.06)	0.318	∞ (-)	-	∞ (-)	-
E5	2.61 (3.59)	0.289	2.60 (4.39)	0.257	0.281 (0.861)	0.198	1.61 (2.90)	−0.970	−0.305 (3.54)	-
E6	0.0552 (0.196)	0.422	−0.0662 (0.337)	-	∞ (-)	-	0.215 (0.621)	−1.26	∞ (-)	-
E8	5.85 (20.0)	0.127	9.27 (46.8)	0.0905	1.18 (2.91)	0.152	−0.151 (0.761)	-	∞ (-)	-
E9	1.98 (3.10)	0.261	2.44 (5.01)	0.211	0.451 (1.57)	0.161	1.05 (2.41)	−0.801	−0.469 (4.25)	-
E11	3.56 (5.91)	0.196	2.12 (4.92)	0.191	0.661 (1.48)	0.193	−0.0161 (0.0701)	-	∞ (-)	-
Mean	2.23 (2.71)	0.341	2.46 (5.96)	-	- (-)	-	- (-)	-	- (-)	-

Approximate standard errors for the number of effective factors are shown in brackets after the estimate along with the average effect (*s*). Overall averages and the standard error of the average are also shown at the base of the Table. Many of the estimates, especially for those traits that were either a small mutational target, or may have had their values both increased and decreased as a result of mutagenesis, were incalculable or negative. If the sampling error of the progenitor lines was large the estimate could be negative, making the calculation of *s* impossible. If the variance amongst the sublines were estimated to be zero, then the estimate for the number of effective factors would be infinite.

are quite low and, despite the large standard errors, are not substantially different from the numbers estimated by Davies *et al.* (1999) or Keightley *et al.* (2000). The Castle-Wright estimator assumes equal effects, but if this assumption is violated then the estimator will underestimate the number of mutations present. Any single large-effect mutation may segregate amongst the sublines produced from a cross. If this occurs, it will lead to a large amount of among-subline variance, making the denominator of the Castle-Wright estimator large and reduce the number of factors estimated; in the extreme case, only one factor would be estimated, even if there were several small-effect mutations segregating in addition to the single large-effect mutation. Therefore, the minimum effect that is possible to detect is highly dependent on the distribution of mutational effects. It is possible to correct for this bias if the variation of effects is known (Zeng 1992); alternatively, a maximum likelihood approach can be used that allows for more than one class of mutation effect.

4.3.2 Likelihood Analysis

The utility of the ML approach described was verified using simulations. The results of these simulations using the one-class ML model are shown in Table 4.4 and the results for the two-class model are shown in Table 4.5. For both the one-class and two-class models, 50 data sets were simulated for each parameter combination and the ML approach was used to estimate the parameter values for each data set. Tables 4.4 and 4.5) show the average estimated parameter values (standard deviations of the estimates are shown in brackets). Mean estimates for all parameters do not differ significantly from the simulated values. However, the estimates of some parameters appear to be noisier than others; estimates of κ (the power of the transformation of the normal distribution) have the largest standard deviations. Since the accuracy of the estimate of κ depends on the number of data points modelled, the two-class model simulations were designed to have a comparable number of data points per simulation to the experimental data. For each simulation, parameter values were estimated from 600 data points (in comparison to 830 data points, for the actual experiment). Over the five sets of simulations, there is a high correlation between the simulated and average estimated values for κ ($r = 0.927$ for one class of mutational effects; $r = 0.997$ for two classes of mutational effects).

The one-class model allows one class of mutational effects and assumes addi-

Table 4.4: Simulation results for maximum likelihood one-class model

Simulated Values				Estimated Values			
λ	s	V_E	κ	λ (SD)	s (SD)	V_E (SD)	κ (SD)
1	0.05	0.001	1	1.10	0.0496	0.000984	0.658
				(0.545)	(0.00840)	(9.28×10^{-5})	(2.39)
1	0.1	0.001	1	0.979	0.0999	0.000979	1.10
				(0.196)	(0.00307)	(8.13×10^{-5})	(1.87)
2	0.05	0.001	1	2.02	0.0487	0.000990	1.04
				(0.498)	(0.00437)	(9.83×10^{-5})	(2.44)
2	0.1	0.001	2	1.98	0.0995	0.000977	1.66
				(0.315)	(0.00192)	(8.86×10^{-5})	(1.72)
2	0.1	0.001	2	2.01	0.100	0.000995	1.85
				(0.305)	(0.00188)	(9.51×10^{-5})	(1.59)

Relative fitness data was simulated according to the models described for the ML analyses. Two sublines for each of 20 p-lines were simulated with three replicate data points per p-line and subline. There were 50 replicates per parameter combination (standard deviations over replicates are in brackets).

tivity; in this respect it is comparable to the Castle-Wright estimator. The number of mutations estimated for traits w, early productivity and total productivity are all similar, low, and not substantially different from the Castle-Wright estimates but have smaller standard errors (Table 4.6). The two-class model allows for two classes of mutations with different effects. It was expected that including variable effects in this way would lead to higher estimates for the number of mutations, with correspondingly lower average effects (Keightley 1998). However, for the three least noisy traits, the most likely mutational model found was a small number ($1.41 \times 0.0884 \simeq 0.12$) of very large-effect mutations (~70%) and a large number ($1.41 \times (1 - 0.0884) \simeq 1.3$) of medium-effect mutations (~20%) (Table 4.6). The large-effect class seems to emerge as a result of the apparent large-effect mutation segregating in line E3 (Fig. 4.3B). With the one-class model, the fitness reduction associated with line E3 can only be explained away with multiple medium-effect mutations; therefore, the number of mutations estimated with the two-class model is lower (albeit not significantly) than that for the one-class model. For all three traits studied, the two-class model fitted significantly better than the one-class model ($p < 0.0001$ in all cases).

The above analysis appears to be dominated by the single large-effect mutation in line E3. Since this may obscure patterns due to smaller-effect mutations in the other

Table 4.5: Simulation results for the maximum likelihood two-class model

Simulated Values						Estimated Values					
λ	s_1	s_2	R	V_E	κ	λ (SD)	s_1 (SD)	s_2 (SD)	R (SD)	V_E (SD)	κ (SD)
1	0.05	0.02	0.4	0.0001	1	0.957	0.0504	0.0201	0.394	9.87×10^{-5}	1.11
						(0.169)	(0.00192)	(0.000956)	(0.108)	(5.79×10^{-6})	(2.57)
4	0.05	0.03	0.6	0.0001	−2	4.08	0.0501	0.0300	0.594	9.86×10^{-5}	−2.14
						(0.390)	(0.000532)	(0.000661)	(0.0661)	(5.40×10^{-6})	(2.71)
2	0.1	0.03	0.6	0.001	1	2.07	0.0993	0.0316	0.588	0.000980	0.876
						(0.379)	(0.00388)	(0.00735)	(0.0943)	(6.38×10^{-5})	(1.13)
1	0.05	0.03	0.4	0.001	2	1.28	0.0475	0.0299	0.354	0.000990	2.27
						(0.495)	(0.0143)	(0.0102)	(0.194)	(6.74×10^{-5})	(1.21)
3	0.05	0.03	0.4	0.001	−1	3.16	0.0534	0.0317	0.422	0.000990	−0.935
						(1.17)	(0.0192)	(0.0138)	(0.263)	(6.31×10^{-5})	(1.28)

Relative fitness data was simulated according to the models described for the ML analyses. Two sublines for each of 30 p-lines were simulated with five replicate data points per p-line and subline. More p-lines and replicates were modelled in this set of simulations than were modelled for the one-class model (Table 4.4) as there were a greater number of parameters to be estimated. There were 50 replicates per parameter combination (standard deviations over replicates are in brackets).

Table 4.6: Maximum Likelihood parameter estimates for both the one-class and two-class models of mutation effects.

Model	Trait	λ (SE)	s (SE)			μ (SE)	V_E (SE)	κ (SE)	*loglik*
One-Class	w	1.64	0.229			1.04	0.0562	1.31	−26.6
		(0.731)	(0.0261)			(0.0330)	(0.00447)	(0.150)	
	Early	1.76	0.162			241	1680	2.25	−1928.9
		(0.706)	(0.0215)			(5.39)	(129)	(0.198)	
	Total	1.49	0.142			278	1930	2.31	−1900.6
		(0.767)	(0.0128)			(4.26)	(158)	(0.187)	
			s_1 (SE)	s_2 (SE)	R (SE)				
Two-Class	w	1.41	0.743	0.213	0.0884	1.03	0.0557	1.32	−21.2
		(0.680)	(0.0589)	(0.0228)	(0.117)	(0.0329)	(0.00444)	(0.150)	
	Early	1.38	0.635	0.163	0.0905	241	1670	2.25	−1920.9
		(0.663)	(0.0408)	(0.0155)	(0.128)	(4.05)	(129)	(0.198)	
	Total	1.81	0.612	0.0814	0.0691	277	1920	2.25	−1891.5
		(0.768)	(0.0496)	(0.0240)	(0.0955)	(5.75)	(158)	(0.183)	

Approximate standard errors are shown in brackets after the parameter estimate. The log-likelihood associated with each parameter combination is shown in the final column. This analysis excludes data for subline E5.2 but includes all data from line E3.

lines, the two-class ML model was applied to the data excluding line E3. The results of this analysis are shown in Figures 4.4A, 4.4B and 4.4C for the trait w; similar results were found for early and total productivity. For w there is virtually no change in log-likelihood above approximately 1.5 mutations (Fig. 4.4A) (lower confidence limit of 0.487 mutations), suggesting that any number of mutations above ~1.5 is equally supported by the data. As this estimate of total mutation number increases, the number of class 1 (medium-effect) mutations in the best fitting model remains constant (at ~1.5); only the number of class 2 (small-effect) mutations increases (Fig. 4.4B), and these have correspondingly lower effects on fitness, such that their total contribution to the average fitness difference remains more or less constant (Fig. 4.4C). The only way to distinguish between a model with a few small effect mutations (*e.g.* four total mutations, ~2.5 of which have very small effects of ~0.8%) and a model with many, very small effect mutations (*e.g.* 20 total mutations, ~18.5 of which have very small effects of ~0.1%), is to use information on the distribution of these mutations amongst the sublines. It is unlikely, given the number of sublines used in this experiment and the level of environmental variation, that it would be possible to distinguish between these distribution patterns. This result was true of all three traits tested. In the case of early productivity the most likely number of mutations tended to infinity and there was no significant change in log-likelihood above 0.542 mutations. For total productivity, the most likely number of mutations was found to be 1.44, and although it is possible to place a lower confidence interval of 0.386, there is no significant change in log-likelihood for increasing numbers of mutations above this. For all traits, when line E3 was removed, a model with two classes of mutations is more likely than a model with one-class, but not significantly so ($p < 0.1$).

Estimates of κ, from the two-class maximum likelihood model including line E3, were tested to see if they increased the fit to normality of the N2 data after it was transformed, using a Ryan-Joiner (correlation-based) normality test (Ryan & Joiner 1976). Since N2 replicates were assumed to have no mutations, the residual data points could be calculated simply by removing the fixed effects estimated from the ML model. N2 data for both w and early productivity departed significantly ($p < 0.025$) from the expectation of a normal distribution without the transformation, but not once transformed ($p > 0.1$). When the same tests were carried out for total productivity, the data did not significantly depart from a normal distribution, with or without the Box-Cox transformation ($p > 0.1$). For relative fitness, a significant increase in the

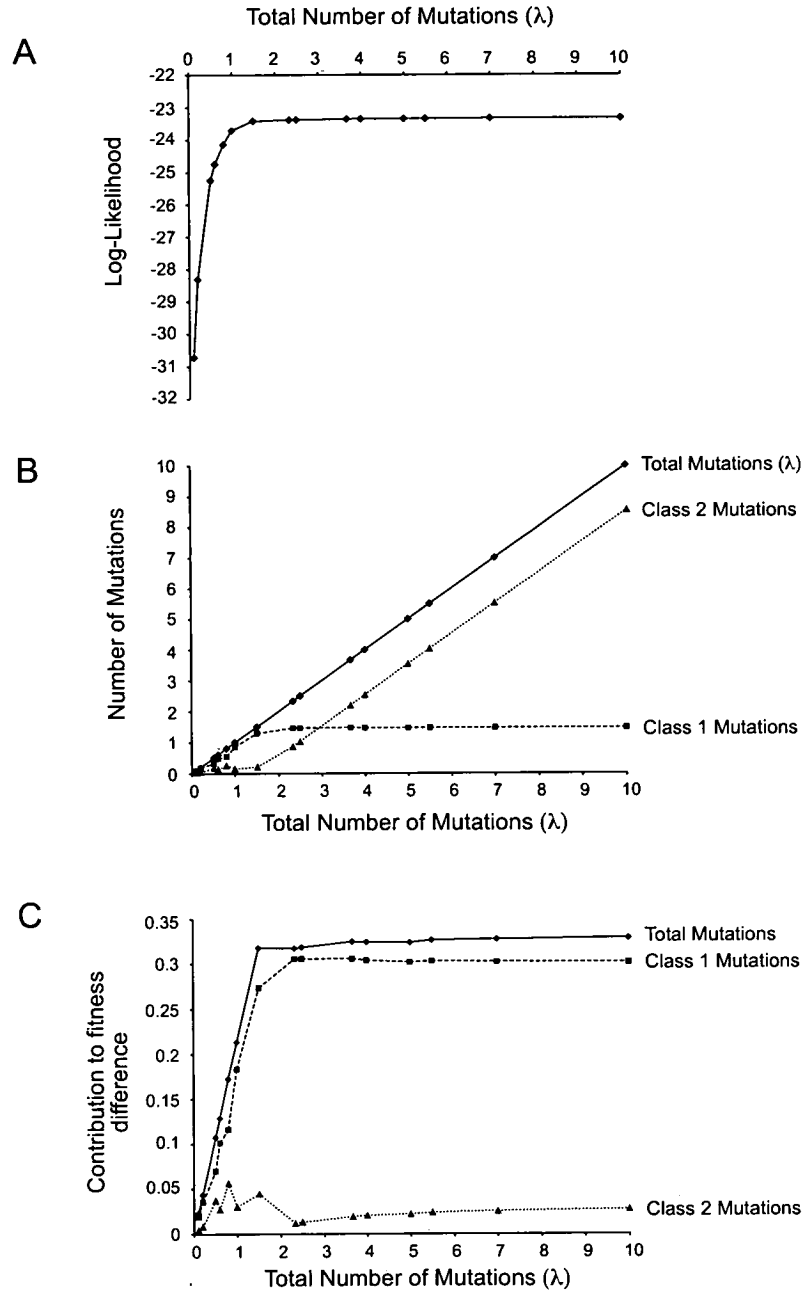


Figure 4.4: Plots of total numbers of mutations against log-likelihood (A), number of mutations (B) and contribution to fitness difference of mutations (C), for class 1 mutations (dashed line with square points), class 2 mutations (dotted line with triangle points) or class 1 + class 2 mutations (solid line with diamond points). The number of class 1 or class 2 mutations was calculated by multiplying the proportion of class 1 or class 2 mutations (R or $1 - R$) by the total number of mutations. The contribution to fitness difference was calculated by multiplying the calculated number of class 1 or class 2 mutations by their estimated effect size (s_1 or s_2).

likelihood ($p = 0.0285$) is obtained when κ is estimated instead of being fixed at 1. The same is true of early productivity ($p < 0.0001$) and total productivity ($p < 0.0001$).

4.3.3 E5.2 and E5 Extra Line Crosses

Unexpectedly, line E5.2 had a significantly lower relative fitness than either of its progenitors, E5 and N2. To investigate this further, 20 sublines were generated from both lines E5.2 and E5 by crossing them to two freshly thawed replicates of the ancestral wild-type (N2A and N2B). Subline number nine generated from line E5.2 was lost during the generations of selfing, due to the extremely low fitness of the line. It was noted that this subline had very slow development time prior to its loss and may, therefore, have failed as a result of fixation of the large-effect deleterious mutation present in E5.2. Even if this lost line is ignored, it is clear that there is one large-effect mutation present in E5.2, which is segregating amongst the sublines (Fig. 4.5A). Sublines 4, 16, 17 and 19, which appear to contain this mutation, are not significantly different from their progenitor line E5.2 but are all significantly different from N2 ($p \leq 0.0001$). Of the remaining 15 sublines, ten are significantly different from both E5.2 and the N2 replicate from which they were generated; only five are not significantly different from line N2. This indicates that there are likely to be some other smaller effect mutations segregating amongst the sublines of this cross.

The number of mutations segregating in line E5.2 was estimated by applying the modified Castle-Wright estimator as described. For relative fitness (w) it was estimated that there were 2.64 mutations segregating in line E5.2 (SE 2.39) with an average effect of 0.645. The number of mutations present in line E5.2 was also estimated using both the one- and two-class maximum likelihood models for relative fitness. Although it was not possible to estimate κ, due to the low number of residual data points, it was possible to get an estimate of the number of mutations. For the one-class model it was estimated that E5.2 contained 1.00 (SE 1.42) mutations with an average effect of 0.957. For the two-class model it was estimated that E5.2 contained 2.00 (SE 2.00) mutations, and that 0.500 of these had an effect of 0.689 whereas the remainder had a lower effect size of 0.278.

All of the extra 20 sublines produced from line E5 had fitness values that were intermediate between the two progenitor lines, and there appears to be no evidence

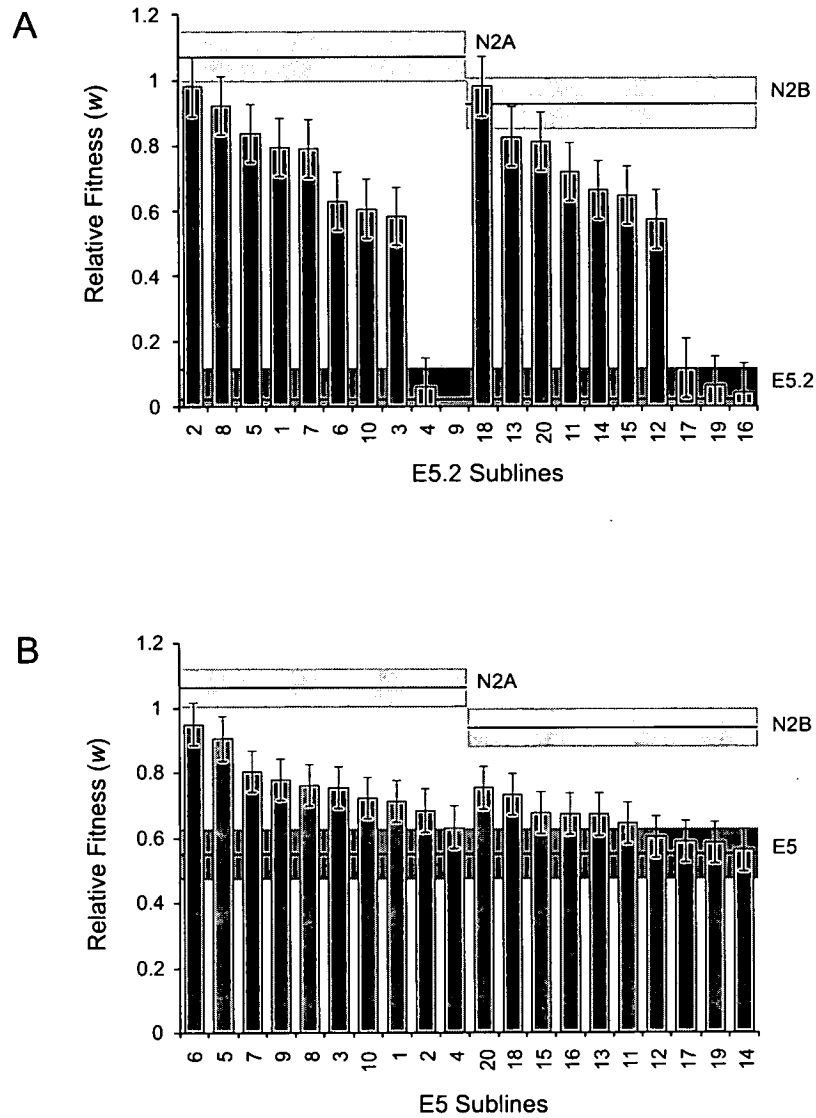


Figure 4.5: Means of the sublines and progenitor lines from the analysis of lines E5 and E5.2. Progenitor lines (E5, E5.2, N2A and N2B) are shown as horizontal bars (± standard error as a grey box) above and below the sublines that they correspond to.

of a single large effect mutation of the size that was observed in the original line E5.2 (Fig. 4.5B). The modified Castle-Wight estimator and the ML approach were applied to the E5 relative fitness data in an attempt to estimate the number of mutations segregating amongst the 20 sublines. Using the Castle-Wright estimator it was estimated that there were 5.88 mutations (SE 8.28) with an average effect of 0.212. Applying the one-class ML model it was found the most likely model contained 8.87 mutations, although this model was not a significantly better fit than any models with more than ~0.5 mutations. The most likely two-class model tended towards the results of the one-class model.

It has thus been established that E5.2 contains a single large effect mutation, but it was not possible to detect this mutation in the progenitor line E5, suggesting that the mutation occurred spontaneously during the generations of inbreeding that produced line E5.2. Alternatively it is possible, although unlikely, that the mutation causing the reduction in fitness is present in line E5 but that another, tightly linked, mutation masked its effects. These mutations may then have been separated after a recombination event during the period of inbreeding that led to line E5.2, but none of the other 29 sublines.

4.4 Discussion

Standard ML methods that attempt to infer the fitness effects of new mutations assume that residual data points (once any mutation effect has been removed) are normally distributed, but this assumption may not be justified. For example, it has been found that the distribution of r, the intrinsic rate of growth of a population, tends to depart strongly from normality (Keightley *et al.* 2000), whereas a related trait, w, has a distribution that is closer to normal. Some efforts have been made to allow for non-normal residuals previously. For example, Vassilieva *et al.* (2000) described the distribution of residual errors for individual replicates and line means, by creating empirical distributions for use in a maximum likelihood analysis. This is not completely satisfactory, however, since discrete (as opposed to continuous) distributions are produced. The inclusion of the Box-Cox exponent as a parameter to be estimated via ML improved the conformation of the data presented here to the assumptions of the ML model. This improvement was tested in two ways: the Ryan-Joiner test of normality and a likelihood ratio test. This method is preferable

to transforming the data set prior to carrying out any analysis because residual values depend on the parameters of the model, and those parameters may depend on the value of κ. Interestingly, despite the significantly improved fit of the model with the transformation, the estimated number of mutations was not substantially affected (for relative fitness under the two-class model, 1.40 mutations were estimated when κ was fixed at 1 as opposed to 1.41 when it was estimated along with the other parameters in the model). It may be possible to extend the ML approach by using other families of power transformations to improve normality and symmetry of residual errors (*e.g.* Yeo & Johnson 2000). Alternative power transformations may also be valid for positive and negative residuals prior to transformation, whereas the Box-Cox transformation requires an additional shift parameter to allow for negative residuals (see Box & Cox 1964).

Davies *et al.* (1999) compared the number of EMS-induced mutations per homozygous line detectable from fitness assays (~2.5 assuming a gamma distribution of effects) to the number estimated to have been induced in the DNA that would be deleterious under natural conditions (~45). The aim of the present experiment was to more accurately estimate the number of induced mutations per EMS-induced mutant line, by producing sublines for a random selection of the mutant lines. The use of sublines allows large-effect mutations to segregate, and it should therefore be possible to determine whether the fitness difference between a wild-type line and a single EMS-induced mutant line is primarily due to few or many mutations with correspondingly large or small effects on fitness. A modification of the Castle-Wright estimator (Castle 1921) has been used to estimate the number of mutations segregating per line and their average effect. With this approach it was estimated that there were 2.23 mutations, on average, affecting relative fitness (SE 2.71) and 2.46, on average, affecting early productivity (SE 5.96). A maximum-likelihood approach to estimate the number of mutations has also been developed, which can allow for variable mutation effects, modelled as two-classes of effects. Under the assumption of two mutation classes, ML estimates of mutation numbers are lower than either the Castle-Wright or ML estimates under a one mutational class model. This surprising result seems to be a consequence of the segregation of a single large-effect mutation in one line (E3), which is modelled as several medium-effect (15–20%) mutations under the one-class model but as a single large-effect mutation under the two-class model. When line E3 was removed from the analysis it was found that the most likely two-class model

consisted of approximately 1.5 medium-effect (~20%) mutations plus a number of smaller effect mutations affecting *w*. However, it proved impossible to determine the number and corresponding effect size of these smaller effect mutations, despite the extra power afforded by producing sublines, since this model was not significantly more likely than a one-class model (for all traits $p > 0.1$). The data presented here are therefore consistent both with a model with several small effect mutations (~3 mutations with an effect size of ~1%) and a model with many very small effect mutations (>20 mutations with an effect size <0.2%). Distinguishing between these models would clearly require very much more data.

The estimates of mutation number are dependent on how variability in effects of mutations is treated. If it is assumed that all mutations have the same effect then it is possible to obtain a concrete estimate of their number, but this is not possible if two classes of effects are assumed. Unfortunately it was not possible to test the fit of a gamma or other continuous distribution because of the limits of computing power, but it is possible that such an analysis could provide greater support for a leptokurtic distribution of mutation effects than the two-class model.

There are at least three possible explanations for the difference between the numbers of mutations estimated to have been induced and the number of mutations detected at the phenotypic level. If the estimate of at least 45 deleterious mutations induced per p-line is correct, then the results presented here suggest that the distribution of mutation effects is highly leptokurtic, and that a large class of mutations have undetectable effects in laboratory assays. This is consistent with several other direct and indirect estimates of the shape of the distribution of mutation effects. For example, transposable elements provide an opportunity to control the number of mutational events at the DNA level, and experiments using these have provided estimates of the distribution of mutation effects. Analyses of the effects of P-element insertions in *D. melanogaster* on metabolic parameters (Clark *et al.* 1995) and bristle numbers (Lyman *et al.* 1996) suggest that mutations with the smallest effects are the most frequent. Similarly, there is direct evidence from Tn10 insertions in *E. coli* for an L-shaped distribution of mutational effects (Elena *et al.* 1998, Elena & Lenski 1997). A second possibility is that each p-line carries many fewer than 45 deleterious mutations, on average, since the estimates of the number carried are indirect. A possible way to resolve this would be to directly estimate the number of mutational events at the DNA sequence level (Denver *et al.* 2000). Finally it is possible that assaying fitness

under standard lab conditions would not reveal every large effect deleterious mutation, and that assaying fitness under a variety of environments could reveal many more potentially large effect deleterious mutations.

If the distribution of mutation effects is L-shaped and the vast majority of deleterious spontaneous mutations have nearly neutral (but still deleterious) effects on fitness then this could have implications for several areas of evolutionary theory. For example, mildly detrimental mutations on the border of neutrality are the most damaging to population viability if the effective population size is larger than a few individuals (Lande 1994). Secondly, if there is a substantial fraction of mutations that are nearly neutral, then deterministic population genetic models of the mutational process might be insufficient necessitating the use of models that incorporate genetic drift as a factor that influences the fate of mutations. For example, the mutation load might be much higher than would be expected under a deterministic model, since many mutations of small effect could drift to fixation. Finally, mutations of very small effect are undetectable in the vast majority of fitness assays, leading to underestimates of the mutation rate, which has implications for the understanding of the evolution of sex. It is thought that that the diploid mutation rate must be above one per generation for sexual reproduction to be maintained by deleterious mutations alone (Kondrashov 1988, 1995). Many estimates of the mutation rate from mutation accumulation experiments that do not account for variability in the effects of mutations fall well below this value (see reviews in Drake *et al.* 1998, Keightley & Eyre-Walker 1999, Lynch *et al.* 1999); however, these may be substantial underestimates if the degree of variation in mutation effects is high.

5 Estimating the Average Dominance Coefficient of Mildly Deleterious Induced Mutations for Fitness Traits in *Caenorhabditis elegans*

The experimental design and implementation of work described in this Chapter were carried out by myself as part of a team headed by Andy Peters. The re-assay to verify heterosis in a subset of the lines was carried out as part of a team, headed by myself. All analyses of the data were performed by myself. This work has been published (Peters *et al.* 2003).

5.1 Introduction

Obtaining estimates of the rate (U) and the average homozygous effect ($\overline{s}$) of mutations has been the aspiration of many evolutionary biologists, and a great deal of research in recent years has attempted to obtain such estimates. However, much less attention has been paid to estimating the heterozygous effects of mutations. The heterozygous effect of a mutation can be related to the homozygous effect by the dominance coefficient. Let the fitness of a wild-type chromosome be 1, and the fitness of a mutant homozygote be $1 - s$, then the fitness of the mutant heterozygote can be represented as $1 - hs$, where h is the dominance coefficient. A mutation acts additively if $h = 0.5$, it is recessive if $0 < h < 0.5$, and dominant if $0.5 < h < 1$. If h is > 1 or < 1 then the mutation is under- or over-dominant respectively.

Knowing the average dominance coefficient of mutations ($\overline{h}$) is crucial to our understanding of evolution, because it is an important parameter for many models in evolutionary biology. For example, $\overline{h}$ is important for theories concerning the maintenance of genetic variability (Charlesworth & Hughes 1999). Under a simple deterministic model of mutation-selection balance, the additive genetic variance for fitness at equilibrium is equal to twice the product of the haploid deleterious mutation

rate (μ) and the average heterozygous effect ($V_A = 2\mu sh$) (Mukai *et al.* 1974). $\bar{h}$ is also an important parameter for models of inbreeding depression, since the number of mutations that can be hidden in the heterozygous state, in a population at mutation-selection balance, decreases as h increases (assuming that deleterious mutations are partially recessive, on average) (Hartl & Clark 1997). The decline in fitness expected when such a population undergoes inbreeding therefore increases as h decreases. In other words, the more recessive deleterious mutations are, the more severe the consequences of inbreeding. Inbreeding depression itself bears importantly on many aspects of evolutionary biology, including the evolution of various aspects of mating system evolution in plants (Lande & Schemske 1985, Charlesworth & Charlesworth 1987, Charlesworth *et al.* 1991) and the conservation of endangered species (*e.g.* Lynch *et al.* 1995*a*,*b*).

Several types of experiments have been carried out to attempt to estimate $\bar{h}$. The most direct of these investigate mutations that have been allowed to accumulate within a genome or chromosome under conditions of relaxed natural selection, a technique pioneered by Terumi Mukai (Mukai *et al.* 1964, 1965, Mukai & Yamazaki 1968, Mukai 1969*a*). Mukai used a balancer chromosome system (*Cy/Pm*) in *Drosophila melanogaster* to maintain wild-type second chromosomes in the heterozygous state, protected from selection for 30–60 generations. This procedure yielded second chromosomes that carried sets of deleterious mutations. To estimate $\bar{h}$ for mildly deleterious mutations, chromosomes with ≥50–60% normal viability ("quasinormals") were selected, and the viabilities of these chromosomes were assayed in the homozygous and heterozygous states, alongside controls that were homozygous for wild-type or wild-type-like second-chromosomes. Similar experiments were subsequently carried out by Ohnishi (1977*b*) (see reviews in Simmons & Crow 1977, Houle *et al.* 1997 and García-Dorado & Caballero 2000).

The results obtained from Mukai and Ohnishi's experiments depended on whether coupling or repulsion heterozygotes were used and also on the method used to calculate $\bar{h}$. Coupling heterozygotes are formed when a chromosome that has accumulated mutations is combined with a chromosome presumed to be free of mutations, whereas repulsion heterozygotes are formed by randomly combining two chromosomes that have both accumulated mutations. Results from the mutation accumulation (MA) experiments also depended on whether a regression or ratio approach was used to calculate $\bar{h}$. $\bar{h}$ can be estimated by calculating the ratio of

the heterozygous effect to the homozygous effect, giving an average weighted by the selection coefficient (s) (Mukai 1969*b*, hereafter referred to as $\bar{h}_1$). Alternatively, $\bar{h}$ can be calculated as the regression coefficient of the heterozygous effect on the sum of two homozygous effects of the parental chromosomes, yielding an estimate of $\bar{h}$ weighted by the square of the selection coefficient (s^2) (Mukai *et al.* 1972, Mukai & Yamaguchi 1974, hereafter referred to as $\bar{h}_2$). There are reasons to believe that the ratio estimate is less biased and less misleading, since the regression approach makes two assumptions that are unlikely to be correct (Caballero *et al.* 1997). These are: (i) that the homozygous effects are known precisely and (ii) that the dominance coefficients of alleles are uncorrelated with their homozygous effect. However, if the regression coefficient is calculated as the ratio of the genetic covariance between heterozygotes and homozygotes to the genetic variance amongst homozygotes estimated from an ANOVA (*e.g.* Mukai & Yamazaki 1968, Mukai *et al.* 1972), then one of these assumptions (that the homozygous effects are known precisely) can be ignored.

The estimates obtained by Mukai and Ohnishi ranged from −0.32 to 0.49 (see reviews in Simmons & Crow 1977, García-Dorado & Caballero 2000) and generally suggest that mutations are partially recessive on average ($0 < \bar{h} < 0.5$). However, there is considerable variability among the estimates. Coupling heterozygotes formed by pairing with the "original" chromosome consistently yielded negative estimates of $\bar{h}$ by either method of calculation ($\bar{h}$ estimates range from −0.32 to −0.09; Mukai *et al.* 1964, Mukai 1969*a*). On the other hand those formed by pairing with non-isogenic chromosomes yielded estimates that were partially recessive ($\bar{h}$ estimates range from 0.09 to 0.13) when calculated using the proportional difference in means approach (Simmons & Crow 1977) or nearly additive when calculated using a regression approach ($\bar{h}$ = 0.27 – 0.56; Mukai *et al.* 1965). Repulsion heterozygotes yielded estimates that were nearly additive by either method of calculation ($\bar{h}$ = 0.36 – 0.46; Mukai & Yamazaki 1968). Ohnishi's estimates of $\bar{h}$ suggest that mutations act additively when $\bar{h}$ is calculated via the proportional difference in means approach ($\bar{h}$ = 0.40 – 0.48 for either coupling or repulsion heterozygotes; Ohnishi 1977*b*) but suggest they are partially recessive when $\bar{h}$ is calculated by regression ($\bar{h}$ = 0.12 – 0.15; García-Dorado & Caballero 2000).

The interpretation of this variation among results has recently been the subject of active debate; García-Dorado & Caballero (2000) have suggested that a non-mutational decline in fitness in Ohnishi's experiment could have inflated his estimate

of $\bar{h}$. On the other hand, Fry (2001) has suggested that the wide variation in estimates of mutational parameters in general might reflect actual differences among strains. In a separate MA experiment (using a standard balancer chromosome system) Houle *et al.* (1997) estimated the mean dominance coefficient for five different life-history traits (other than viability) and found mutations to be partially recessive on average (using a regression approach, $\bar{h}_2 = 0.12$). It should also be noted, however, that Houle *et al.* included non-quasinormal chromosomes in their analysis, which might be expected to bias the estimate of $\bar{h}$ downward, since highly deleterious mutations tend to be highly recessive (Caballero & Keightley 1994).

Vassilieva *et al.* (2000) recently reported several estimates of $\bar{h}$ from 73 MA lines in *Caenorhabditis elegans*. In this experiment, independent lines were allowed to accumulate mutations for 170 generations, after which time they were crossed to cryopreserved controls, and the F1 performance for six life-history characters was assayed alongside the parental MA lines. $\bar{h}$ was then calculated, using the regression approach, by dividing the covariance between heterozygote and homozygote line mean phenotypes by the variance among the homozygote mutant lines. Estimates of $\bar{h}_2$ ranged from -0.10 to 0.69 for the 6 traits measured (0.38 on average), although standard errors for individual traits were relatively large (0.21 on average).

Estimates of $\bar{h}$ have also been obtained using induced mutations (Mukai 1970, Ohnishi 1977*b*), P-element insertions in *Drosophila* (MacKay *et al.* 1992, Lyman *et al.* 1996), chromosomes extracted from equilibrium populations (Mukai *et al.* 1972, Mukai & Yamaguchi 1974, Watanabe *et al.* 1976, Eanes *et al.* 1985, Hughes 1995) and by crossing individuals from inbred plant populations (Johnston & Schoen 1995). Estimates obtained using ethylmethane sulphonate (EMS) induced mutations are in line with those from spontaneous mutations, and suggest partial recessivity (Mukai obtained estimates of $\bar{h}$ from 0.022 to 0.037; Ohnishi's estimates ranged from 0.27–0.47). Induced mutations have the advantage that the effects of many mutations can potentially be tested, without the need for long periods of MA. However, it is possible that the results obtained are not reflective of spontaneous mutations. Experiments testing the effects of P-element insertions in *Drosophila* have suggested that these inserts have completely recessive effects on viability on average (MacKay *et al.* 1992, Lyman *et al.* 1996), although again this may not be applicable to spontaneous mutations. Estimates from chromosomes in equilibrium populations and from inbred natural populations suggest that mutations tend to be partially recessive on average.

However, since mutations with low dominance are expected to contribute more to genetic variation at equilibrium than those with higher dominance, such estimates will be underestimates of $\bar{h}$ for newly arising mutations. Additionally, if genetic variance in equilibrium populations is not solely due to mutation-selection balance (for example, if it is maintained by selection) this may bias results by an unknown amount (Drake *et al.* 1998).

If a general conclusion can be drawn from these estimates of $\bar{h}$, it might be that newly arising mutations tend to be partially recessive, although, there is wide variation among current estimates. It therefore seems clear that further estimates are desirable, preferably from other fitness-related traits in a variety of species. The aim of this experiment was to estimate the average dominance coefficient of mildly deleterious EMS-induced mutations in *C. elegans*. EMS-induced mutations provide a powerful system for this purpose, since the effects of many mutations can be studied without the need for many generations of mutation accumulation, and it has been argued that their effects are similar to those of spontaneous mutations (Davies *et al.* 1999). In addition, using *Caenorhabditis elegans* as the study organism makes it relatively straightforward to assay heterozygotes and both parental homozygotes contemporaneously, under the same conditions.

5.2 Materials and Methods

A subset of the EMS-induced mutant lines generated by Davies *et al.* (1999) (see Subsection 2.2.1) were used to measure dominance in this experiment. Lines with particularly low fitness may contain mutations of large effect, and these types of mutation tend to have small dominance coefficients (see Simmons & Crow 1977, Caballero & Keightley 1994). Including lines with low fitness could therefore bias the estimate of $\bar{h}$ (for mildly deleterious mutations) towards recessivity and so only lines that performed well in previous assays (Davies *et al.* 1999, Keightley *et al.* 2000) were chosen. The chosen lines were thawed from storage at −85°C (Sulston & Hodgkin 1988) in 4 batches. In each case a single wild-type strain (N2) was thawed for use as a control for that particular block. Thirty three lines were initially selected for this experiment, although 14 could not be used, due to production of inviable males or a failure to produce a 1:1 sex ratio when crossed, which suggests that the planned crossing had not occurred. Bacterial and fungal contamination that

had occurred during freezing and thawing were controlled using both streptomycin-containing plates (see Subsection 2.2.3) and alkaline hypochlorite treatment of thawed adult worms (see Subsection 2.2.2). Worm lines were then maintained at 20°C for up to eight generations, to remove any effects associated with thawing, by transferring one hermaphrodite from each line to a fresh plate every generation.

Crosses to produce heterozygous worms were carried out in both directions, *i.e.* with the hermaphrodite parent as the mutant (M-heterozygotes) and with mutant paternal parents (P-heterozygotes). It was therefore necessary to generate male worms for the wild-type line (N2) and all mutant lines. Males were generated by heat-shocking hermaphrodites at 25.5°C (see Subsection 2.2.5). After heat shock, any resulting male progeny were returned to 20°C and crossed to hermaphrodites from the same line in the ratio four to six males to one hermaphrodite, to maximise the crossing rate. All worm lines that produced viable males (including N2) were then maintained at 20°C for three to eight generations as both hermaphrodite and mixed-sex lines. Mixed-sex lines were maintained by setting up crosses each generation between four to six males and one hermaphrodite. Hermaphrodite lines were maintained by transferring one hermaphrodite each generation to a fresh plate.

Assays were carried out for two fitness-related traits, viability and productivity (see Figure 5.1 for a diagram illustrating the methodology for both assays). For each of these traits, homozygotes (wild-type and mutant) and both M- and P-heterozygotes were assayed simultaneously. All assayed worms (including the homozygotes) were produced by crossing, to remove any effects associated with selfing. This was ensured by placing four to six young adult males on a plate with a single late L4 stage hermaphrodite. A previous pilot experiment has shown that in such a cross, the sex ratio of offspring is normally 1:1 after the 3rd day (see Figure 5.2), suggesting that progeny produced after this point are unlikely to be the result of selfing. Six M-heterozygote, six P-heterozygote and ten homozygote crosses were set up for each mutant line, alongside 50 N2 homozygote crosses for each assay.

On the third day after setting up crosses, adult hermaphrodites were moved onto fresh agar plates for six hours, to lay eggs, and then removed. These eggs were used for both the viability and productivity assays. Viability was measured by counting the number of eggs laid and number of larval and adult offspring surviving after three days. Pilot studies revealed that eggs laid around the edge of a plate can be difficult

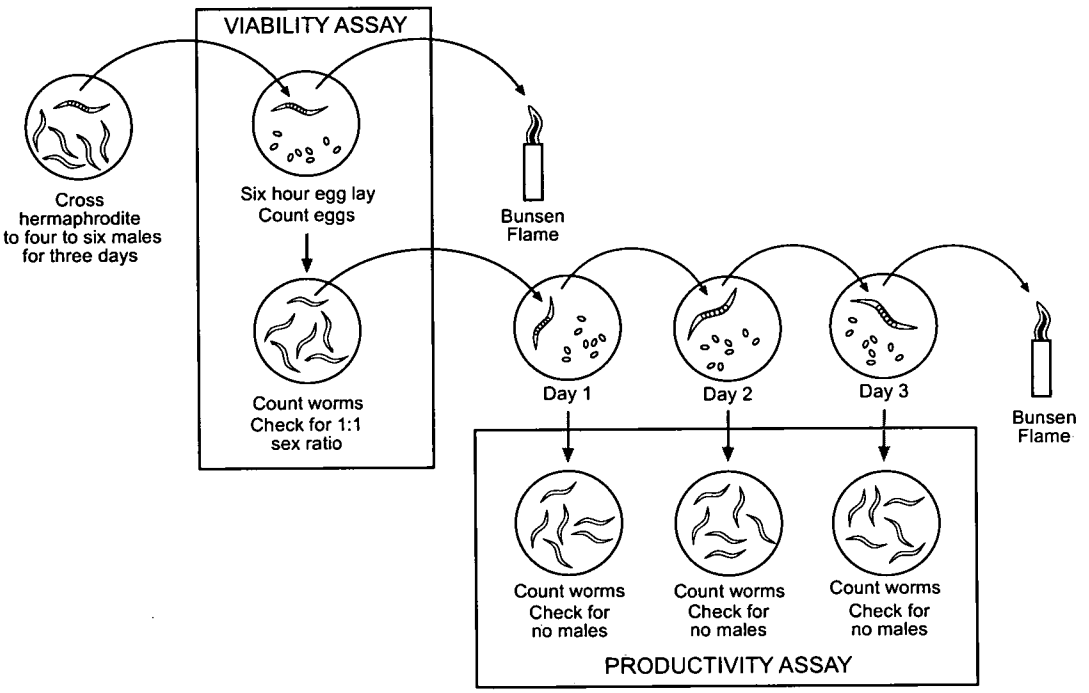


Figure 5.1: The methodology for the viability and productivity assays. Grey straight lines indicate where plates are incubated and individuals are allowed to develop from eggs to worms. Curved lines indicate the transfer of a worm between plates or where a worm is removed and killed.

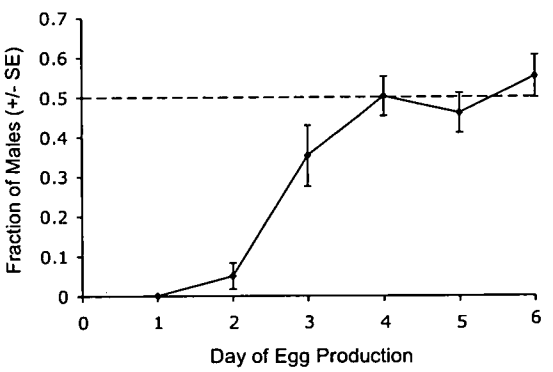


Figure 5.2: Results of a pilot study showing the fraction of male offspring produced over a number of days after crosses were set up. Crosses were set up using single L3 hermaphrodites and four to six young adult males.

to count, which can lead to undercounts of the number of eggs produced. In order to obtain more accurate egg counts, hermaphrodites were forced to lay eggs in the centre of the plate using double-walled enclosures, which were placed over hermaphrodites immediately after transfer to the agar plate. In the event that a worm escaped from the enclosure, plates were excluded from the viability assay but the eggs laid were used in the productivity assay. Following egg-laying, the plates were incubated at 20°C for three days, after which time the numbers of male and hermaphrodite larval and adult offspring were counted. At this stage the offspring sex-ratio was checked to ensure that it did not significantly differ from 1:1 using a χ^2 test. Plates with significantly more hermaphrodites than males were excluded from both assays.

A number of L3 stage hermaphrodite offspring were taken from the six hour egg-lay plates and moved on to fresh plates to assay the productivity of individual worms. L3 stage hermaphrodites were taken from egg-lay plates, to prevent crossing with males on the same (egg-lay) plate, and to avoid missing any of the worm's productivity. If sufficient offspring were available, 36 homozygote offspring and 18 M-heterozygote and 18 P-heterozygote offspring were taken for each mutant line along with 96 wild-type offspring. If possible, these worms were taken from a range of different egg-lay plates. A stock of extra worms (6–20) was maintained in parallel to replace the main stock if males were found on the productivity plates, if the hermaphrodite was accidentally killed, or if any of the plates became contaminated. Worms were allowed to lay eggs for three days and were moved to fresh plates every 24 hours during this period. After three to four days, productivity was measured by counting the number of offspring produced by each individual worm. If more than two males were produced by a single hermaphrodite over the three days of recorded productivity, the data for that hermaphrodite was ignored, since these offspring could have been the result of crossing. Productivity was only assayed for three days because this incorporates the majority of a worm's production of eggs, and productivity on days four and five contribute little to relative fitness.

Viability was defined as the ratio of the number of adult worms to the number of eggs counted and this trait was analysed using mixed models (proc GLIMMIX and proc MIXED) in SAS (SAS Institute Inc. 1997, Littell *et al.* 1996). Viability for individual worms is either zero or one, so viability for a single plate is expected to be binomially distributed and should therefore be analysed assuming a binomial error structure and a logit link function. This assumption requires that viability is not greater

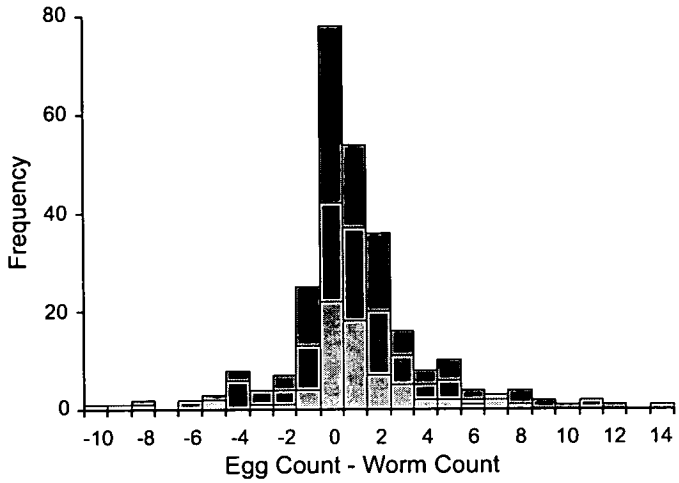


Figure 5.3: Distribution of egg minus worm counts for homozygote (dark shading), heterozygote (intermediate shading) and wild-type plates (light shading). Results should all be positive (or zero) if egg counts are perfect.

than one, or, in other words, that egg counts are perfect. However, it was found that for 53 out of 272 of the egg-lay plates, the worm count exceeded the egg count (in 27 of these cases the worm count exceeded the egg count by more than one). Viability was therefore analysed excluding all plates where the worm count exceeded the egg count, using a generalised linear mixed model, and assuming a binomial error structure and logit link function. However, ignoring plates where eggs were undercounted could potentially bias the results, if the proportion of plates with egg undercounts is different for heterozygotes *vs.* mutant homozygotes *vs.* wild-type homozygotes. Indeed the ratios of undercounted worms for these genotypes varies (13% of wild-type worms, 26% of heterozygotes and 18% of homozygotes) (see Figure 5.3). To asses the bias introduced by excluding plates with egg undercounts, viability was also analysed including data where eggs were undercounted, using a mixed model assuming a normal error structure. Fixed effects included in the models were zygosity (zyg), (*i.e.* heterozygote *vs.* homozygote), and reference-parent genotype (refpar), (*i.e.* whether the mutant parental worm was maternal or paternal, nested within zyg). Random effects in the model were assay, reference genotype (refgt) (*i.e.* the line of the mutant parent) and the interaction between refgt and zyg.

Total productivity is defined here as the unweighted sum of worms produced over the three days of recorded productivity. Relative fitness (w) is calculated as a weighted sum of the three days of productivity, such that early days are weighted more heavily.

This quantity is proportional to the expected fitness of a population with a stable age-structure and is calculated as $w = \sum_x e^{-r_c x} l_x m_x$ (Charlesworth 1994), where l_x and m_x are the survivorship to, and productivity at, day x respectively. Mean fitness of the wild-type was set at 1 for each individual assay, so r_c could be calculated for each assay by solving $1 = \sum_x e^{-r_c x} l_x m_x$, using the average $l_x m_x$ across all the wild-type replicates within a given assay. Individual estimates of w were then calculated by substituting r_c, for the relevant assay, into the above formula.

Productivity and w were analysed using a general linear mixed model (SAS Proc MIXED) (SAS Institute Inc. 1997, Littell *et al.* 1996). Fixed effects in the model were zygosity (zyg) and reference-parent genotype (refpar) (nested within zyg). Random effects included in the model were assay (1–4), counter (the person who counted the worms, nested within assay) (1–4), reference genotype (refgt), the interaction between refgt and zyg, and family (*i.e.* offspring from the same cross, (nested in assay, refgt, zyg and refpar). Due to the fact that random effects were estimated by REML, Z-tests were used to test significance between means assuming a normal distribution of residuals.

5.2.1 Verification of Heterosis

Three lines (E11, E13 and E25) showed significant evidence of heterosis for relative fitness. These lines were re-assayed using a modified version of the productivity assay described above. In this re-assay, two separate samples of wild-type (N2) worms (N2A and N2B) (instead of just one) and one sample of each of the lines E11, E13 and E25 were thawed. Males were generated from all lines (including both N2A and N2B), and the lines were maintained as four to six mixed sex and four to six pure hermaphrodite families for three or more generations prior to crossing. Fifteen to twenty homozygote crosses, and seven to ten each of M-heterozygote and P-heterozygote crosses were set up for each mutant line, to yield 80–82 homozygous individuals and 39–41 individuals of each type of heterozygote (M and P). Unfortunately, line E25 failed to outcross, and so was excluded from further analysis. The other two lines outcrossed successfully and were assayed for productivity alongside 50 worms from each N2 sample. Productivity and relative fitness for the re-assay were analysed using SAS Proc MIXED. Fixed effects included in the model were zygosity (zyg) and reference-parent genotype (refpar), (nested within zyg). Random effects were counter, reference

genotype (refgt), the interaction between refgt and zyg, N2-type (*i.e.* N2A *vs.* N2B, nested within refpar and zyg) and family (nested within refgt, refpar and zyg).

5.2.2 Calculation of the Average Dominance Coefficient

The average dominance coefficient was calculated as the proportional reduction in a fitness related trait value among heterozygotes relative to homozygotes (averaged over all lines).

$$\bar{h}_1 = \frac{\overline{(X_{WT} - X_{het})}}{\overline{(X_{WT} - X_{hom})}} \tag{5.1}$$

This gives an estimate of dominance weighted by the selection coefficient ($\bar{h}_1 = \sum sh / \sum s$) (Mukai 1969*b*). Dominance can also be estimated as the regression of the trait value in heterozygotes on that in homozygotes. This gives an estimate of $\bar{h}_2$ weighted by the squared selection coefficient ($\bar{h}_2 = \sum s^2 h / \sum s^2$) (Mukai *et al.* 1972, Mukai & Yamaguchi 1974). As discussed in Section 5.1, estimates from the regression approach may be biased or misleading (Caballero *et al.* 1997).

5.3 Results

5.3.1 Mutational Effects on Viability

A total of 6197 eggs and 5933 worms from a total of 273 plates were counted across all viability assays. However, for 53 of the plates, the number of worms counted exceeded the number of eggs, due to undercounts of eggs. When plates with egg undercounts were included, the estimated mean viability of wild-type (N2), heterozygous and homozygous worms was 0.962 (0.0189), 0.987 (0.0172) and 0.941 (0.0181) respectively, and none of these means were significantly different. When viabilities were calculated on a line-by-line basis, only two estimates differed significantly from the wild-type (see Figure 5.4). These were E37 heterozygotes and E29 heterozygotes, both of which had significantly higher estimates of viability, although this effect was solely due to the inclusion of several plates where the worm count was much greater than the egg count. None of the fixed or random effects included in the mixed model were significant. Similarly, when egg undercounts were

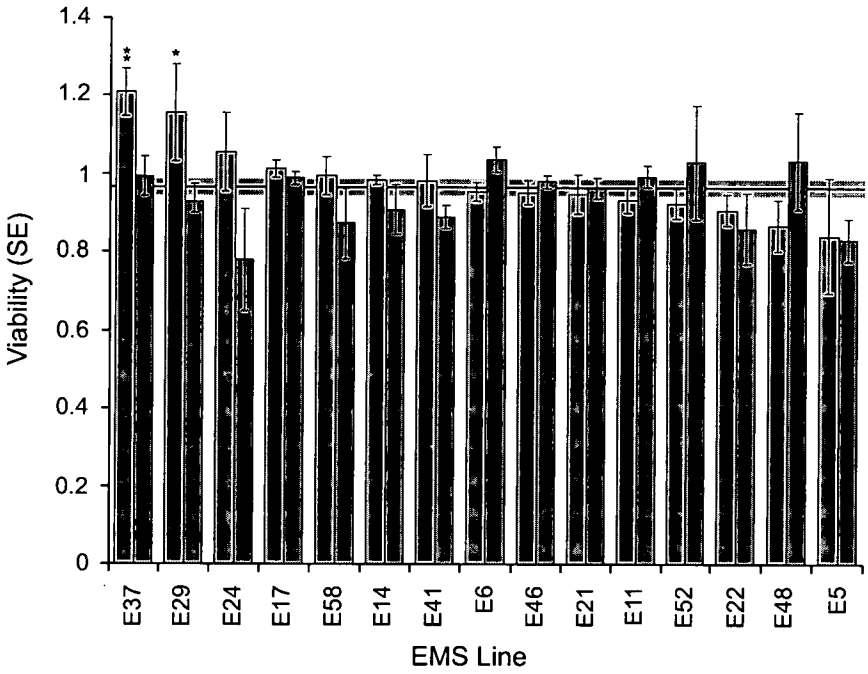


Figure 5.4: Estimates of heterozygote (lightly shaded bars) and homozygote (darkly shaded bars) viability (±SE) for individual lines. The mean of the wild-type (N2) is shown as a horizontal bar (±SE grey box).

excluded, none of the factors included in the generalised linear mixed model were significant (see Table 5.1 for ANOVA tables from proc MIXED and proc GLIMMIX for data with and without egg undercounts respectively). The mean viability of wild-type, heterozygous and homozygous worms was 0.920 (0.0122), 0.917 (0.0126) and 0.905 (0.0133) respectively, and again none were significantly different from one another. Unfortunately, there were too little data to reliably estimate viability for individual lines when plates with egg undercounts were excluded.

The results from this assay are imprecise due to the small number of independent data points and unreliable egg counts; not only do a large fraction of plates have larger worm counts than egg counts, but in many cases the worm count exceeds the egg count by more than one (see Figure 5.3). In addition, the average viability of heterozygotes varies substantially, depending on whether plates with egg undercounts are included or not. This assay can therefore not be used to estimate the average dominance coefficient for viability. Nevertheless, it appears that homozygotes have lower viability than wild-type worms, as would be predicted (this is true whether or not egg undercounts are included), although the differences are non-significant in both cases. Furthermore,

Table 5.1: ANOVA table for proc MIXED analysis of viability (including egg undercounts) and proc GLIMMIX analysis of viability (excluding egg undercounts).

Analysis	Effect	Variance (SE)	df_{num}	df_{den}	F	Z
MIXED	zyg		1	14	2.03	
	refpar		1	208	2.43	
	refgt	0.000117 (0.00267)				0.04
	refgt × zyg	0.00324 (0.00336)				0.97
	assay	0.0 (-)				-
	residual	0.0267 (0.00260)				10.28***
GLIMMIX	zyg		1	13	0.31	
	refpar		1	159	2.10	
	refgt	0.00122 (0.00106)				1.15
	refgt × zyg	0.000172 (0.000988)				0.17
	assay	0.0 (-)				-
	residual	0.0120 (0.00134)				8.98 ***

Random effects were estimated by REML and significance was tested using Z scores. Significance of fixed effects was tested using F statistics. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

it seems that the overall effect of mutagenesis on viability is small compared to the other traits analysed (either in the homozygous or heterozygous state). In this sense these results are consistent with those of Vassilieva *et al.* (2000) who showed that the related trait, survival to maturity, was subject to low mutation rates ($U = 0.003$ (SE 0.0011) per genome per generation). Productivity assays should, therefore, not be greatly affected by differences in the viability of F2 progeny, which may have variable genotypes depending on the type of cross.

5.3.2 Mutational Effects on Productivity and Relative Fitness (w)

A total of 1,606 parent worms were used to measure productivity. On average, 33 homozygote, 14 M-heterozygote and 18 P-heterozygote worms were assayed for productivity for each line, alongside 360 wild-type worms. A total of 318,310 worms were counted for all four productivity assays. For both productivity and w, heterozygote and wild-type worms performed significantly better than homozygotes ($p < 0.0001$ in all cases). However, heterozygotes do not appear to be significantly different from wild-type for either productivity ($p = 0.391$) or w ($p = 0.597$),

Table 5.2: Means and standard errors for total productivity and *w*.

Parameter	Prod. estimate (SE)	*w* estimate (SE)
WT Mean	214 (3.34)	1.00 (0.0265)
Het Mean	211 (2.46)	0.988 (0.0184)
Hom Mean	184 (2.52)	0.790 (0.0193)
Het – WT	−3.49 (4.06)	−0.0169 (0.0319)
Hom – WT	−30.2 (4.06) **	−0.215 (0.0322) **
Hom – Het	−26.7 (3.35) **	−0.198 (0.0261) **
Het – (Hom + WT)/2	11.6 (3.12) *	0.0907 (0.0243) *
$\bar{s}$ (Hom – WT)/WT)	0.14 (0.019)	0.22 (0.032)
$\bar{h}_1$ (Het – WT)/(Hom – WT))	0.12 (0.12)	0.079 (0.14)

Standard errors for $\bar{s}$ and $\bar{h}_1$ were calculated using the delta method (Lynch & Walsh 1998). Het, heterozygote; Hom, homozygote; WT, wild-type. *$p < 0.001$, **$p < 0.0001$

suggesting that the mildly-deleterious induced mutations are recessive, on average. Additionally, heterozygotes perform significantly better than would be predicted under additivity for total productivity and *w* (the mean trait value for heterozygotes is significantly higher than the mean of wild-types and mutant homozygotes), also indicating that the suites of mutations tested are recessive, on average (see Table 5.2). The average coefficient of dominance for *w* and total productivity was calculated as the ratio of the average heterozygous effect to the average homozygous effect ($\bar{h}_1$) and standard errors for this estimate were calculated using the delta method (Lynch & Walsh 1998). For both *w* and total productivity estimates of $\bar{h}_1$ are low, 0.079 (SE 0.14) and 0.12 (SE 0.12) respectively. These estimates are both significantly different from additivity (0.5), but not significantly different from zero. Estimates of $\bar{h}$ were also calculated using the regression approach ($\bar{h}_2 = \sigma_{het,hom}/\sigma^2_{hom}$ where σ and σ^2 are genetic covariances and variances respectively), standard errors were estimated by bootstrapping 10,000 times by line. Estimates of $\bar{h}_2$ are somewhat higher than $\bar{h}_1$: 0.49 (SE 0.105) for *w* and 0.24 (SE 0.101) for total productivity (see Figure 5.5 for the regression of heterozygous effect on homozygous effect for *w* and total productivity). The estimate of $\bar{h}_2 = 0.49$ for *w* is significantly different from zero but not from additivity, whereas the estimate of $\bar{h}_2 = 0.24$ for total productivity is significantly different from both zero and additivity.

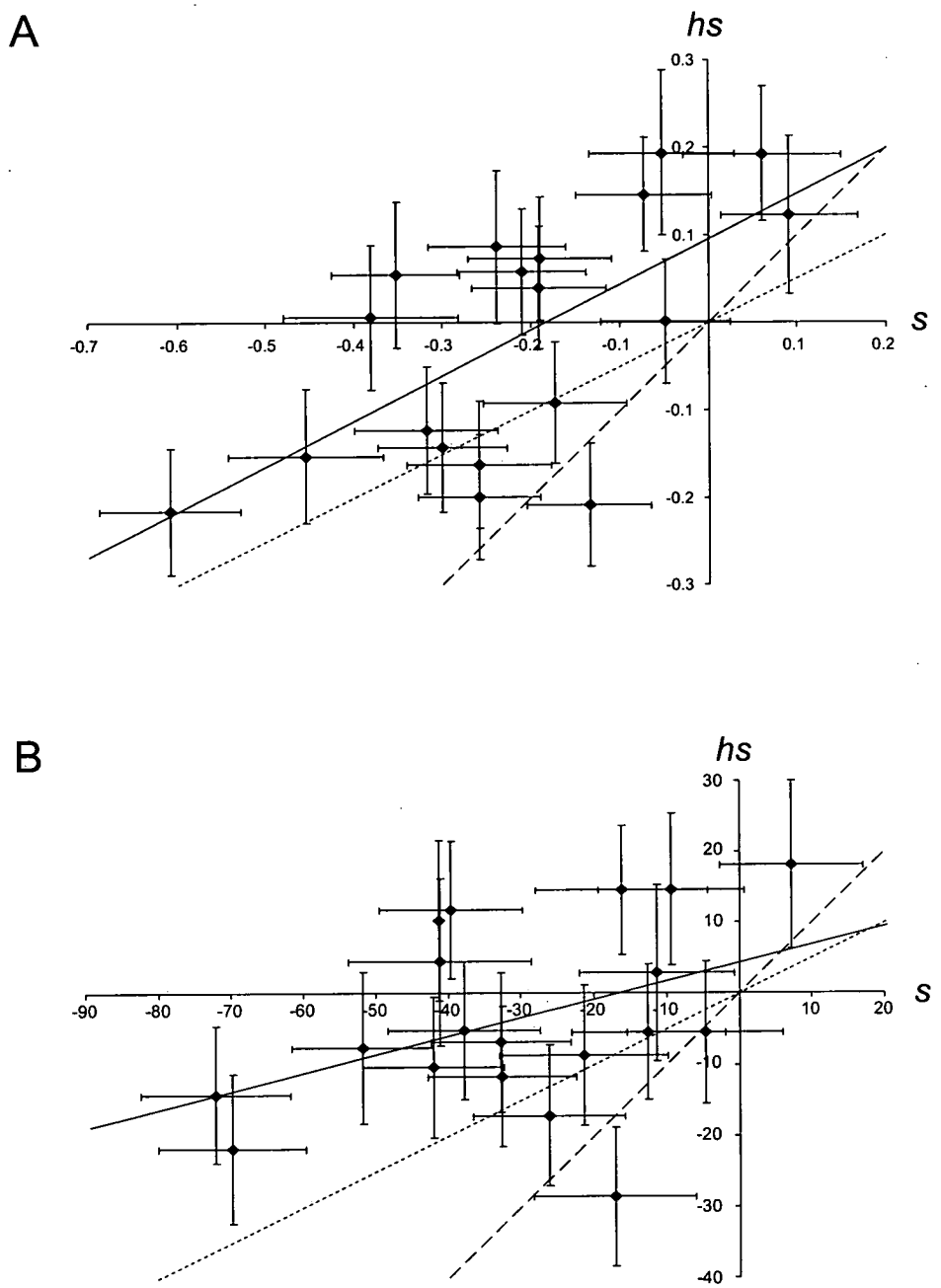


Figure 5.5: Plot of heterozygous genotypic effect (*hs*) against homozygous genotypic effect (*s*) of individual lines for relative fitness (*w*) (A) and total productivity (B). The solid line shows the linear regression of *hs* on *s*, the dashed line shows complete dominance ($\bar{h} = 1$) and the dotted line shows additivity ($\bar{h} = 0.5$).

The magnitude of the difference between heterozygotes and mutant homozygotes varies significantly among individual lines in two respects. Firstly, the interaction between refgt and zyg is significant (see Table 5.3), and secondly, the average dominance coefficient for individual lines varies across lines ranging from recessivity to complete dominance and possible overdominance (see Figures 5.6 and 5.7). To test whether the amount of variation between lines was significant, estimates of $\bar{h}_l$ were bootstrapped by line 10,000 times and the variance calculated among each of the bootstrapped datasets. The point estimate for the variance of $\bar{h}_l$ between lines was 1.89 for w (0.768 for total productvity) and the 95% confidence interval for this estimate (obtained by bootstrapping) was 0.35–3.7 (0.20–1.4 for total productivity), suggesting that the variance amongst lines was indeed significant for both traits. Interestingly, heterozygote means for individual lines appear to be evenly distributed around the wild-type mean (both for w and productivity). In fact three lines show significant evidence of heterosis for w, *i.e.* heterozygote means for these lines are significantly greater than the wild-type mean. These are lines E11 ($p = 0.0258$), E13 ($p = 0.0136$) and E25 ($p = 0.0413$) (p-values are not corrected for multiple tests).

There was a highly significant effect of family (*i.e.* the egg-lay plate from which the parental worm was derived) for both productivity and w (see Table 5.3). This is somewhat surprising, as it was expected that worms originating from different families would be synchronised, since they were all laid during the same six-hour period. However, as in previous studies, it was noted here that older worms tend to lay more advanced eggs, so eggs laid at the same time may not be at the same stage of development. This could give rise to variation amongst families in the time at which worms mature and begin to lay eggs. Since the calculation of w weights the first day of productivity most heavily, it is more sensitive to subtle variation in the timing of the start of egg production. It might therefore be expected that the family effect would be larger for w than for total productivity and this is found to be true (the Z-statistic associated with the family effect is larger for w than for total productivity).

5.3.3 Verification of Heterosis

Two lines (E11 and E13) were re-assayed for productivity to test the hypothesis that they were heterotic, using a modification of the standard productivity assay. The two replicates of N2 used in this experiment did not differ significantly for either

Table 5.3: ANOVA table for proc MIXED analysis of total productivity and relative fitness (w).

Trait	Effect	Variance (SE)	df_{num}	df_{den}	F	Z
Prod.	zyg		1	18	16.72 ***	
	refpar		1	204	0.01	
	refgt	144 (136)				1.06
	refgt × zyg	276 (129)				2.14 **
	assay	688 (592)				1.16
	counter	38.5 (23.8)				1.62
	family	283 (54.6)				5.19 ***
	residual	1350 (52.5)				25.77 ***
w	zyg		1	18	18.0 ***	
	refpar		1	204	1.62	
	refgt	0.0167 (0.00963)				1.74
	refgt × zyg	0.0136 (0.00676)				2.02 *
	assay	0.0 (-)				-
	counter	0.00180 (0.000920)				1.96 *
	family	0.0245 (0.00323)				7.61 ***
	residual	0.0390 (0.00152)				25.74 ***

Random effects were estimated by REML and significance was tested using Z scores. Significance of fixed effects was tested using F statistics. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

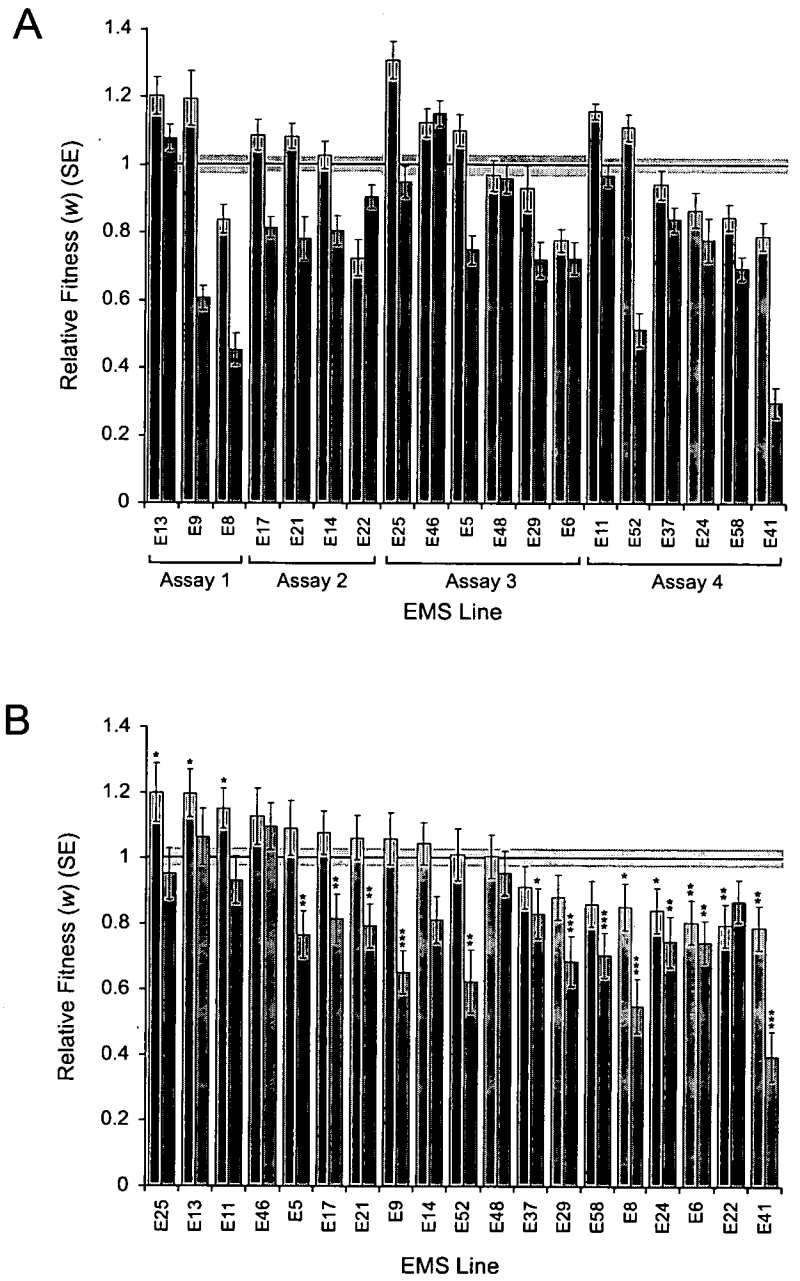


Figure 5.6: Estimates of heterozygote (lightly shaded bars) and homozygote (darkly shaded bars) relative fitness (w) (±SE) for individual lines. Both raw means (A) and least-square means (B) are shown. Raw means are ordered by assay and the wild-type mean is shown as a horizontal bar (±SE grey box), calculated for individual assays. Least-square means were calculated with the generalised linear mixed model and the wild-type mean is shown as a horizontal bar (±SE grey box) calculated across all assays. Asterisks above bars (in B) correspond to the significance of the difference between that bar and the wild-type. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

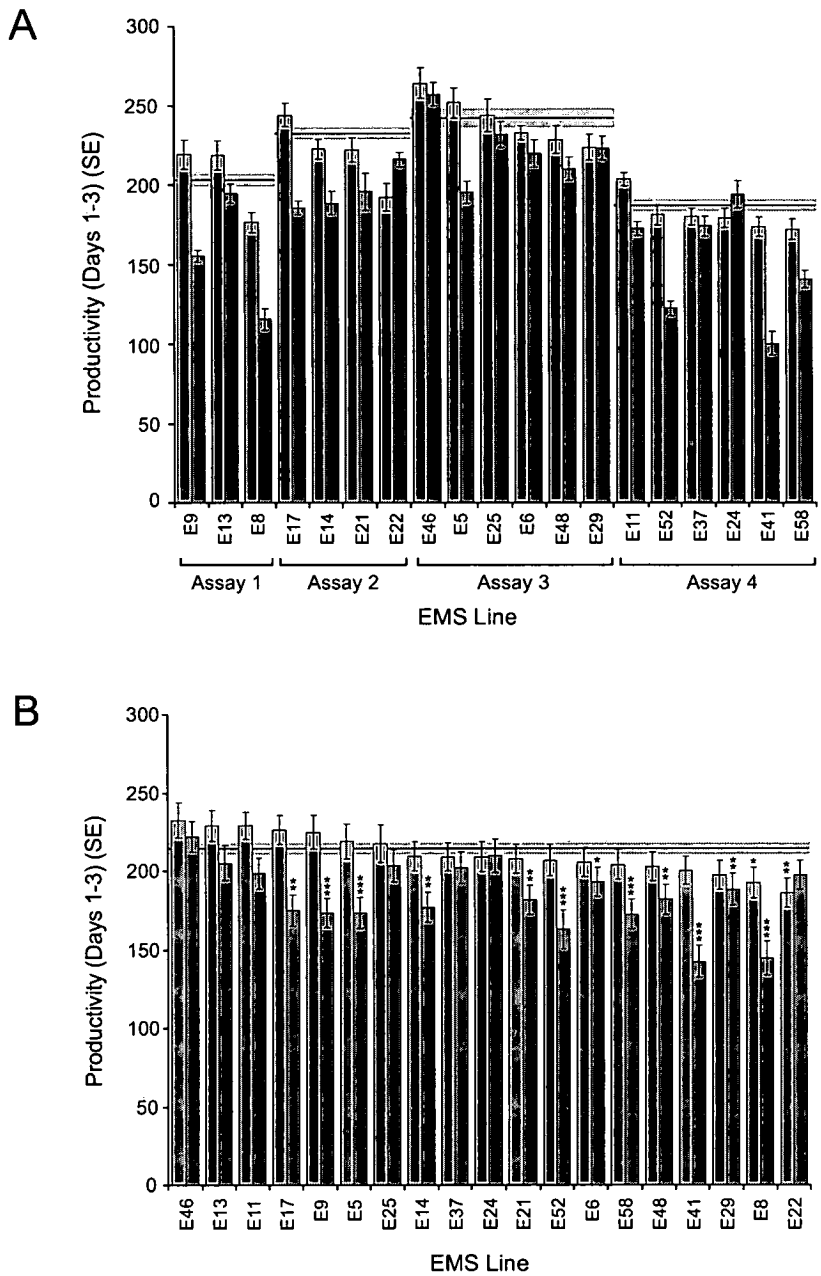


Figure 5.7: Estimates of heterozygote (lightly shaded bars) and homozygote (darkly shaded bars) total productivity (±SE) for individual lines. Both raw means (A) and least-square means (B) are shown. Raw means are ordered by assay and the wild-type mean is shown as a horizontal bar (±SE grey box), calculated for individual assays. Least-square means were calculated with the generalised linear mixed model and the wild-type mean is shown as a horizontal bar (±SE grey box) calculated across all assays. Asterisks above bars (in B) correspond to the significance of the difference between that bar and the wild-type. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

w ($p = 0.137$), or productivity ($p = 0.191$) and the random effect "N2-type" was non-significant for both traits. However, as for the main assay, family was a significant effect for both traits and again this is likely to be due to subtle variation in the age of parental worms from different egg-lay plates.

Heterozygotes performed better than the wild-type for both w and productivity confirming the findings above (this is significant for productivity ($p = 0.0399$) but not for w ($p = 0.114$)). Furthermore, in accordance with the main assays, heterozygotes performed significantly better than homozygotes for both w ($p = 0.0464$) and productivity ($p = 0.0343$) but there was no significant difference between the wild-type and homozygotes ($p = 0.480$ and 0.888 for w and total productivity respectively). On an individual line basis, the E11 heterozygote mean was significantly higher than N2 for productivity ($p = 0.0318$) but was marginally non-significantly higher for w ($p = 0.0557$) and although the E13 heterozygote also performed better than the wild-type for both w and total productivity the differences were both non-significant ($p = 0.418$ and $p = 0.0900$ respectively). This result is in slight contrast to the result from the main assay, since the difference between E11 heterozygotes and the wild-type was less significant than that between E13 heterozygotes and the wild-type. This difference could potentially be due to some assay effect, reflecting an element of genotype × environment interaction. Nonetheless, the results from the re-assay are in broad agreement with those from the main assays and suggest that E11 and E13 are indeed heterotic.

5.4 Discussion

The estimates of $\bar{h}_1 \simeq 0.1$ for both w and total productivity (calculated via the proportional difference in means approach) and $\bar{h}_2 \simeq 0.5$ and 0.2 for w and total productivity respectively (calculated via the regression approach) are broadly consistent with previous results from *Drosophila* and *C. elegans*, which suggest that mildly deleterious mutations are partially recessive, on average (Mukai *et al.* 1964, 1965, Mukai & Yamazaki 1968, Mukai 1970, Mukai *et al.* 1972, Ohnishi 1977*b*, Hughes 1995, Houle *et al.* 1997, García-Dorado & Caballero 2000). The estimates of $\bar{h}_1$ are consistent with complete recessivity ($\bar{h} = 0$) but not with additivity ($\bar{h} = 0.5$); whereas the estimates of $\bar{h}_2$ are consistent with partial recessivity (the estimate for w is consistent with additivity but not recessivity, whereas the estimate for total

Table 5.4: Means and effect estimates (SE) for total productivity and *w* from the verification of heterosis reassay.

Parameter	Total Productivity (SE)	*w* (SE)
WT Mean	196 (4.61)	0.979 (0.0467)
N2A	190 (5.86)	0.902 (0.0608)
N2B	202 (5.71)	1.06 (0.0598)
Het Mean	217 (4.00)	1.11 (0.0397)
E11	222 (5.00)	1.19 (0.0500)
E13	213 (4.91)	1.04 (0.0481)
Hom Mean	197 (3.81)	0.932 (0.0388)
E11	195 (4.69)	0.991 (0.0465)
E13	199 (4.92)	0.874 (0.0502)
Het – WT	21.3 (6.10)**	0.136 (0.0613)
E11	25.9 (6.80)**	0.208 (0.0685)*
E13	16.6 (6.74)*	0.0628 (0.0670)
Hom – WT	0.853 (5.56)	−0.0468 (0.0583)
E11	−1.20 (6.22)	0.0115 (0.0638)
E13	2.91 (6.36)	−0.105 (0.0663)
Hom – Het	−20.4 (5.52)**	−0.182 (0.0556)**
E11	−27.1 (6.86)**	−0.197 (0.0612)**
E13	−13.7 (6.95)	−0.168 (0.0618)*

Standard errors were calculated for the specific counter and assay effects included in this experiment. *$p < 0.1$; **$p < 0.05$. Het, heterozygote; Hom, homozygote; WT, wild-type.

productivity is consistent with neither additivity nor recessivity).

The results presented here therefore seem to depend somewhat on the method of analysis, and this was also found to be true of experiments performed by Mukai *et al.* (1965) and Ohnishi (1977*b*) (see Simmons & Crow 1977, García-Dorado & Caballero 2000). Mukai *et al.* (1965) formed coupling heterozygotes using non-isogenic chromosomes and estimated $\bar{h}$ using the proportional difference in means method ($\bar{h}$ = 0.09–0.13), whereas the same results analysed using the regression approach yields slightly lower estimates ($\bar{h}$ = 0.053–0.075, García-Dorado & Caballero 2000). Similarly, the results of (Ohnishi 1977*b*) vary depending on the method of calculation, an estimate of $\bar{h}$ = 0.42–0.48 was obtained using proportional difference in means, but a lower estimate of $\bar{h}$ = 0.12–0.15 was found when the data were analysed using the regression approach (García-Dorado & Caballero 2000). The estimates of $\bar{h}_1$ presented here therefore agree reasonably well with those of Mukai *et al.* (1965), using either method of calculation, but, peculiarly, the estimates of $\bar{h}_2$ align better with the $\bar{h}_1$ estimates of Ohnishi (1977*b*) and *vice versa*, although Ohnishi's estimates of $\bar{h}_1$ have been questioned. García-Dorado & Caballero (2000) suggested that there may have been a non-mutational decline in viability during the first half of the experiment (that affected both homozygotes and heterozygotes), and this would bias the estimate of $\bar{h}_1$ upwards, whilst it would not affect the estimate of $\bar{h}_2$. This is because the viability decline is not used when estimating $\bar{h}$ with the regression method.

Estimates of $\bar{h}_1$ (but not $\bar{h}_2$) are fairly consistent with previous estimates using EMS-induced mutations. Ohnishi (1977*b*), Mukai (1970) and R. G. Temin (unpublished results, Simmons & Crow 1977) all found evidence for low dominance. Ohnishi estimated $\bar{h}$ = 0.27, for high doses of EMS whereas Mukai's estimates were lower ($\bar{h} \simeq 0.03$) and R. G. Temin found that $\bar{h}$ varied from 0.06 to 0.24 for coupling heterozygotes with a non-isogenic homologue. However, the estimates of $\bar{h}_1$ calculated here are inconsistent with Mukai's estimates of $\bar{h}$ using repulsion heterozygotes (when calculated with either method, estimated dominance coefficients were consistently close to additive, $\bar{h}$ = 0.36–0.46, Mukai & Yamazaki 1968), although the estimates of $\bar{h}_2$ presented here are more consistent with Mukai's results

Interestingly, the estimates of $\bar{h}_1$ in this study differ substantially from those of Vassilieva *et al.* (2000) for similar traits in the same strain of *C. elegans*; their regression-based estimate for productivity was $\bar{h}$ = 0.64 as compared to the means-

based estimate of $\bar{h}$ = 0.12 reported here, although the error on both of these estimates is substantial (SE = 0.18 and 0.12 respectively). However, it is possible that this difference is due to the method of calculation. The estimate from the regression approach for w ($\bar{h}_2$ = 0.49) is comparable to that obtained by Vassilieva *et al.* for productivity ($\bar{h}$ = 0.64) and for "intrinsic rate of increase" ($\bar{h}$ = 0.55), although the estimate of $\bar{h}$ for productivity presented here is still somewhat lower ($\bar{h}_2$ = 0.24). Arguably, the proportional difference in means approach is a more appropriate measure of the average dominance coefficient, since it has been suggested that the regression approach is likely to be more susceptible to bias (Caballero *et al.* 1997). It would therefore be interesting to determine what the difference-in-means approach would yield if it were applied to the data of Vassilieva *et al.*, unfortunately this analysis has not been performed.

It is unlikely that the differences between the estimates of Vassilieva *et al.* and those here are due to this experiment preferentially using fitter lines. Although Vassilieva *et al.* used lines chosen almost at random, high-fitness lines are expected to bias estimates toward higher values of $\bar{h}$ (García-Dorado & Caballero 2000). Although the lines tested here carry EMS-induced mutations, whereas those of Vassilieva *et al.* carry spontaneous mutations, this seems unlikely to be the reason for the difference in $\bar{h}$. The average number and effects of mutations are similar between the EMS lines used here and the MA lines of Vassilieva *et al.* (2000). For productivity, it has been estimated that these EMS lines carry, on average, a low number of mutations with moderate effects (~1.5 detectable mutations per line, with an average effect of ~20%; Keightley *et al.* 2000), although there could be very many smaller effect mutations (Halligan *et al.* 2003), whilst Vassilieva *et al.* (2000) estimate that their 214-generation lines carry ~3.2 mutations (0.015 per generation × 214 generations) with an average effect of ~22%.

If the average dominance coefficient is low for newly arising mutations (as is suggested by the estimates of $\bar{h}_1$ in this study), then effects of newly arising mutations (which will be mostly in the heterozygous state) will be small. Low dominance of mutations is expected to lead to high inbreeding depression in outcrossing and partially self-fertilizing populations (Charlesworth & Charlesworth 1987). Since the level of inbreeding depression is a critical parameter in models determining the outcome of selection on alleles affecting selfing rates (Lloyd 1979), it also has consequences for the evolution of mating systems. A potential increase in the amount

of expected inbreeding depression also has implications for conservation biology. Furthermore, in combination with the conclusion that the distribution of deleterious mutation effects may be highly L-shaped and leptokurtic (see Davies *et al.* 1999, Keightley *et al.* 2000, Halligan *et al.* 2003 and Chapter 4), it seems likely that many newly arising deleterious mutations may have very small effects indeed. This could affect, for example, the speed of Muller's Ratchet (Charlesworth & Charlesworth 1997). It has also been suggested that very weak selection against newly arising mutations might change the predictions of deterministic models of mutations and the evolution of sex (Peters & Keightley 2000).

One interesting finding of this study is the apparent significant variation in the degree of dominance between individual lines. This is in agreement with another study, which also found substantial variation in dominance amongst lines, ranging from underdominant, through to recessive, to dominant (Fernández & López-Fanjul 1996). Furthermore, in at least two of the 19 lines tested here, there is evidence for heterosis, and this result appears to be repeatable. It is possible that this heterosis could be caused by some of the induced mutations having advantageous effects on the traits measured. If deleterious mutations are very nearly recessive on average, but advantageous mutations are dominant (to the same extent), then heterosis may be observed if only a small fraction of mutations are advantageous (Fry 1994). However, it seems unlikely that even a small fraction of induced mutations would be advantageous given that the strain exposed to EMS has been adapted to the lab environment and that the traits were measured were related to fitness. It is also possible that the observed heterosis may also be caused by epistatic interactions between loci, such that the advantageous effect of a dominant allele at one locus is suppressed in homozygotes by a recessive allele at another linked locus. Alternatively, it could be due to overdominance amongst a subset of the mutations induced. If the distribution of mutational effects is skewed, then the estimated dominance for any one line may be highly dependent on the effects of only a small number of mutations. In Chapter 4 it was estimated that the fitness difference between N2 and a random selection of 8 homozygous mutant lines was primarily due to a small number (~1.5–2.5) of major effect mutations. Therefore, even if the vast majority of mutations were partially recessive it might still be possible to observe the effects of a small number of overdominant mutations in some lines in this experiment. Interestingly, overdominance for spontaneous and induced mutations has been previously reported (Wallace

1957, Wallace & Dobzhansky 1962, Wallace 1963, Mukai *et al.* 1966, Maruyama & Crow 1975), although only on certain genetic backgrounds, and it is possible that the controls used in some experiments may have accumulated mutations themselves, which would give artifactual evidence of overdominance (if the deleterious mutations present were partially recessive). Evidence for overdominance has not been found since these early studies, although many experiments that have been performed were not designed to detect overdominance. If overdominance is only a property of a small fraction of mutations, their effects may well be swamped by other partially dominant and recessive mutations, when averages are calculated. It is therefore desirable to test the effects of a *single locus* if overdominance is suspected, such an experiment has been performed in maize, where individual mutations were tested against a highly isogenic background (Schuler 1954). Schuler analysed a total of 12 mutants and found that five appeared to be heterotic for all (or almost all) traits that were tested. However, it was shown that for two of the mutants tested the background was not totally isogenic, so it was not possible to conclude that the the observed effects were due to the single locus being tested. A further study by Schuler & Sprague (1956) also provided inconclusive evidence for overdominance at a single locus as it could not be firmly established that the background was totally isogenic (Schuler & Sprague 1956).

The possibility of overdominance makes inferences about the expected level of inbreeding depression more complicated, since overdominance can also lead to inbreeding depression (Charlesworth & Charlesworth 1987). Furthermore, if a reasonably large fraction of mutations are overdominant, this could affect the amount of genetic variation expected at equilibrium. Overdominant mutations would persist over greater periods of time than partially recessive mutations and may therefore make a significant contribution to genetic variation. This potentially provides some support for the "balanced" theory over the "classical" theory in the ongoing debate regarding the maintenance of genetic variation.

In summary, the results presented here are consistent with two general patterns; (i) that mildly deleterious mutations are partially recessive on average, but (ii) that there is substantial variation in the degree of dominance. Both of these patterns may have substantial implications for evolutionary processes that are driven by deleterious mutations. The present result also reinforces the conclusion that newly arising mutations tend to be weak in their effects, but that there is high variability among

6 Discussion and Conclusions

6.1 Summary

Mutations are the source of genetic variability, which is necessary for evolution to proceed. Understanding the properties of new mutations can therefore help us to understand evolution. Indeed, parameters associated with new mutations, particularly deleterious mutations, are thought to be crucial to many aspects of evolutionary theory. In this thesis, various parameters associated with new mutations have been estimated, and the effect of mutations in different regions of non-coding DNA has been assessed.

In Chapter 3, the level of functional constraints in the non-coding DNA of *Drosophila* was estimated by comparing the non-coding DNA of two related *Drosophila* species. Since the majority of the eukaryotic genome is non-coding, quantifying the level of constraint is vital for determining the genome wide mutation load due to deleterious mutations. The aim of the work presented in Chapters 4 and 5 was to estimate directly parameters associated with new deleterious mutations, using lines containing suites of EMS-induced mutations. In Chapter 4 the average number of mutations per line was estimated using inbred sublines. This has direct relevance for determining the shape of the distribution of mutational effects. The aim of the work in Chapter 5 was to estimate the average dominance coefficient for a selection of lines containing mildly deleterious induced mutations.

The DNA sequences of two related *Drosophila* species (*D. simulans* and *D. melanogaster*) were used in Chapter 3 to estimate functional constraint in intronic and intergenic DNA. Functional constraint (C) was defined as the fraction of missing substitutions (Kimura 1983), estimated by comparing the expected numbers of substitutions, obtained from putatively neutral sequences, to the observed number of substitutions in a stretch of non-coding DNA. The results obtained by averaging across many loci suggest that there is substantial functional constraint in intergenic sequences within at least 500bp upstream and downstream of a coding sequence. This result

corresponds well with previous studies that have suggested that intergenic sequences show quite high levels of conservation (Britten 1986, Li & Graur 1991, Oeltjen *et al.* 1997, Jareborg *et al.* 1999, Shabalina & Kondrashov 1999, Bergman & Kreitman 2001, Waterston *et al.* 2002, Thomas *et al.* 2003, Dermitzakis *et al.* 2003). However, the results for intronic sequences show that constraint is negative, on average, if four-fold sites are assumed to be neutral. In other words, four-fold sites appear to be evolving slower and are more constrained than intronic sites. This conflicts with previous studies (Jareborg *et al.* 1999, Shabalina & Kondrashov 1999, Bergman & Kreitman 2001), and does not support the view that most introns contain functionally important sequences. This difference could reflect real differences between species or could be an artefact of the different methods used to estimate the degree of conservation. These results also suggest that four-fold degenerate synonymous sites could be evolving under reasonably high levels of constraint, since they appear to be evolving slower than intronic sites (when splice control regions are removed). This could be due to selection associated with codon-usage bias.

Despite the lack of constraint observed in intronic sequences, the results from Chapter 3 suggest that there are substantial amounts of DNA under functional constraint outside coding regions. Rough calculations suggest that the amount of constrained intergenic DNA may even be on the same order as the amount of constrained coding sequence. There are, therefore, many more sites at which mutations would be deleterious than those in coding sequences alone. Estimates of the deleterious mutation rate based purely on analysis of coding sequences could therefore be substantial underestimates (*e.g.* Eyre-Walker & Keightley 1999). Additionally, given that there is little difference in the numbers of genes between highly divergent eukaryotes, it is possible that non-coding DNA makes a substantial contribution to the observed differences in complexity.

In Chapters 4 and 5, direct estimates of parameters associated with induced mutations were obtained. In both chapters EMS-induced mutant lines, generated by Davies *et al.* (1999), were studied. These EMS-induced lines provide a powerful system for studying the properties of new mutations for a number of reasons. Firstly, it is possible to calculate approximately the number of mutations induced for a given dose of EMS. Davies *et al.* conservatively estimated that they had induced ~45 point mutations per line that would be deleterious under natural conditions. Secondly, it has been argued that EMS mutations are representative of spontaneous mutations (see

Davies *et al.* 1999). Finally, the effects of many mutations can be studied without the need for long periods of mutation accumulation.

In Chapter 4, the number of mutations detectable on the basis of fitness assays was estimated using inbred sublines. Sublines were produced for a total of eight EMS-induced homozygous mutant lines, by crossing the lines to the wild-type (N2), and inbreeding a random selection of ten F1 offspring to homozygosity. The original lines, their corresponding sublines and the N2 line were then assayed for several fitness correlates, including relative fitness (w). Such inbred sublines provide a powerful method for determining the number and corresponding average effect of mutations present in a given line, because individual mutations should segregate amongst sublines. The results of the fitness assays were analysed using a modification of the Castle-Wright estimator and a maximum-likelihood (ML) method. The ML method allows for variation in mutational effects by fitting either one or two classes of mutational effect, and uses a Box-Cox power transformation of residual values to account for a skewed distribution of residuals. Both the Castle-Wright and the ML analyses suggest that most of the variation among sublines was due to a few (~1.5–2.5 on average) large-effect mutations. Under an ML model with two classes of effects, including a class with small effects fitted the data better than a single class model, although not significantly so. Nonetheless, given that it has been estimated that each line contains on average ~45 deleterious mutations, these results suggest that many induced mutations have very small (but still deleterious) effects on fitness, and that the distribution of mutation effects is highly leptokurtic and L-shaped. This confirms previous findings using the same EMS-induced mutant lines (Davies *et al.* 1999, Keightley *et al.* 2000). These very small effect mutations are undetectable using standard fitness assays, and would be overlooked if the variability in mutation effects were ignored. This has important implications for theories concerning the evolution of sex, since it is thought that sexual reproduction could be maintained purely by spontaneous deleterious mutations, if they occur at a rate greater than one per diploid genome per generation. Although current estimates obtained from mutation accumulation experiments fall well below this value (see reviews in Drake *et al.* 1998, Keightley & Eyre-Walker 1999, Lynch *et al.* 1999), the preponderance of mildly deleterious mutations could mean that the true deleterious mutation rate is above one for many species.

In Chapter 5 the average dominance coefficient ($\bar{h}$) for mildly deleterious induced

mutations was estimated using a selection of 19 EMS-induced lines that performed well in previous assays. The viability, productivity and relative fitness (w) was estimated for heterozygotes, homozygotes and the wild-type (N2). These results were then used to calculate the average dominance coefficient using two different methods. Dominance was estimated as the average proportional reduction in the heterozygote trait value relative to that in homozygotes, and also as the regression coefficient of the heterozygous trait values on the homozygous trait values for individual lines. The mutational effect on viability was found to be small, and estimates of viability were prone to error, due to difficulty in counting eggs, and so only productivity and w were used to estimate dominance.

It was found that mutations were partially recessive on average (*i.e.* $0 < \bar{h} < 0.5$), although the level of recessivity depended on the method used to calculate $\bar{h}$. The proportional difference in means approach gave an estimate of $\bar{h} \simeq 0.1$ (for both w and productivity), but the regression approach resulted in somewhat higher estimates ($\bar{h} \simeq 0.5$ and 0.2 for w and productivity respectively). This conclusion is in broad agreement with previous studies (Mukai *et al.* 1964, 1965, Mukai & Yamazaki 1968, Mukai 1970, Mukai *et al.* 1972, Ohnishi 1977*b*, Hughes 1995, Houle *et al.* 1997, García-Dorado & Caballero 2000). Combined with the conclusion that many induced mutations have very small but deleterious effects on fitness in the homozygous state, the estimates of $\bar{h}$ imply that new mutations could potentially have very small effects indeed. Secondly, there appears to be significant variation in the dominance coefficient for the mutations in individual lines; two lines even showed repeatable evidence of heterosis. These two conclusions have implications for the amount of genetic variability expected at mutation-selection balance. Low dominance of new mutations leads to large amounts of genetic variability at equilibrium, but the possibility that even a few mutations may be overdominant could increase the expected amount of genetic variability even further, because overdominant mutations would accumulate in stable equilibrium at intermediate frequencies. Even a low frequency of overdominant *mutations* becomes, at equilibrium, a high frequency of overdominant *segregating alleles* (Lewontin 1974).

6.2 Future Directions

The conclusions from Chapter 3 were limited by the lack of available non-coding DNA sequence data for *D. simulans* (particularly intronic and 3' intergenic sequences).

Furthermore, the sequences that were used were not randomly selected. Completion of the *D. simulans* genome sequence will therefore provide an opportunity to clarify whether or not the patterns observed in Chapter 3 are genuine, by analysing a completely random sample of many more loci than were used here. Furthermore, completion of the *D. yakuba* genome sequence will make it possible to apply a parsimony approach and infer the direction of individual mutations, facilitating lineage specific estimates of substitution rates. This would be useful for verifying whether or not selection has been relaxed along the lineage leading to *D. melanogaster* since the split with *D. simulans*, a theory which is supported by other recent data (Akashi 1995, 1996, McVean & Vieira 2001, Halligan *et al.* 2004).

With the extra data afforded by the genomic sequences, further comparisons could potentially be made. For example, it should be possible to test whether UTR sequences (sequences that are transcribed and form part of the mRNA but are untranslated) explain the high levels of constraint observed close to coding sequences. It should also be feasible to compare the constraint observed according to the position of introns within a coding sequence (*i.e.* to ask whether introns close to the start of a coding sequence show the highest constraint, as has been observed in mouse / rat comparisons; Keightley & Gaffney 2003). Extra sequence data should also make it possible to study changes in the level of functional constraint at greater distances from coding sequences. Both of these genome sequences are near completion, and the data will become publicly available shortly [1].

The conclusion that many mutations have very small, undetectable effects on fitness (Chapter 4) is based mainly on the assumption that there are approximately 45 deleterious mutations per line as a result of EMS-mutagenesis. However, this assumption is based purely on the expected rate of EMS-mutagenesis, which itself is calibrated from experiments to measure forward mutation rates and suppressor-induced reversion mutation rates. The estimate of 45 deleterious mutations is therefore indirect and potentially inaccurate. For that reason, it would be desirable to determine directly and independently how many homozygous mutations are present per line. One possible approach would be to sequence random blocks of genomic DNA from all EMS-induced mutant lines, to identify individual point and indel mutations, an approach that has previously been used to estimate the mitochondrial mutation rate

[1] Information about the current state of the *D. simulans* and *D. yakuba* genome projects can be found at `http://www.dpgp.org/sim_yak/`

in *Caenorhabditis elegans* (Denver *et al.* 2000). However, this approach would be very expensive. An alternative is to use DHPLC (Denaturing High Pressure Liquid Chromatography) to detect differences between two homologous sections of DNA, for example, between a PCR product from a wild-type line, and an EMS-induced mutant line. Single base pair differences between these PCR products can be detected by mixing *single-stranded* DNA from both products (by melting the double-stranded DNA molecules and allowing them to re-anneal). If the original PCR products do differ by a single base pair, then after re-annealing, a certain fraction of DNA molecules (50%) will contain a mismatch. These mismatch molecules would move at a different speed along the chromatography column to non-mismatch molecules and this difference can be easily detected. If applied to many homologous sections of DNA, the fraction of sections containing a mismatch could be used to obtain a direct estimate of the number of mutations induced at the DNA level. This approach is currently being applied by Prof. Peter Keightley and coworkers to estimate the genomic mutation rate in spontaneous mutation accumulation lines of *Drosophila melanogaster.*

With a direct estimate of the average number of mutations per EMS-induced mutant line, it would be possible to use the two-class ML model described in Chapter 4 to obtain a concrete estimate of the number and effect size of the class of small effect mutations. Alternatively, with a direct estimate of the average number of mutations per p-line, it would be possible to use a previously developed maximum likelihood approach (Davies *et al.* 1999, Keightley *et al.* 2000) to estimate the parameters of the best fitting gamma distribution of mutation effects. This estimate would also have relevance for the results from Chapter 5, since it would be interesting to determine whether the heterosis observed in a few lines was due to the effects of a few or many individual mutations.

One other possibility for future research, identified in Chapter 5, would be to look for evidence of overdominance for a small fraction of spontaneous (or induced) mutations. In Chapter 5, three of 19 of the original lines assayed showed significant evidence of heterosis (not correcting for multiple tests), and this proved to be repeatable for two of the lines. It is possible that this could be due to overdominance of one or several of the induced mutations in those lines. Whilst the average dominance coefficient for new mutations is probably partially recessive, this result raises the interesting possibly that a significant fraction of mutations could be overdominant.

However, very few experiments have been designed such that overdominance of a small fraction of mutations could bė detected. Instead most previous studies have simply estimated the average dominance coefficient, where the partially recessive effects of many mutations may swamp overdominant effects of a few.

Bibliography

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. *et al.* (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics* **139**, 1067–1076.

Akashi, H. (1996). Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* **144**, 1297–1307.

Anderson, P. (1995). *Caenorhabditis elegans: Modern Biological Analysis of an Organism*. Epstein, H. F. and Shakes, D. C., eds. Academic Press, London.

Bachtrog, D., Agis, M., Imhof, M. & Schlötterer, C. (2000). Microsatellite variability differs between dinucleotide repeat motifs—evidence from *Drosophila melanogaster*. *Mol. Biol. Evol* **17**, 1277–1285.

Bataillon, T. (2000). Estimation of spontaneous genome-wide mutation rate parameters: whither beneficial mutations? *Heredity* **84**, 497–501.

Bateman, A. J. (1959). The viability of near-normal irradiated chromosomes. *Int. J. Radiat. Biol.* **1**, 170–180.

Begun, D. J. (2001). The frequency distribution of nucleotide-variation in *Drosophila simulans*. *Mol. Biol. Evol.* **18**, 1343–1352.

Bejsovec, A. & Anderson, P. (1988). Myosin heavy-chain mutations that disrupt *Caenorhabditis elegans* thick filament assembly. *Genes Dev.* **2**, 1307–1317.

Bell, G. (1982). *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*. University of California Press, Berkley.

Bell, G. (1988). Recombination and the immortality of the germ line. *J. Evol. Biol.* **1**, 67–82.

Bergman, C. M. & Kreitman, M. (2001). Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res.* **11**, 1335–1345.

Blackman, R. K. & Meselson, M. (1986). Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of the *Drosophila* hsp82 gene. *J. Mol. Evol.* **188**, 499–515.

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *J. R. Stat. Soc. Ser. B* **26**, 211–252.

Britten, R. J. (1986). Rates of DNA-sequence evolution differ between taxonomic groups. *Science* **231**, 1393–1398.

Bulmer, M. G. (1989). Maintenance of genetic variability by mutation-selection balance: a child's guide through the jungle. *Genome* **31**, 761–767.

Butcher, D. (1995). Muller's ratchet, epistasis and mutation effects. *Genetics* **141**, 431–437.

Caballero, A. & Keightley, P. D. (1994). A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* **138**, 883–900.

Caballero, A., Keightley, P. D. & Turelli, M. (1997). Average dominance for polygenes: drawbacks of regression estimates. *Genetics* **147**, 1487–1490.

Case, T. J. & Bender, E. A. (1981). Is recombination advantageous in fluctuating and spatially heterogeneous environments? *J. Theor. Biol.* **90**, 181–190.

Castle, W. (1921). An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. *Proc. Natl. Acad. Sci. USA* **81**, 6904–6907.

Cavener, D. R. (1992). Transgenic animal studies on the evolution of genetic regulatory circuitries. *BioEssays* **14**, 237–244.

Charlesworth, B. (1987). *Sexual Selection: Testing the Alternatives*, chap. The heritability of fitness, pp. 21–40. Bradbury, J. W. and Andersson, M. B., eds. Wiley, Chichester, UK.

Charlesworth, B. (1990). Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**, 199–221.

Charlesworth, B. (1994). *Evolution in Age-Structured Populations*. Cambridge University Press, Cambridge, CB2 1RP.

Charlesworth, B. & Charlesworth, D. (1997). Rapid fixation of deleterious alleles can be caused by Muller's ratchet. *Genet. Res.* **70**, 63–73.

Charlesworth, B. & Charlesworth, D. (1998). Some evolutionary consequences of deleterious mutations. *Genetica* **103**, 3–19.

Charlesworth, B. & Charlesworth, D. (1999). The genetic basis of inbreeding depression. *Genet. Res.* **74**, 329–340.

Charlesworth, B. & Hughes, K. A. (1996). Age-specific inbreeding depression and components of genetic variance in relation to the evolution of senescence. *Proc. Natl. Acad. Sci. USA* **93**, 6140–6145.

Charlesworth, B. & Hughes, K. A. (1999). *Evolutionary Genetics: from Molecules to Morphology, Vol. 1*, chap. The maintenance of genetic variation in life-history traits, pp. 369–392. Singh, R. S. and Krimbas, C. B., eds. Cambridge University Press.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1991). Multilocus models of inbreeding depression with synergistic selection and partial self-fertilization. *Genet. Res.* **57**, 177–194.

Charlesworth, D. & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.* **18**, 237–268.

Charlesworth, D., Morgan, M. T. & Charlesworth, B. (1990). Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* **44**, 1469–1489.

Charlesworth, D., Morgan, M. T. & Charlesworth, B. (1993). Mutation accumulation in finite outbreeding and inbreeding populations. *Genet. Res.* **61**, 39–56.

Chasnov, J. R. (2000). Mutation-selection balance, dominance and the maintenance of sex. *Genetics* **156**, 1419–1425.

Chen, F. C. & Li, W. H. (2001). Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68**, 444–456.

Cherry, J. L. (2002). Deleterious mutation and the evolution of eusociality. *Evolution* **56**, 2359–2367.

Clark, A. G. (2001). The search for meaning in noncoding DNA. *Genome Res.* **11**, 1319–1320.

Clark, A. G., Wang, L. & Hulleberg, T. (1995). Spontaneous mutation rate of modifiers of metabolism in *Drosophila*. *Genetics* **139**, 767–779.

Cockerham, C. C. (1986). Modifications in estimating the number of genes for a quantitative character. *Genetics* **114**, 659–664.

Crow, J. F. (1958). Some possibilities for measuring selection intensities in man. *Human Biology* **30**, 1–13.

Crow, J. F. (1997). The high spontaneous mutation rate: is it a health risk? *Proc. Natl. Acad. Sci. USA* **94**, 8380–8386.

Crow, J. F. & Kimura, M. (1965). Evolution in sexual and asexual populations. *Am. Nat.* **99**, 439–450.

Davies, E. K., Peters, A. D. & Keightley, P. D. (1999). High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* **285**, 1748–1751.

den˜ Dunnen, J. T., Vanneck, J. W., Cremers, F. P. M., Lubsen, N. H. & Schoenmakers, J. G. G. (1989). Nucleotide-sequence of the rat gamma-crystallin gene region and comparison with an orthologous human region. *Gene* **78**, 201–213.

Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L. & Thomas, W. K. (2000). High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**, 2342–2344.

Dermitzakis, E. T., Reymond, A., Lyle, R., Scamuffa, N., Ucla, C., Deutsch, S., Stevenson, B. J., Flegel, V., Bucher, P., Jongeneel, C. V. *et al.* (2002). Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* **420**, 578–582.

Dermitzakis, E. T., Reymond, A., Scamuffa, N., Ucla, C., Kirkness, E., Rossier, C. & Antonarakis, S. E. (2003). Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science* **302**, 1033–1035.

Dickinson, W. J. (1991). The evolution of regulatory genes and patterns in *Drosophila*. *Evol. Biol.* **25**, 127–173.

Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* **148**, 1667–1686.

Duret, L. & Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila* and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **96**, 4482–4487.

Duret, L., Dorkeld, F. & Gautier, C. (1993). Strong conservation of noncoding sequences during vertebrates evolution—potential involvement in posttranscriptional regulation of gene-expression. *Nucleic Acids Res.* **21**, 2315–2322.

Eanes, W. F., Hey, J. & Houle, D. (1985). Homozygous and hemizygous viability variation on the X chromosome of *Drosophila melanogaster*. *Genetics* **111**, 831–844.

Elena, S. F. & Lenski, R. E. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398.

Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A. & Lenski, R. E. (1998). Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Genetica* **103**, 349–358.

Epp, T. A., Dixon, I. M. C., Wang, H. Y., Sole, M. J. & Liew, C. C. (1993). Structural organization of the human cardiac alpha-myosin heavy-chain gene (myh6). *Genomics* **18**, 505–509.

Eyre-Walker, A. & Bulmer, M. (1995). Synonymous substitution rates in enterobacteria. *Genetics* **140**, 1407–1412.

Eyre-Walker, A. & Keightley, P. D. (1999). High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347.

Eyre-Walker, A., Keightley, P. D., Smith, N. G. & Gaffney, D. (2002). Quantifying the slightly deleterious mutation model of molecular evolution. *Mol. Biol. Evol.* **19**, 2142–2149.

Fernández, J. & López-Fanjul, C. (1996). Spontaneous mutational variances and covariances for fitness-related traits in *Drosophila melanogaster*. *Genetics* **143**, 829–837.

Fisher, J. A. & Maniatis, T. (1986). Regulatory elements involved in *Drosophila Adh* gene expression are conserved in divergent species and seperate elements mediate expression in different tissues. *EMBO J.* **5**, 1275–1289.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford Univeristy Press, Oxford.

Fry, J. D. (1994). How common are overdominant mutations? *Genetics* **167**, 1031–1032.

Fry, J. D. (2001). Rapid mutational declines of viability in *Drosophila*. *Genet. Res.* **77**, 53–60.

Fry, J. D., Keightley, P. D., Heinsohn, S. L. & Nuzhdin, S. V. (1999). New estimates of the rates and effects of mildly deleterious mutation in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **96**, 574–579.

García-Dorado, A. (1997). The Rate and Effects Distribution of Viability Mutation in *Drosophila*: Minimum Distance Estimation. *Evolution* **51**, 1130–1139.

García-Dorado, A. & Caballero, A. (2000). On the average coefficient of dominance of deleterious spontaneous mutations. *Genetics* **155**, 1991–2001.

García-Dorado, A., Lopez-Fanjul, C. & Caballero, A. (1999). Properties of spontaneous mutations affecting quantitative traits. *Genet. Res.* **74**, 341–350.

Gilbert, D. G. (2002). euGenes: a eukaryote genome information system. *Nucleic Acids Res.* **30**, 145–148.

Glazko, G. V., Koonin, E. V., Rogozin, I. B. & Shabalina, S. A. (2003). A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. *Trends Genet.* **19**, 119–124.

Goldstein, D. B. (1992). Heterozygote advantage and the evolution of a dominant diploid phase. *Genetics* **132**, 1195–1198.

Gottgens, B., Barton, L. M., Gilbert, J. G., Bench, A. J., Sanchez, M. J., Bahn, S., Mistry, S., Grafham, D., McMurray, A., Vaudin, M. *et al.* (2000). Analysis of vertebrate *SCL* loci identifies conserved enhancers. *Nat. Biotechnol.* **188**, 181–186.

Gu, X. & Li, W. H. (1995). The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. *J. Mol. Evol.* **40**, 464–473.

Halligan, D. L., Peters, A. D. & Keightley, P. D. (2003). Estimating numbers of EMS-induced mutations affecting life history traits in *Caenorhabditis elegans* in crosses between inbred sublines. *Genet. Res.* **19**, 191–205.

Halligan, D. L., Eyre-Walker, A., Andolfatto, P. & Keightley, P. D. (2004). Patterns of evolutionary constraint in intronic and intergenic DNA of *Drosophila*. *Genome Res.* **14**, 273–279.

Hamilton, W. D. (1966). The moulding of senescense by natural selection. *J. Theor. Biol.* **12**, 12–45.

Hardison, R. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet.* **16**, 369–372.

Hardison, R. & Miller, W. (1993). Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol. Biol. Evol.* **10**, 73–102.

Hartl, D. L. & Clark, A. G. (1997). *Principles of Population Genetics*, chap. 6, pp. 236–239. Sinauer Associates, Inc., Sunderland, MA, 3rd edn.

Hedrick, P. W. & Kalinowski, S. T. (2000). Inbreeding depression in conservation biology. *Annu. Rev. Ecol. Syst.* **31**, 139–162.

Hellmann, I., Zollner, S., Enard, W., Ebersberger, I., Nickel, B. & Paabo, S. (2003). Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res.* **13**, 831–837.

Hodgkin, J. (1985). Novel nematode amber suppressors. *Genetics* **111**, 287–310.

Hodgkin, J. (1988). *The Nematode Caenorhabditis elegans*, chap. Sexual dimorphism and sex determination, pp. 243–279. Wood W. B., ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Houle, D., Hughes, K. A., Assimacopoulos, S. & Charlesworth, B. (1997). The effects of spontaneous mutation on quantitative traits. II. Dominance of mutations with effects on life-history traits. *Genet. Res.* **70**, 27–34.

Hughes, K. A. (1995). The inbreeding decline and average dominance of genes affecting male life-history characters in *Drosophila melanogaster*. *Genet. Res.* **65**, 41–52.

Jan, E., Yoon, J. W., Walterhouse, D., Iannaccone, P. & Goodwin, E. B. (1997). Conservation of the *C. elegans tra-2* 3'UTR translational control. *Embo J.* **16**, 6301–6313.

Jareborg, N., Birney, E. & Durbin, R. (1999). Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Res.* **9**, 815–824.

Jenkins, C. D. & Kirkpatrick, M. (1995). Deleterious mutation and the evolution of genetic life cycles. *Evolution* **49**, 512–520.

Johnston, M. O. & Schoen, D. J. (1995). Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. *Science* **267**, 226–229.

Judson, O. P. & Normark, B. B. (1996). Ancient asexual scandals. *Trends Ecol. Evol.* **11**, 41–46.

Kawecki, T. J. (1997). Sympatric speciation via habitat specialization driven by deleterious mutations. *Evolution* **51**, 1751–1763.

Kawecki, T. J., Barton, N. H. & Fry, J. D. (1997). Mutational collapse of fitness in marginal habitats and the evolution of ecological specialisation. *J. Evol. Biol.* **10**, 407–430.

Keightley, P. D. (1994). The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**, 1315–1322.

Keightley, P. D. (1996). Nature of deleterious mutation load in *Drosophila*. *Genetics* **144**, 1993–1999.

Keightley, P. D. (1998). Inference of genome-wide mutation rates and distributions of mutation effects for fitness traits: a simulation study. *Genetics* **150**, 1283–1293.

Keightley, P. D. & Bataillon, T. M. (2000). Multigeneration maximum-likelihood analysis applied to mutation-accumulation experiments in *Caenorhabditis elegans*. *Genetics* **154**, 1193–1201.

Keightley, P. D. & Caballero, A. (1997). Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **94**, 3823–3827.

Keightley, P. D. & Eyre-Walker, A. (1999). Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* **153**, 515–523.

Keightley, P. D. & Eyre-Walker, A. (2000). Deleterious mutations and the evolution of sex. *Science* **290**, 331–333.

Keightley, P. D. & Gaffney, D. J. (2003). Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents. *Proc. Natl. Acad. Sci. USA* **23**, 13402–13406.

Keightley, P. D. & Johnson, T. (2004). MCALIGN: stochastic alignment of noncoding DNA sequences based on an evolutionary model of sequence evolution. *Genome Res. Submitted* .

Keightley, P. D., Davies, E. K., Peters, A. D. & Shaw, R. G. (2000). Properties of ethylmethane sulfonate-induced mutations affecting life-history traits in *Caenorhabditis elegans* and inferences about bivariate distributions of mutation effects. *Genetics* **156**, 143–154.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kimura, M. & Maruyama, T. (1966). The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351.

Kirby, D. A., Muse, S. V. & Stephan, W. (1995). Maintenance of pre-mRNA secondary structure by epistatic selection. *Proc. Natl. Acad. Sci. USA* **92**, 9047–9051.

Kirkwood, T. B. L. & Rose, M. R. (1991). Evolution of senescence: late survival sacrificed for reproduction. *Philos. T. Roy. Soc. B* **332**, 15–24.

Kondo, K., Makovec, B., Waterston, R. H. & Hodgkin, J. (1990). Genetic and molecular analysis of eight tRNA(Trp) amber suppressors in *Caenorhabditis elegans*. *J. Mol. Biol.* **215**, 7–19.

Kondrashov, A. S. (1982). Selection against harmful mutations in large sexual and asexual populations. *Genet. Res.* **40**, 325–332.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440.

Kondrashov, A. S. (1995). Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *J. Theor. Biol.* **175**, 583–594.

Kondrashov, A. S. (1998). Measuring spontaneous deleterious mutation process. *Genetica* **103**, 183–197.

Kondrashov, A. S. & Crow, J. F. (1991). Haploidy or diploidy: which is better? *Nature* **351**, 314–315.

Kondrashov, A. S. & Crow, J. F. (1993). A molecular approach to estimating the human deleterious mutation rate. *Hum. Mutat.* **2**, 229–234.

Kondrashov, A. S. & Shabalina, S. A. (2002). Classification of common conserved sequences in mammalian intergenic sequences. *Hum. Mol. Genet.* **11**, 669–674.

Koop, B. F. (1995). Human and rodent DNA sequence comparisons: a mosaic model of genomic evolution. *Trends Genet.* **11**, 367–371.

Koop, B. F. & Hood, L. (1994). Striking sequence similarity over almost 100 kilobases of human and mouse t-cell receptor DNA. *Nat. Genet.* **7**, 48–53.

Koop, B. F., Richards, J. E., Durfee, T. D., Bansberg, J., Wells, J., Gilliam, A. C., Chen, H. L., Clausell, A., Tucker, P. W. & Blattner, F. R. (1996). Analysis and comparison of the mouse and human immunoglobulin heavy chain J_H-$C\mu$-$C\delta$ locus. *Mol. Phylogen. Evol.* **5**, 33–49.

Lande, R. (1981). The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**, 541–553.

Lande, R. (1994). Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**, 1460–1469.

Lande, R. (1995). Mutation and conservation. *Conserv. Biol.* **9**, 782–791.

Lande, R. & Schemske, D. W. (1985). The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. *Evolution* **39**, 24–40.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

Leicht, B. G., V., M. S., Hanczyc, M. & G., C. A. (1995). Constraints on intron evolution in the gene encoding the myosin alkali light-chain in *Drosophila*. *Genetics* **139**, 299–308.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. Columbia University Press, 1st edn.

Li, W. H. (1997). *Molecular Evolution*. Sinauer Associates, Inc., Sunderland, MA, USA.

Li, W. H. & Graur, D. (1991). *Fundamentals of Molecular Evolution.* Sinauer Associates, Inc., Sunderland, MA.

Li, W. H. & Sadler, L. A. (1991). Low nucleotide diversity in man. *Genetics* **129**, 513–523.

Liew, C. C., Sole, M. J., Yamauchitakihara, K., Kellam, B., Anderson, D. H., Lin, L. & Liew, J. C. (1990). Complete sequence and organization of the human cardiac betamyosin heavy-chain gene. *Nucleic Acids Res.* **18**, 3647–3651.

Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D. (1996). *SAS System for Mixed Models.* SAS Institute Inc., Cary, NC.

Lively, C. M. & Lloyd, D. G. (1990). The cost of biparental sex under individual selection. *Am. Nat.* **135**, 489–500.

Lloyd, D. G. (1979). Some reproductive factors affecting the selection of self fertilization in plants. *Am. Nat.* **113**, 67–79.

Loison, A., FestaBianchet, M., Gaillard, J. M., Jorgenson, J. T. & Jullien, J. M. (1999). Age-specific survival in five populations of ungulates: evidence of senescence. *Ecology* **8**, 2539–2554.

Ludwig, M. Z. & Kreitman, M. (1995). Evolutionary dynamics of the enhancer region of *even-skipped* in *Drosophila. Mol. Biol. Evol.* **12**, 1002–1011.

Lyman, R. F., Lawrence, F., Nuzhdin, S. V. & Mackay, T. F. C. (1996). Effects of single P-element insertions on bristle number and viability in *Drosophila melanogaster. Genetics* **143**, 277–292.

Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Inc., MA.

Lynch, M., Buerger, R., Butcher, D. & Gabriel, W. (1993). The mutational meltdown in asexual populations. *J. Hered.* **84**, 339–344.

Lynch, M., Conery, J. & Bürger, R. (1995*a*). Mutation accumulation and the extinction of small populations. *Am. Nat.* **146**, 489–518.

Lynch, M., Conery, J. & Bürger, R. (1995*b*). Mutational meltdowns in sexual populations. *Evolution* **49**, 1067–1080.

Lynch, M., Blanchard, J., Houle, D., Kibota, T., Schultz, S., Vassilieva, L. & Willis, J. (1999). Perspective: spontaneous deleterious mutation. *Evolution* **53**, 645–663.

MacKay, T. F. C., Lyman, R. F. & Jackson, M. S. (1992). Effects of P-element insertions on quantitative traits in *Drosophila melanogaster. Genetics* **130**, 315–332.

Majewski, J. & Ott, J. (2002). Distribution and characterization of regulatory elements in the human genome. *Genome Res.* **12**, 1827–1836.

Maruyama, T. & Crow, J. F. (1975). Heterozygous effects of x-ray-induced mutations on viability of *Drosophila melanogaster*. *Mutat. Res.* **27**, 241–248.

Maynard-Smith, J. (1971). *Group Selection*, chap. The origin and maintenance of sex, pp. 163–175. Williams, G. C., ed. Aldine Atherton, Chicago.

Maynard-Smith, J. (1978). *The Evolution of Sex*. Cambridge University Press, Cambridge, 1st edn.

McClure, M. A., Vasi, T. K. & Fitch, W. M. (1994). Comparative analysis of multiple protein-sequence alignment methods. *Mol. Biol. Evol.* **11**, 571–592.

McVean, G. A. T. & Vieira, J. (2001). Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. *Genetics* **157**, 245–257.

Medawar, P. B. (1946). Old age and natural death. *Mod. Q.* **2**, 30–49.

Medawar, P. B. (1952). *An Unsolved Problem in Biology*. H. K. Lewis, London.

Miller, R. D., Duan, S., Lovins, E. G., Kloss, E. F. & Kwok, P. Y. (2003). Efficient high-throughput resequencing of genomic DNA. *Genome Res.* **13**, 717–720.

Moore, W. S. & Hines, W. G. S. (1981). Sex in random-environments. *J. Theor. Biol.* **92**, 301–316.

Mukai, T. (1964). The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**, 1–19.

Mukai, T. (1969*a*). The genetic structure of natural populations of *Drosophila melanogaster*. VI. Further studies on the optimum heterozygosity hypothesis. *Genetics* **61**, 479–495.

Mukai, T. (1969*b*). The genetic structure of natural populations of *Drosophila melanogaster*. VIII. Natural selection on the degree of dominance of viability polygenes. *Genetics* **63**, 467–478.

Mukai, T. (1970). Viability mutations induced by ethyl methanesulphonate in *Drosophila melanogaster*. *Genetics* **65**, 335–348.

Mukai, T. (1988). *Proceedings of the 2nd International Conference on Quantitative Genetics*, chap. Genotype-environment interaction in relation to the maintenance of genetic variability in populations of *Drosophila melanogaster*, pp. 21–31. Weir, B. S., Eisen, E. J., Goodman, M. M. and Namkoong, G., eds. Sinauer Associates, Inc., Sunderland, MA.

Mukai, T. & Yamaguchi, O. (1974). The genetic structure of natural populations of *Drosophila melanogaster*. XI. Genetic variability in a large local population. *Genetics* **76**, 339–366.

Mukai, T. & Yamazaki, T. (1968). The genetic structure of natural populations of *Drosophila melanogaster*. V. Coupling-repulsion effect of spontaneous mutant polygenes controlling viability. *Genetics* **59**, 513–535.

Mukai, T., Chigusa, S. & Yoshikawa, I. (1964). The genetic structure of natural populations of *Drosophila melanogaster*. II. Overdominance of spontaneous mutant polygenes controlling viability in homozygous genetic background. *Genetics* **50**, 711–715.

Mukai, T., Chigusa, S. & Yoshikawa, I. (1965). The genetic structure of natural populations of *Drosophila melanogaster*. III. Dominance effect of spontaneous mutant polygenes controlling viability in heterozygous genetic backgrounds. *Genetics* **52**, 493–501.

Mukai, T., Yoshikawa, I. & Sano, K. (1966). The genetic structure of natural populations of *Drosophila melanogaster*. IV. Heterozygous effects of radiation-induced mutations on viability in various genetic backgrounds. *Genetics* **53**, 513–527.

Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. (1972). Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335–355.

Mukai, T. M., Cardellino, R. A., Watanabe, T. K. & Crow, J. F. (1974). The genetic variance for viability and its components in a local population of *Drosophila melanogaster*. *Genetics* **78**, 1195–1208.

Muller, H. J. (1928). The problem of gene modification. In *Verhandlung des V Internationalen Kongesses für Vererbungswissenschaft (Suppl. I, Z. induckt. Abtamm. Vererbungsl.)*, pp. 234–260.

Muller, H. J. (1932). Some genetic aspects of sex. *Amer. Natur.* **8**, 118–138.

Muller, H. J. (1950). Our load of mutations. *Am. J. Hum. Genet.* **2**, 111–176.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutat. Res.* **I**, 2–9.

Nelder, J. & Mead, R. (1965). A simplex method for function minimization. *Comput. J.* **7**, 308–313.

Nordborg, M., Charlesworth, B. & Charlesworth, D. (1996). The effect of recombination on background selection. *Genet. Res.* **67**, 159–174.

Oeltjen, J. C., Malley, T. M., Muzny, D. M., Miller, W., Gibbs, R. A. & Belmont, J. W. (1997). Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains. *Genome Res.* **7**, 315–329.

Ohnishi, O. (1977*a*). Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. II. Homozygous effect of polygenic mutations. *Genetics* **87**, 529–545.

Ohnishi, O. (1977*b*). Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. III. Heterozygous effect of polygenic mutations. *Genetics* **87**, 547–556.

Orr, H. A. & Otto, S. P. (1994). Does diploidy increase the rate of adaptation? *Genetics* **136**, 1475–1480.

Otto, S. P. & Goldstein, D. B. (1992). Recombination and the evolution of diploidy. *Genetics* **131**, 745–751.

Otto, S. P. & Whitton, J. (2000). Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–438.

Pamilo, P., Nei, M. & Li, W.-H. (1987). Accumulation of mutations in sexual and asexual populations. *Genet. Res.* **49**, 135–146.

Paquin, C. & Adams, J. (1983). Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. *Nature* **302**, 495–500.

Partridge, L. & Barton, N. H. (1993). Optimality, mutation and the evolution of ageing. *Nature* **362**, 305–311.

Partridge, L. & Gems, D. (2002). Mechanisms of ageing: public or private. *Nat. Rev. Genet.* **3**, 165–175.

Perrot, V., Richerd, S. & Valéro, M. (1991). Transition from haploidy to diploidy. *Nature* **351**, 315–317.

Peters, A. D. & Keightley, P. D. (2000). A test for epistasis among induced mutations in *Caenorhabditis elegans*. *Genetics* **156**, 1635–1647.

Peters, A. D., Halligan, D. L., Whitlock, M. C. & Keightley, P. D. (2003). Dominance and overdominance of mildly deleterious induced mutations for fitness related traits in *Caenorhabditis elegans*. *Genetics* **165**, 589–599.

Pletcher, S. D., Houle, D. & Curtsinger, J. W. (1999). The evolution of age-specific mortality rates in *Drosophila melanogaster*: Genetic divergence among unselected lines. *Genetics* **153**, 813–823.

Ryan, T. & Joiner, B. (1976). Normal probability plots and tests for normality. Technical report, The Pennsylvania State University.

SAS Institute Inc. (1997). The MIXED Procedure. In *SAS/STAT Software: Changes and Enhancements through Release 6.12*, pp. 571–701. SAS Institute Inc., Cary, NC.

Schuler, J. F. (1954). Natural mutations in inbred lines of maize and their heterotic effect. I. Comparison of parent, mutant and their F1 hybrid in a highly isogenic background. *Genetics* **39**, 908–922.

Schuler, J. F. & Sprague, G. F. (1956). Natural mutations in inbred lines of maize and their heterotic effect. II. Comparison of mother line versus mutant when outcrossed to related inbreds. *Genetics* **41**, 281–291.

Shabalina, S. A. & Kondrashov, A. S. (1999). Pattern of selective constraint in *C. elegans* and *C. briggsae* genomes. *Genet. Res.* **74**, 23–30.

Shabalina, S. A., Ogurtsov, A. Y., Kondrashov, V. A. & Kondrashov, A. S. (2001). Selective constraint in intergenic regions of human and mouse genomes. *Trends Genet.* **17**, 373–376.

Sharp, P. A. (1994). Split genes and RNA splicing. *Cell* **77**, 805–815.

Shaw, F. H., Geyer, C. J. & Shaw, R. G. (2002). A comprehensive model of mutations affecting fitness and inferences for *Arabidopsis thaliana*. *Evolution* **56**, 453–463.

Shen, J. C., Rideout, W. M. & Jones, P. A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22**, 972–976.

Simmons, M. J. & Crow, J. F. (1977). Mutations affecting fitness in *Drosophila* populations. *Annu. Rev. Genet.* **11**, 49–78.

Stephan, W. & Kirby, D. A. (1993). RNA folding in *Drosophila* shows a distance effect for compensatory fitness interactions. *Genetics* **135**, 97–103.

Subramanian, S. & Kumar, S. (2003). Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res.* **13**, 838–844.

Sulston, J. & Hodgkin, J. (1988). *The Nematode Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Takano, T. S. (1998). Rate variation of DNA sequence evolution in the *Drosophila* lineages. *Genetics* **149**, 959–970.

The *C. elegans* Sequencing Consortium (1998). Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science* **282**, 2012–2018.

Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G., Beckstrom-Sternberg, S. M., Margulies, E. H., Blanchette, M., Siepel, A. C., Thomas, P. J., McDowell, J. C. *et al.* (2003). Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788–793.

Thorne, J. L., Kishino, H. & Felsenstein, J. (1991). An evolutionary model for maximum-likelihood alignment of DNA sequences. *J. Mol. Evol.* **33**, 114–124.

Turelli, M. (1984). Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theor. Popul. Biol.* **25**, 138–193.

Vassilieva, L. L. & Lynch, M. (1999). The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* **151**, 119–129.

Vassilieva, L. L., Hook, A. M. & Lynch, M. (2000). The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* **54**, 1234–1246.

Wallace, B. (1957). The effect of heterozygosity for new mutations on viability in *Drosophila*: A preliminary report. *Proc. Nat. Acad. Sci. USA* **43**, 404–407.

Wallace, B. (1963). Further data on the overdominance of induced mutations. *Genetics* **48**, 633–651.

Wallace, B. & Dobzhansky, T. (1962). Experimental proof of balanced genetic loads in *Drosophila*. *Genetics* **47**, 1027–1042.

Watanabe, T. K., Yamaguchi, G. & Mukai, T. (1976). The genetic variability of third chromosomes in a local population of *Drosophila melanogaster*. *Genetics* **82**, 63–82.

Waterston, R. H. (1981). A second informational suppressor, SUP-7 X, in *Caenorhabditis elegans*. *Genetics* **97**, 307–325.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P. *et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562.

Webb, C. T., Shabalina, S. A., Ogurtsov, A. Y. & Kondrashov, A. S. (2002). Analysis of similarity within 142 pairs of orthologous intergenic regions of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Nucleic Acids Res.* **30**, 1233–1239.

Weir, B. (1996). *Genetic Data Analysis II*. Sinauer Associates, Inc., Sunderland, MA.

West, S. A., Lively, C. M. & Read, A. F. (1999). A pluralist approach to sex and recombination. *J. Evol. Biol.* **12**, 1003–1012.

Williams, G. C. (1957). Pleiotropy, natural selection and the evolution of senescense. *Evolution* **11**, 398–411.

Williams, G. C. (1971). *Group Selection*, chap. Introduction, pp. 1–15. Aldine Atherton, Chicago.

Williams, G. C. (1975). *Sex and Evolution.* Princeton University Press, Princeton, N. J.

Willis, J. H. (1999*a*). The contribution of male-sterility mutations to inbreeding depression in *Mimulus guttatus*. *Heredity* **83**, 337–346.

Willis, J. H. (1999*b*). The role of genes of large effect on inbreeding depression in *Mimulus guttatus*. *Evolution* **53**, 1678–1691.

Wright, F. (1990). The effective number of codons used in a gene. *Gene* **87**, 23–29.

Wright, S. (1968). *Evolution and the Genetics of Populations, Vol. 1. Genetic and Biometric Foundations.* Univ. Chicago Press, Chicago.

Wright, S. (1977). *Evolution and the Genetics of Populations, Vol. 3, Experimental Results and Evolutionary Deductions.* Univ. Chicago Press, Chicago.

Yeates, D. K. & Wiegmann, B. M. (1999). Congruence and controversy: toward a higher-level phylogeny of diptera. *Annu. Rev. Entomol.* **44**, 397–428.

Yeo, I. K. & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954–959.

Zdobnov, E. M., von Mering, C., Letunic, I., Torrents, D., Suyama, M., Copley, R. R., Christophides, G. K., Thomasova, D., Holt, R. A., Subramanian, G. M. *et al.* (2002). Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* **298**, 149–159.

Zeng, Z. B. (1992). Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. *Genetics* **131**, 987–1001.

Zeyl, C. & DeVisser, J. A. G. M. (2001). Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics* **157**, 53–61.

Zeyl, C., Vanderford, T. & Carter, M. (2003). An evolutionary advantage of haploidy in large yeast populations. *Science* **299**, 555–558.

A Supplementary Material

Table A.1: List of loci from which introns were extracted, along with the intronic sequence lengths

Locus index	Locus name	No. Introns	Length (bp)
1	Est-6	1	51
2	G6pd	2	147
3	Adh	2	119
4	Pgi	4	710
5	Amyrel	1	56
6	janB	3	176
7	ocn	2	109
8	sry beta	1	58
9	Pgm	2	137
10	vermilion	5	354
11	zeste	1	62
12	CecC	1	69
13	Anp	1	62
14	rux	1	90
15	period	2	128
16	Cu-Zn superoxide dismutase	1	720
17	Mlc1	2	410
18	fru	2	379
19	CG17061	2	185
20	EG0007.9	1	77
21	sn	3	195
22	AP50	2	118
23	DS06238.4	1	62
24	BACR44L22.3	1	53
25	Mth	4	271
26	Rel	1	268
27	Tpi	1	65
28	ciD	1	53
29	Cen190	1	57
30	Osbp	1	71
31	hyd	5	297
32	notch	1	71
33	Idgf1	1	71

Table A.1: (continued)

Locus index	Locus name	No. Introns	Length (bp)
34	csw	3	257
35	Dsor1	3	595
36	drk	3	355
37	ph	3	197
38	100G10.2	1	69
39	h	2	1161
40	Vha68-1	1	87
41	ND75	2	139
42	Gpdh	3	462
43	bcd	3	1126
44	rp49	1	62
45	Acp53Ea	1	62
46	dec-1	1	73
47	prod	2	133
48	T-cp1	1	63
49	p3	2	121
50	tld	1	56
51	sqh	1	73
52	pit	2	137
53	otu	1	68
54	nos	1	73
55	mei-9	2	125
56	bnb	1	73
57	ras2	1	56
58	ras1	2	217
59	Su(P)	1	64
60	Lsp1-gamma	1	65
61	Fbp2	1	59
62	runt	1	409
63	CYP4D2	4	256
64	CYP4D1	2	194
65	GLD	1	73
66	white	4	425
67	tra	2	305
68	Mst26Aa	1	56
69	ref(2)p	2	686
70	yp2	1	68
71	Pgd	1	75
72	spalt	1	57
73	plu	2	116
74	Sos	1	68

Table A.1: (continued)

Locus index	Locus name	No. Introns	Length (bp)
75	Ddc	2	1103
76	cytochrome c oxidase polypeptide VIIc	1	108
77	cytochrome c oxidase polypeptide VIIa	1	55
78	cytochrome c oxidase polypeptide VIc	1	67
79	cytochrome c oxidase polypeptide IV	1	59
80	CG10198	6	632
81	attacin C	1	63
82	OS-E	2	112
83	OS-F	3	562
84	Trap100	2	200
85	CG5446	1	81
86	Hop	1	75
87	CG2947	3	371
88	CHIP	4	337
89	dpp	1	1723
90	anon1G5	1	61
91	ldgf3	2	337
Total		163	20513

Table A.2: List of loci from which 5' or 3' regions were extracted, along with the their sequence lengths

Locus index	Locus name	5' length (bp)	3' length (bp)
1	Plu	449	422
2	H4	1002	444
6	Amyrel	79	53
7	Csw		64
9	Dsor1	598	
10	Drk	620	65
11	ldgf3	506	399
12	Hairy	1036	1032
13	ND75	213	
15	Bicoid	445	978
17	JanB	73	347
19	EG0007.9	686	
20	DS06238	586	
21	BACR44L22.4	723	42
22	H3-Like-Cid-gene / cid	426	
23	G-protein-coupled receptor	483	

Table A.2: (continued)

Locus index	Locus name	5' length (bp)	3' length (bp)
24	RpL32	836	552
25	gstD1	643	
28	Ocn	419	
29	Rdl	546	
31	Fru	666	
32	H3	365	404
33	Hexokinase-t1	465	
34	Hexokinase-t2	237	
35	Hex-C	608	
36	Hex-A	309	
37	Prod	639	562
38	Amy-proximal	657	154
39	Ras1/Ras85D	601	232
41	Anon73B1/CG4101	458	80
43	Lsp1-gamma	655	
44	Acp29AB	538	465
46	CecA1	732	1775
47	CecB	571	333
48	Fbp2	787	258
50	Dpp	564	481
51	Tpi		134
52	Scute	592	
54	CYP4D1	565	130
55	Anon1G5/carravagio	601	
56	Gld	579	111
57	Est-6	1691	181
58	White	675	136
59	Eve	1822	
60	Cu-Zn-Superoxidase-dismutase	1010	
61	Mst26Aa/Acp26Aa	684	
63	Hsp82 / Hsp83	2149	
64	Adh	2189	399
65	Achaete / ac	991	116
66	Ref(2)p	146	146
67	Vermillion	639	
68	Pgi	562	
69	Sal	612	110
70	Mlc1	586	150
71	Hsc70	593	
72	Pgd	558	621
74	Mtn	660	691

Table A.2: (continued)

Locus index	Locus name	5' length (bp)	3' length (bp)
75	CecC	99	
76	Anp	652	
77	Acp53Ea	633	
78	Ras3	579	
79	Acp70A	543	218
82	Acp62	547	
84	Ap-50	592	
88	Crq	585	
89	Cs	470	
90	Dec-1	626	
91	JanA	378	103
93	Sry-alpha	326	
94	Sqh	652	238
96	Idgf1	218	407
97	Lethal of Scute	102	
98	Amy-distal	250	195
99	Transformer	127	105
100	CG10198	48	
101	Ddc (brain)	1682	
102	Attacin C	581	194
103	Os-e	2322	1557
104	Os-f	3128	
105	Drosocin	926	
106	Defensin	1123	
107	Metchnikowin	1496	
108	Bnb		70
Total		52100	8707

Table A.3: List of primer sequences and lengths of DNA that were successfully amplified

Index	Locus Name	Length (bp)	5'/3'	Sequence
9	Dsor1	657	5'	AAAGCGTATCGAATAGTTTA
			3'	CAACAGTTGCCTCCGTAT
10	Drk	662	5'	GCCCTAGATTGCCAAATGT
			3'	CGTCGCAGAGAAATCGTGT
11	Idgf3	575	5'	ATGAACAGTAGCGACCTT
			3'	ACAAACCAGATTAGGAGC
20	DS06238	697	5'	TTTTGGCTCGTTTGTTTC
			3'	TCGTGCGATGTTAGATGC

Table A.3: (continued)

Index	Locus Name	Length (bp)	5'/3'	Sequence
21	BACR44L22.4	612	5'	GTAGGAAAGGTAAGCAGCAAT
			3'	TAGTATGATGTCGTCTTCTTG
22	H3-Like-Cid-gene	613	5'	CTCCGCAAATCCACTGACCAG
			3'	GCTTGGCTCTGCTGTGTCGTG
23	G-protein-coupled receptor	635	5'	GGGATTGGTTTGGAGTGG
			3'	CCGTCAGGTTCACCGTAT
24	RpL32	625	5'	ACGCAAAGACCACCCTAT
			3'	AACGCAGTTCAACTCAAAA
25	gstD1	655	5'	GTAAGTTCCCCACAGCAA
			3'	GGGCAGGTAGTAGAAGTCAA
28	Ocn	638	5'	TGGTGGAAGGATTCTCAA
			3'	GAAACAGGTGCCAAAAGG
29	Rdl	697	5'	CTGCTGCCACTGGACTGA
			3'	GCCGCTTGCACTGTATGGA
31	Fru	601	5'	TTAACAGTTAGCCAGCAG
			3'	TCCAGCGCAAGCAGAATT
33	Hexokinase-t1	622	5'	AGTTCGTTCTCCGCTCTG
			3'	TCTCAAGGTCGTCTATGCTC
34	Hexokinase-t2	600	5'	TGCGACACGGTCACAAAT
			3'	CCTCGGTCATCTTCTCCTG
35	Hex-C	697	5'	GCCTTATCTCATCGTAACTAA
			3'	CTGGTAATCACTCAACACAAA
36	Hex-A	699	5'	CACCAGTCGGCGAGTGAA
			3'	CAGGATGCGGTAGCACAG
38	Amy-proximal	681	5'	ACTGCGACAGGAAGGAAA
			3'	CAATGCTCTTAGCCAGAAAC
39	Ras1/Ras85D	695	5'	GCGAAAGTGACGTGAAAC
			3'	GGATTAGCTGGATGGTGA
43	Lsp1-gamma	671	5'	AATAAATGTCGGAGACCC
			3'	AGAGCCAGTATCACAAGG
47	CecB	647	5'	CCATTGCGAACTAAGTGA
			3'	TCTTTCCCAGCTTCCTAA
50	Dpp	659	5'	GCCGATGTGCAAGTGAGA
			3'	CACTGCGAGGAGTAGAAGC
52	Scute	679	5'	CGAATGTCTAATAAGGAGGAT
			3'	GTTGCCGAATTGATGGTC
54	CYP4D1	647	5'	CACACATCAACTCAATCAAAATC
			3'	CTGGTATCGTCCTCATGTTCC
55	Anon1G5/carravagio	672	5'	GTCACCTCAATCGGAAAT
			3'	AGTGGCTAGTAAAGCAAACA
56	Gld	662	5'	AAAGCCAAGCTCGGATGA

Table A.3: (continued)

Index	Locus Name	Length (bp)	5'/3'	Sequence
			3'	GCACGCCCACCAAACAAT
58	White	697	5'	ACAACAGGCGGATTAACA
			3'	TAGCTCCTGATCCTCTTGG
60	Cu-Zn-Superoxidase-dismutase	651	5'	AGCAGCAAACAAAGCAAA
			3'	ATGACAGTGGCGGGAAAT
61	Mst26Aa/Acp26Aa	685	5'	CGTGACTTCAATGATCTC
			3'	CTTAAATGTGCTCAAAAT
64	Adh	605	5'	ACAAGTAGTGCGAACTCAA
			3'	AGACCGATGCCTCCCAGA
68	Pgi	603	5'	CTGACAGCCCTGAACACC
			3'	GAACGCTGCCTCCTGGTT
69	Sal	640	5'	TACCGAAAGACCTCACAA
			3'	AAGTTAGAAGACGCACCA
70	Mlc1	602	5'	TCCCGCCTGAATGAGAAA
			3'	CGACTGGGCACTCACCAT
71	Hsc70	637	5'	AAACACGGCCAGGTGAGA
			3'	ATGCCAACCGCTGGTAAT
76	Anp	652	5'	GATAGGTTCGTTTAGGGTTA
			3'	CAGGACGACAAGGACCAC
77	Acp53Ea	674	5'	GGTCGTGAACAACCAAAG
			3'	CAGCCAGTAAGCTGAACAAT
78	Ras3	645	5'	GAGCGAGAACGAGAACGA
			3'	TCCGAGGACCACGATTTT
79	Acp70A	646	5'	TGGCGATATTCTGGGTCT
			3'	TCGTAGGCTTCCTGTTCC
82	Acp62	624	5'	TCTGACAGGTTGAGGAGGTA
			3'	CATGGAGTCGATGGGTTT
84	Ap-50	642	5'	TTGTCAAATCGCAGAGTC
			3'	CTCGCCCTTGTGGTTGTA
88	Crq	671	5'	CGTCAAGTGAACCCATTA
			3'	GCCAGAAGACCACGATTA
89	Cs	645	5'	CTTCTTCATCGCCCTTCG
			3'	TATCCGCCTGTAGCACCC
90	Dec-1	675	5'	CGCTTGAAGGGCAGACAC
			3'	GCCTGGACGACAAGGAGC
91	JanA	643	5'	TCGGCTATCAGGACTTTC
			3'	TGAAGCATCTCCACCATC
93	Sry-alpha	655	5'	AGAAGAAGTACCAGTGCGACAA
			3'	GGGCCAATAGCTGTTCCAT
94	Sqh	659	5'	CCGCAGTTCAACAGCATT
			3'	CGGCGGTCTTACGGGATG

B Publications

Research Focus

How many lethal alleles?

Daniel L. Halligan and Peter D. Keightley

Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

Knowledge of the frequency of lethal mutant alleles in a population is important for our understanding of population genetics and evolution, and yet there have been few attempts to measure their number in wild populations. A new study has revealed unexpectedly low numbers of segregating lethal alleles in two species of fish. More experiments are needed, however, to know whether this result is general.

There have been very few attempts to estimate the mean number of lethal alleles per individual in wild populations (R, see Glossary), and until recently there was only one reliable estimate available from any vertebrate species. In a recent article, McCune *et al.* [1] provided new estimates from populations of two different species of fish, bluefin killifish (*Lucania goodei*) and zebrafish (*Danio rerio*). The estimates were both unexpectedly low, if it is assumed that R should scale to genome size or gene number across taxa. Below we discuss previous results and the potential impact of these latest findings.

Mutations with positive effects on fitness are necessary for adaptive evolution, but the vast majority of spontaneous mutations have negative effects on fitness in all taxa studied [2]. New mutant alleles arise spontaneously, and although they are purged from a population by natural selection, a MUTATION–SELECTION BALANCE (see Glossary) is expected to exist, in which the frequency of deleterious alleles fluctuates around an equilibrium. These deleterious mutations could explain many observed phenomena in evolutionary biology. For instance, INBREEDING DEPRESSION is widely believed to be caused by recessive deleterious mutations becoming homozygous in the offspring of related individuals. Knowledge of the frequency and nature of deleterious alleles could help in predicting levels of inbreeding depression, which could have an impact in conservation genetics. Furthermore, the information has relevance for human genetics in the context of genetic counselling in cases of consanguineous marriages.

Despite the potential importance of deleterious mutations, there are still very few estimates of the number of deleterious alleles segregating in individuals from wild populations. One problem in obtaining such estimates is that the majority of deleterious mutations have very small or effectively undetectable effects on fitness [3]. It is much more straightforward to estimate the frequency of recessive mutations with very large homozygous effects. In particular, it should be possible to estimate the number of recessive lethal alleles per individual in wild populations objectively and unambiguously. Furthermore, in *Drosophila*, lethal alleles are thought to contribute as much to inbreeding depression as minor effect deleterious alleles [4,5].

The majority of published estimates of the number of lethal alleles per individual in wild populations come from various *Drosophila* species. The general method for detecting lethal genes in *Drosophila* was suggested by Muller [6], and involves the use of BALANCER CHROMOSOMES to test for lethal alleles in a specific chromosome (Fig. 1). It is then possible to estimate the number of lethal alleles per individual (R), by accounting for the proportion of the genome in the chromosome tested. Much of the data have been summarized and converted to estimates of R by Lewontin [7] (see also [8]). In 18 such experiments [7], all but one estimate of R fell in the range 0.5 to 3.

Reliable estimates from other taxa are needed to make general conclusions about R, and to know whether R is correlated to or affected by genome size, the number and length of coding regions, EFFECTIVE POPULATION SIZE or other demographic factors. Unfortunately, other than for *Drosophila*, there are very few estimates of R, and not all are reliable. In humans, there are no good quantitative data, although it has been suggested that there are at most 1.4 LETHAL EQUIVALENTS per individual [9], which is suggestive of a low R. However, it is very difficult to quantify the effects of recessive deleterious alleles that manifest themselves before birth in humans, and their effects could be substantial. Many species of fish and amphibians fertilize their eggs externally, which provides an excellent opportunity for estimating mortality after inbreeding: Any maternal effects should be small, owing to the fact that eggs do not develop inside the mother (if there

Glossary

Balancer chromosome: A chromosome that suppresses recombination by having multiple large inversions, and is identifiable by phenotype when heterozygous or homozygous. They are often chosen to be lethal when homozygous although this is not necessary.
Effective population size (N_e): A number reflecting the size of an idealized population (i.e. large, random mating, even sex ratio, non-overlapping generations) that is affected by drift and selection to the same extent as the population under consideration.
Gynogenesis: A form of female parthenogenesis in which the embryo only contains maternal chromosomes, owing to the sperm failing to fuse with the egg's nucleus.
Inbreeding depression: The reduction in fitness due to increasing homozygosity.
Lethal equivalent: A group of mutant genes that would cause on average one genetic death.
Mutation–selection balance: the equilibrium formed between spontaneous mutation introducing new deleterious mutations, and natural selection removing them.
***R*:** The mean number of recessive lethal alleles carried by an individual in a population.

Corresponding author: Peter D. Keightley (p.keightley@ed.ac.uk).

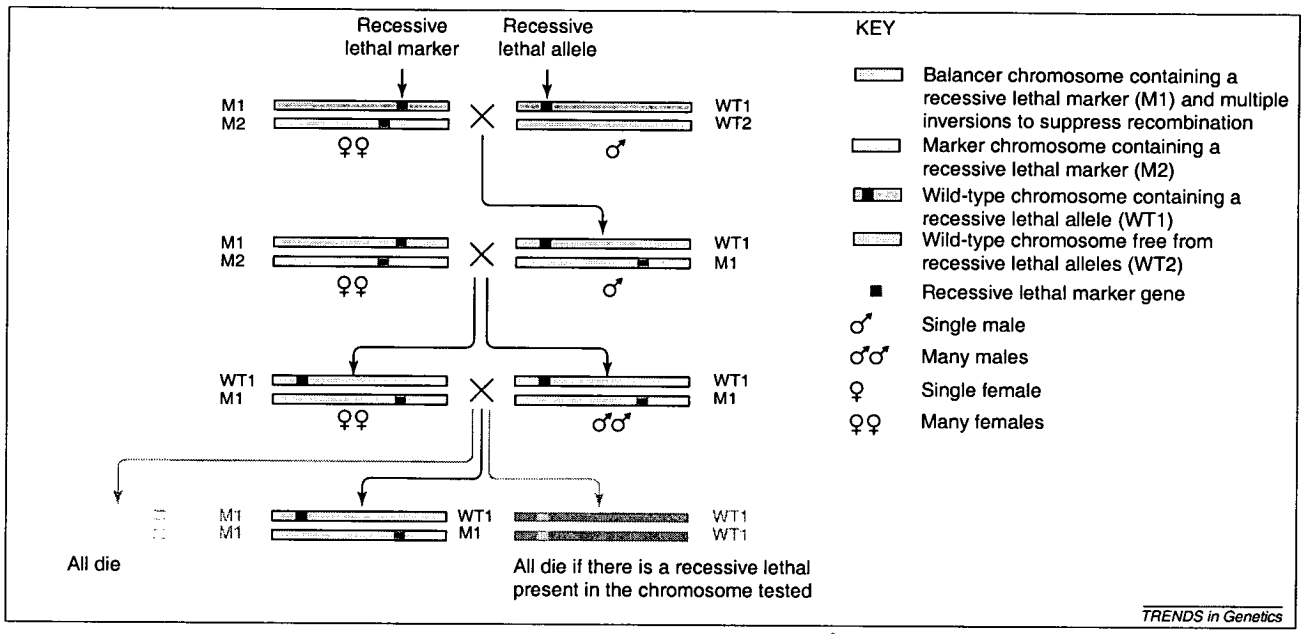


Fig. 1. Balancer chromosome crossing scheme for the detection of recessive lethal alleles on a particular chromosome of interest in *Drosophila* (adapted from [7]). This scheme allows the detection of recessive lethal alleles in one chromosome from a wild-type individual by crossing the individual to a balanced marker stock population. A single wild-type male (carrying two homologous wild-type chromosomes) is crossed to many balanced marker stock females. The balanced marker stock have two different dominant marker genes (M1 and M2) on homologous chromosomes. One of these chromosomes (M1) also contains recombination suppressing inversions that keep the wild-type chromosome intact. A single male is selected from the F1 offspring on the basis of having the M1 heterozygous phenotype, thereby choosing one wild-type chromosome to study, and backcrossed to the marker stock. The backcrossed offspring are intercrossed, producing many offspring, which are scored. If a recessive lethal is present on the wild-type chromosome tested, as in this case, then only heterozygous individuals will be produced from the final cross, providing a simple and objective scheme to test for the presence of a recessive lethal on a random wild chromosome.

are large maternal effects, then only an upper limit for R can be estimated [10]). Amphibians and fish also produce large numbers of offspring, allowing expectations of mendelian ratios to be tested, and offspring that fail to develop can be counted directly. An experiment of this type was carried out in wild-caught *Xenopus laevis* [11]. The experimental design used GYNOGENESIS and inbreeding, to detect the effects of rare recessive lethal alleles in their homozygous state. Fourteen mutants were recovered from eight females giving an estimate of R of 1.875, which is similar to the estimates from *Drosophila* discussed above.

There have been a number of reports of estimates of R in species other than *Drosophila*, but their validity has been questioned [1]. An estimate of $R = 1.6$ from the Mexican salamander (*Ambystoma mexicanum*) [12,13] was reported in [11], although it was not possible to recover this estimate from the original papers [1]. A large estimate of R was also reported in the pacific oyster *Cassostrea gigas* [14], although it has been suggested that departures from mendelian ratios could be caused by factors other than recessive lethal alleles in this species [1]. Another estimate of $R = 3–6$ has been reported from Loblolly pine (*Pinus taeda*) [15], although this estimate was based on data from only one individual.

Until the latest experiment, therefore, the study in *Xenopus* provided the only reliable estimate from a vertebrate species. Recently, however, McCune *et al.* studied bluefin killifish and zebrafish [1] with an experimental design suggested by Timoféeff-Ressovski [16] to estimate R, based on the idea that offspring from crosses between related parents are expected to have fewer viable offspring than crosses between unrelated parents if the related parents share recessive deleterious alleles. Wild-caught parents were mated and their offspring (F1 sibships) were used in brother–sister matings (Fig. 2). Recessive lethal alleles would reveal themselves in most cases as severe morphological mutants in expected mendelian ratios in 25% of brother-sister crosses (Fig. 2). For each species, McCune *et al.* estimated R using a maximum-likelihood method. The estimated numbers of recessive lethal alleles in both the species ($R = 1.87$ for *L. goodei* and $R = 1.43$ for *D. rerio*) are entirely consistent with the *Xenopus* estimate and fall in the middle of the range of the *Drosophila* estimates.

This similarity of R estimates across *Drosophila* and vertebrate taxa is perhaps surprising, given that the *Drosophila* genome is substantially smaller and is thought to have fewer genes than the vertebrate genome. Furthermore, numbers of deleterious mutations that arise in the protein-coding genes appear to be positively correlated with the generation time of a species, and *Drosophila* have a substantially shorter generation time than the fishes in question [17]. All else being equal, therefore, higher numbers of segregating lethal alleles are expected in vertebrate populations. There are several possible factors that could reduce the numbers of segregating lethal alleles in vertebrates. Greater selection against heterozygotes in vertebrates would reduce the frequency of segregating recessive lethal alleles. Selection against heterozygotes potentially accounts for the majority of selection against lethal alleles, because they are not completely recessive, on average, in *Drosophila* [8]. Furthermore, with incomplete recessivity, effective population size of populations and

bottlenecks in the past could have a major role in determining the frequency of lethal alleles [18]. However, there are no data to suggest that this selection is stronger in vertebrates [1]. Alternatively, a lower fraction of essential loci in vertebrates could explain the observation, but proportions of essential loci in *Drosophila* and humans are not, apparently, dissimilar (~20% in both) [1]. More plausible explanations are higher levels of inbreeding in the vertebrate populations or population subdivision; both these factors can have the effect of increasing the rate of purging of recessive deleterious mutations [19,20]. It is also possible that population size and structure could have changed recently, and that historically inbreeding was more prevalent, purging mutations and reducing the number present today. Unfortunately data about the population structure at present, or in the past, are very limited, making these hypotheses difficult to test.

An approximate constancy for R across taxa would have implications for our understanding of inbreeding depression. In *Drosophila*, recessive lethal alleles cause about a half of the inbreeding load [5]. If vertebrates typically carry a similar number of lethal alleles then we could expect these to cause a similar level of inbreeding depression as in *Drosophila*. However, the contribution from more minor effect mutations, to inbreeding depression in vertebrates is still unknown. This contribution could potentially be estimated using an appropriate experimental design. For example, it is expected that lethal alleles will be purged much faster from a population by inbreeding than minor effect mutations, so a comparison of the level of inbreeding depression before and after purging would allow the inference of the proportional contribution of lethal alleles to the overall inbreeding load [21].

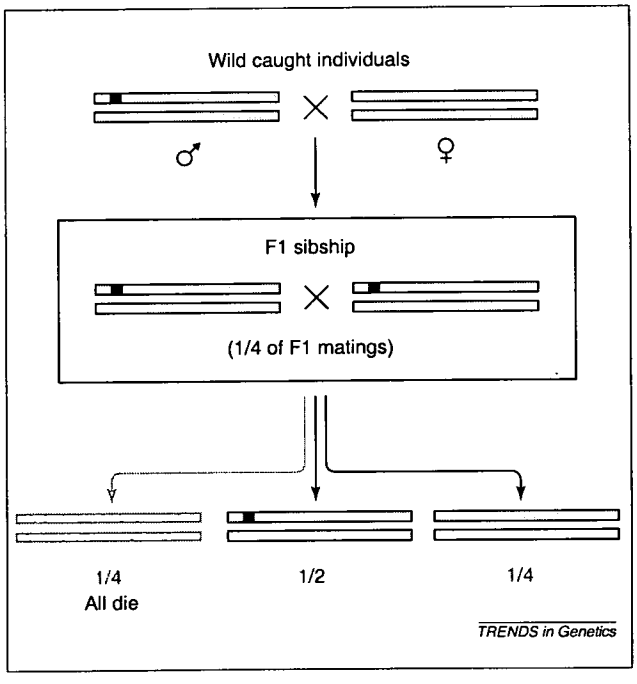


Fig. 2. Crossing scheme used to infer the number of recessive lethal alleles present in two unrelated wild-caught parents. Sibships are produced by mating the two wild-caught parents, and brother–sister crosses are carried out within these sibships. If there is a recessive lethal allele (red) present in either parent, then 25% of these brother–sister crosses (both the brother and the sister have to be heterozygous) will be able to reveal it. In such a cross, 25% of the offspring are expected to show the effects of the recessive lethal allele by failing to survive to reproductive age.

In summary, with only a handful of reliable estimates of R outside of *Drosophila*, it is still unclear whether we should expect to observe similar frequencies of lethal alleles in other outbreeding species. We will therefore need more data in order to draw general conclusions about the frequency of recessive lethal alleles across different taxa. With further information of the kind outlined above, it might then be possible to make more general inferences about the contribution of lethal alleles to inbreeding depression and how this varies across taxa.

Acknowledgements
We would like to thank Brian Charlesworth for useful comments.

References

1 McCune, A.R. *et al.* (2002) A low genomic number of recessive lethals in natural populations of bluefin killifish and zebrafish. *Science* 296, 2398–2401
2 Keightley, P.D. and Lynch, M. Towards a realistic model of mutations affecting fitness. *Evolution* (in press)
3 Keightley, P.D. and Eyre-Walker, A. (1999) Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* 153, 515–523
4 Crow, J.F. and Simmons, M.J. (1983) The mutation load in *Drosophila*. In *The Genetics and Biology of Drosophila* (Vol. 3C) (Ashburner, M. *et al.*, eds), pp. 1–35, Academic Press
5 Charlesworth, B. and Charlesworth, D. (1987) Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.* 18, 237–268
6 Muller, H.J. (1928) The problem of gene modification. In *Verhandlung des V Internationalen Kongesses für Vererbungswissenschaft*, pp. 234–260, (Suppl. I, Z. induckt. Abtamm. Vererbungsl.)
7 Lewontin, R.C. (1974) *The Genetic Basis of Evolutionary Change*, Colombia University Press
8 Simmons, M.J. and Crow, J.F. (1977) Mutation affecting fitness in *Drosophila* populations. *Annu. Rev. Genet.* 11, 49–78
9 Bittles, A.H. and Neel, J.V. (1994) The costs of human inbreeding and their implications for variations at the DNA level. *Nat. Genet.* 8, 117–121
10 Kärkkäinen, K. *et al.* (1999) Why do plants abort so many developing seeds: bad offspring or bad maternal genotypes? *Evol. Ecol.* 13, 305–317
11 Krotoski, D.M. *et al.* (1985) Developmental mutants isolated from wild-caught *Xenopus laevis* by gynogenesis and inbreeding. *J. Exp. Zool.* 233, 443–449
12 Humphrey, R.R. (1975) The axolotl, *Ambystoma mexicanum*. In *Handbook of Genetics* (King, R.C., ed.), pp. 3–18, Plenum Press
13 Humphrey, R.R. (1977) A lethal mutant gene in the Mexican axolotl. *J. Hered.* 68, 407–408
14 Launey, S. and Hedgecock, D. (2001) High genetic load in the pacific oyster *Crassostrea gigas*. *Genetics* 159, 255–265
15 Remington, D.L. and O'Malley, D.M. (2000) Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolly pine family. *Genetics* 155, 337–348
16 Timoféeff-Ressovsky, H.A. and Timoféeff-Ressovsky, N.W. (1927) Genetische analyse einer freilebenden *Drosophila melanogaster* – population. *Wilhelm Roux' Arch. Entwicklungsmech. Org.* 109, 70–109
17 Keightley, P.D. and Eyre-Walker, A. (2000) Deleterious mutations and the evolution of sex. *Science* 290, 331–333
18 Hedrick, P.W. (2002) Lethals in finite populations. *Evolution* 56, 654–657
19 Wang, J.L. *et al.* (1999) Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. *Genet. Res.* 74, 165–178
20 Whitlock, M.C. (2002) Selection, load and inbreeding depression in a large metapopulation. *Genetics* 160, 1191–1202
21 Willis, J.H. (1999) The role of genes of large effect on inbreeding depression in *Mimulus guttatus*. *Evolution* 53, 1678–1691

PII: S0168-9525(02)00045-8

Dominance and Overdominance of Mildly Deleterious Induced Mutations for Fitness Traits in *Caenorhabditis elegans*

A. D. Peters,*,†,[1] D. L. Halligan,* M. C. Whitlock† and P. D. Keightley*

**Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom and* †*Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada*

Manuscript received November 29, 2002
Accepted for publication June 17, 2003

ABSTRACT

We estimated the average dominance coefficient of mildly deleterious mutations ($\bar{h}$, the proportion by which mutations in the heterozygous state reduce fitness components relative to those in the homozygous state) in the nematode *Caenorhabditis elegans*. From 56 worm lines that carry mutations induced by the point mutagen ethyl methanesulfonate (EMS), we selected 19 lines that are relatively high in fitness and estimated the viabilities, productivities, and relative fitnesses of heterozygotes and homozygotes compared to the ancestral wild type. There was very little effect of homozygous or heterozygous mutations on egg-to-adult viability. For productivity and relative fitness, we found that the average dominance coefficient, $\bar{h}$, was ~0.1, suggesting that mildly deleterious mutations are on average partially recessive. These estimates were not significantly different from zero (complete recessivity) but were significantly different from 0.5 (additivity). In addition, there was a significant amount of variation in $\bar{h}$ among lines, and analysis of average dominance coefficients of individual lines suggested that several lines showed overdominance for fitness. Further investigation of two of these lines partially confirmed this finding.

EXPERIMENTS to estimate the average dominance coefficient of new mutations ($\bar{h}$) for fitness-related characters have a long and venerable history. The earliest experiments (WALLACE 1957, 1963; FALK 1961; MULLER and FALK 1961; review in LEWONTIN 1974) were aimed at distinguishing between the "classical" and "balance" theories of genetic variation as defined by DOBZHANSKY (1955), that is, between the maintenance of genetic variation (and its concomitant load) by partially recessive deleterious mutations or by overdominance. More recently, the ubiquity of partially recessive deleterious mutations in several models of evolution (MULLER 1964; HAIGH 1978; KONDRASHOV 1984, 1988; LANDE and SCHEMSKE 1985; CHARLESWORTH 1990; CHARLESWORTH *et al.* 1991; LYNCH *et al.* 1995a; review in CHARLESWORTH and CHARLESWORTH 1998) has fueled efforts to quantify parameters associated with deleterious mutations, including dominance coefficients.

If natural selection is characterized primarily by the purging of recurrent deleterious mutations, then the average dominance coefficient for fitness ($\bar{h}$) is expected to affect the mean fitness of populations at mutation-selection or mutation-selection-drift balance, unless mutations act multiplicatively across loci (KONDRASHOV 1982) or populations are completely panmictic (WHITLOCK 2002). Perhaps more interestingly, $\bar{h}$ is expected to affect the relative fitnesses of outcrossing *vs.* inbreeding populations (CHARLESWORTH *et al.* 1991), as well as those of sexual *vs.* asexual populations (HAIGH 1978; KONDRASHOV 1982; CHARLESWORTH 1990). As such, the mean dominance coefficient of deleterious mutations is an important parameter in models of variation in mating systems and the evolution of sex and recombination.

Despite the recent concentration on studying the parameters of partially recessive deleterious mutations, the question of the relative importance of segregation load arising from overdominant loci *vs.* mutation load caused by recurrent deleterious mutation has never been completely answered. If even a small proportion of mutations impart a fitness advantage in the heterozygous state, then a large portion of the genetic load in outcrossing populations could result from balancing selection (DOBZHANSKY 1955). This would have consequences not only for the evolution of sexual reproduction and the nature of inbreeding depression, but also for the very mechanisms by which genetic variation is maintained in populations. Thus, the quantitative effects of the dominance coefficient of partially recessive deleterious mutations may be less important than the qualitative consequences if an appreciable fraction of mutations are in fact overdominant.

Evidence regarding the distribution of dominance coefficients has been equivocal, both from experiments intended to discriminate qualitatively between balanced and mutation loads and from those aimed at quantifying $\bar{h}$. Of the former, much classic evidence for overdominance came from heterozygote superiority in crosses involving inversion polymorphisms in *Drosophila pseudo-*

[1]*Corresponding author:* Department of Zoology, University of British Columbia, 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada. E-mail: peters@zoology.ubc.ca

obscura (DOBZHANSKY 1954); these results are as easily explained by linkage among partially recessive deleterious mutations (associative overdominance) as by overdominance within loci (OHTA 1971). Of the latter, a few experiments based on induced or laboratory-accumulated mutations in *D. melanogaster* have concluded that $\bar{h}$ is negative (that is, that the average new mutation is overdominant; WALLACE 1957; MUKAI *et al.* 1964; MUKAI 1969), although it is possible that these results were an artifact of the controls used. The majority of experiments yielding estimates of $\bar{h}$ for fitness components in *D. melanogaster* and *Caenorhabditis elegans* have concluded that mutations fall on average between complete recessivity and additivity ($0 \leq \bar{h} \leq 0.5$; SIMMONS and CROW 1977; see reviews in GARCÍA-DORADO and CABALLERO 2000; see also VASSILIEVA *et al.* 2000).

Here we present an experiment to estimate the average dominance coefficient of ethyl methanesulfonate (EMS)-induced mutations in *C. elegans*. We have argued previously that the effects of the GC → AT transitions induced by EMS are similar to those of spontaneous point mutations (DAVIES *et al.* 1999). This makes EMS-induced mutation lines a powerful system for testing the parameters of newly arising deleterious mutations in *C. elegans*.

MATERIALS AND METHODS

Experimental strains and culture conditions: All worm strains were derived from the standard *C. elegans* lab strain N2, originally obtained from the *Caenorhabditis* Genetics Center (CGC). Mutation lines used were a subset of those described previously (DAVIES *et al.* 1999; KEIGHTLEY *et al.* 2000). These lines had been generated by exposing N2 worms to 50 mM EMS for 4 hr and then maintaining by selfing for >10 generations to yield mutations in the homozygous state (for a full description of methods, see DAVIES *et al.* 1999; KEIGHTLEY *et al.* 2000). After ~12 generations of selfing, these lines had been frozen at −85° using standard techniques (SULSTON and HODGKIN 1988).

Because we were interested primarily in the dominance of mildly deleterious mutations, we chose lines that had relatively high fitness as measured in the original experiment (DAVIES *et al.* 1999; KEIGHTLEY *et al.* 2000). These lines, as well as the ancestral N2 strain, were thawed in batches and decontaminated using alkaline hypochlorite (SULSTON and HODGKIN 1988). A total of 33 lines were tested; 14 of these were unusable for crossing (due to either the production of inviable males or the failure to produce a 50:50 sex ratio when crossed). Thus, a total of 19 mutant lines were assayed, split into four blocks; N2 (wild-type) worms were thawed separately three times during the experiment and assayed contemporaneously with each block. Mutant lines were thawed in batches and were maintained for up to eight generations (including initial production of males, as described below) before being crossed and assayed.

Except where specifically noted, worms were maintained at 20° on 3.5-cm MYOB agar plates seeded with *Escherichia coli* strain OP50, using standard techniques (SULSTON and HODGKIN 1988). To generate males for crossing, several dozen individuals of each line were heat-shocked at 25.5° (HODGKIN 1988). The resulting male progeny were then returned to 20° and crossed to hermaphrodites from the same line in a ratio of 4–6 males:1 hermaphrodite; in most cases, such crosses yield a ~50:50 sex ratio among the offspring. Lines were then maintained at 20° for three to eight generations, as both pure hermaphrodite and mixed-sex families; mixed-sex families were produced every generation by crossing males from the previous generation to hermaphrodites from pure-hermaphrodite families in a ratio of 4–6 males:1 hermaphrodite. Pure-hermaphrodite families were maintained by transferring single individuals every generation.

Experimental crosses and life-history assays: We assayed egg-to-adult viability and productivity of four classes of hermaphrodite: homozygous wild type (N2), homozygous mutant, M-heterozygotes (heterozygotes in which the hermaphrodite parent was a mutant), and P-heterozygotes (heterozygotes in which the male parent was a mutant). A total of 1606 worms were assayed for productivity, composed of an average of 33 worms per homozygous mutant line, 14 per M-heterozygote, 18 per P-heterozygote, and 360 wild types. These yielded a total offspring count of 318,310. To remove any inherent differences between offspring produced by self-fertilization and those produced by crossing, all assayed individuals were produced by crossing. To ensure that individuals chosen for assay were the result of crossing (rather than selfing), we placed four to six young (larval stage L4 to young adult) males on a plate with a single late (L4) larval stage hermaphrodite. Pilot experiments suggested that, although offspring produced in the first day after such a cross is set up tend to have hermaphrodite-biased sex ratios, offspring produced on the second and third days show sex ratios that do not differ significantly from 50:50 (data not shown). This suggests that eggs laid on the second or third day of a cross are very unlikely to be the result of self-fertilization, although we also subsequently checked for this in each cross. Thus, on the third day after setting up the crosses, we moved each hermaphrodite to a fresh plate for 6 hr to lay eggs. These eggs, and the worms that developed from them, were used for viability and productivity assays.

Egg-to-adult viability was assayed by counting the number of eggs laid by an adult hermaphrodite in a 6-hr period and then counting the number of larval and adult offspring 3 days later. Eggs laid along the edge of a plate are difficult to see, and egg counts on plates with large numbers of such eggs are likely to be substantial underestimates. To minimize such occurrences, we placed an enclosure (a double-walled screw-top microcentrifuge tube cap) over the parent worm during the egg-lay period, forcing her to lay all eggs in the central part of the plate. The enclosure and parental worm were both removed after 6 hr. Plates on which the worm had escaped from the enclosure were excluded from viability analysis but used in the productivity analysis described below.

Three days following the egg lay, larval worms on each egg-lay plate were counted. At this stage, offspring on the egg-lay plates were sexed, and plates that deviated from a 50:50 sex ratio were excluded from both viability and productivity analyses. Hermaphrodite offspring were removed from the egg-lay plate and placed on a fresh plate for productivity assays. These worms were allowed to lay eggs for 3 days and were moved to new plates every 24 hr during this period. Eggs on these plates were allowed to hatch and grow to an advanced (L4) larval stage, at which point they were counted, to give individual daily productivities for days 1–3. Although *C. elegans* normally produces offspring over a period of 5 days, the majority of offspring are produced over the first 3 days. These plates were repeatedly checked for males over the following 2 days; if more than two male offspring were laid by a hermaphrodite, it was assumed that the parent had undergone cross-fertilization and she was excluded from the analysis (the normal rate of

male production due to spontaneous X-chromosome nondisjunction is ~1/1000; HODGKIN 1988). A number (6–20) of extra worms of every genotype were maintained in parallel with the primary set of plates, and worms that produced an excess of male offspring or were accidentally killed were replaced from this stock.

Daily productivities were included in the calculation of two fitness components: productivity (the unweighted sum of daily productivities over all 3 days) and relative fitness, w. This quantity is proportional to the expected fitness of a population with a stable age structure and is calculated as $w = \Sigma_x e^{-rx} l_x m_x$ (CHARLESWORTH 1994), where $l_x m_x$ is the product of survivorship to and productivity at day x. Since the fitness of the wild type is by definition one in this case, r_C was calculated by setting mean fitness for wild-type worms at $\bar{w}_C = 1$ and solving using $l_x m_x$ across all the wild types within a given assay. The calculated value of r_C was then substituted into the above formula for the calculation of individual estimates of w. Note that this calculation requires only one estimate of r per assay, using the mean $l_x m_x$ table across all the wild-type worms within the assay; this value (r_C) is then used to weight productivity by a negative exponential function over time. Since r is never calculated for individual worms, this measure (w) is defined for worms with $\Sigma l_x m_x = 0$.

Analysis: Viability was assayed as the ratio of the number of adult worms to the number of eggs counted on an egg-lay plate. This trait was analyzed using a generalized linear mixed model via the GLIMMIX macro of SAS 6.12 (LITTELL *et al.* 1996), assuming binomial error structure and a logit link function. This package requires that the numerator be less than or equal to the denominator (*i.e.*, in our case, it requires the assumption that the egg count on each plate was perfect). Due to the uncertainty in counting eggs on the nonuniform agar substrate, there were several cases in which the number of worms exceeded our estimate of the number of eggs. In cases where the worm count exceeded the egg count by one (53 of 273 total plates), we set the egg count equal to the worm count (and therefore the ratio to 1); in cases where the worm count was greater than the egg count by more than one (25 of 273 plates), we discarded the data. Fixed factors in the analysis were "zygosity" (*i.e.*, heterozygous *vs.* homozygous) and maternity (nested within zygosity; *i.e.*, M- *vs.* P-heterozygotes); random factors were reference genotype (*i.e.*, the line from which the mutant parent derived, or N2, the wild type), reference genotype × zygosity, and assay. For comparison, and in an attempt to quantify the bias introduced by correcting for egg undercounts, we also calculated the ratio of worms to eggs for each genotype directly from estimates of the two values, without any correction; standard errors on these estimates were calculated by the delta method (LYNCH and WALSH 1998).

Productivity-related traits (productivity and w) were analyzed by a general linear mixed model using SAS Proc Mixed (LITTELL *et al.* 1996; SAS INSTITUTE 1997). The same effects were fitted for this analysis as for the viability analysis above, with the addition of two random effects: family (offspring from the same cross—that is, a single mother and four to six fathers—are coded as being from the same family; this effect is nested within assay, reference genotype, zygosity, and maternity) and counter (nested within assay). Significance of random effects was tested by Z-statistics, under the assumption that residuals are normally distributed.

Average dominance coefficients ($\bar{h}$) were estimated primarily as the proportional reduction in trait value Z among heterozygotes relative to that among homozygotes, $\bar{h}_1 = (\bar{Z}_{het} - \bar{Z}_{N2})/(\bar{Z}_{hom} - \bar{Z}_{N2})$, where the numerator and denominator are least-squares estimates of differences in trait values, derived from the mixed-model analysis described above. This yields an estimate weighted by the homozygous effect s ($\bar{h}_1 = \Sigma sh/\Sigma s$) (GARCÍA-DORADO and CABALLERO 2000). For comparison, we also estimated $\bar{h}$ as the regression of the trait value in heterozygote lines to that in homozygote lines, $\bar{h}_2 = \sigma_{het,hom}/\sigma^2_{hom}$, where σ and σ^2 are genetic covariances and variances, respectively (CABALLERO *et al.* 1997); estimates from this approach are weighted by the square of the homozygous effect ($\bar{h}_2 = \Sigma s^2 h/\Sigma s^2$). Arguably, the former approach is more appropriate, since genetic variance at mutation-selection balance is proportional to sh (MUKAI *et al.* 1974); furthermore, the regression approach is likely to be more susceptible to bias (CABALLERO *et al.* 1997).

Reassay of selected lines: Three lines (E11, E13, and E25) showed significant evidence of overdominance in the first assay (see RESULTS); these lines were tested again with a modification of the methods described above. Two separate samples of wild-type (N2) worms, as well as one sample each of lines E11, E13, and E25, were thawed, and males were generated again using the methods described above. Four to six families each of males and hermaphrodites of each mutant line plus N2 were maintained separately for three or more generations prior to crossing. Fifteen to 20 homozygote crosses and 7–10 each of M- and P-heterozygote crosses were set up for each mutant line to yield 80–82 homozygotes and 39–41 of each type of heterozygote. Despite repeated attempts, line E25 did not outcross on any of the reassay plates, so data were obtained only for lines E11 and E13. These worms, plus 50 worms from each N2 sample, were assayed for productivity and relative fitness as described above.

Productivity and relative fitness data were analyzed by SAS Proc Mixed, with maternal and paternal treatments (mutant or wild type) and their interaction as fixed effects, and maternal and paternal line (nested within maternal or paternal treatment) and their interaction, maternal and paternal family (nested within maternal or paternal line), sibship, and counter as random effects. The two homozygous N2 (control) replicates differed significantly from each other (see RESULTS), so individual comparisons between heterozygotes and controls were made between crosses involving the same N2 replicate, and more general comparisons involving N2 were made using the replicate (rather than the individual) as the unit of replication.

RESULTS

Mutational effects on viability: The mean viability of wild-type (N2) worms was 93.6% (SE 1.1); that of heterozygotes was 93.0% (SE 1.2); and that of homozygote mutants was 93.4% (SE 1.4). None of the factors included in the generalized linear mixed model significantly affected viability on average. Contrasts involving individual genotypes (Figure 1) reveal three genotypes that show significantly lower viability than the wild type: E5 heterozygotes ($P = 0.02$) and homozygotes ($P = 0.001$), and E24 homozygotes ($P = 0.003$). In addition, the E14 homozygote showed significantly lower viability than the E14 heterozygote ($P = 0.04$), suggesting that the suite of mutations carried by line E14 is on average partially recessive (although neither the heterozygote nor the homozygote showed a significant difference in viability compared to the wild type). A Bonferroni correction for multiple comparisons (28 comparisons) renders all but the E5 homozygote *vs.* wild-type difference ($P = 0.028$) nonsignificant. Direct calculation of worm-

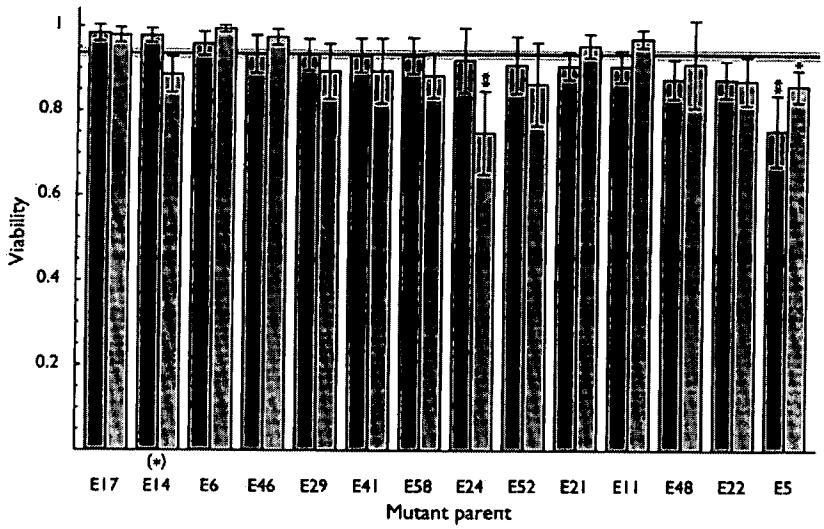


FIGURE 1.—Estimates of heterozygote (darkly shaded bars) and homozygote (lightly shaded bars) viabilities ±SE by line, compared to wild type (solid line) ±SE (lightly shaded field). Asterisks above axis labels correspond to the significance of the dominance effect (het − hom); asterisks above individual bars correspond to the significance of the difference between the value of the given genotype and the wild type. $*P < 0.05$; $**0.001 < P < 0.01$.

to-egg ratios for every genotype gave qualitatively similar results (all point estimates close to one, none significantly different from any other), although the inclusion of plates on which the eggs were undercounted yielded very high standard errors for these estimates.

Overall, these results suggest that mutations, whether in the heterozygous or homozygous state, have very little effect on viability on average. This is particularly important in the context of the present experiment, since differences in viability among heterozygotes, mutant homozygotes, and wild-type individuals in the F_2 progeny would make interpretation of the productivity-based results (below) difficult. More generally, it suggests that viability under benign conditions is a relatively small target for deleterious mutation in worms. This is partially consistent with the results of VASSILIEVA *et al.* (2000), who inferred that a related trait, "survival to maturity," is subject to low mutation rates (~0.003/genome/generation), but that the average effect of each mutation is quite high (~39%). Direct comparisons between these results should be made with care, however, since they measure viability over slightly different time frames: here, viability is a measure of survivorship from egg to late larval stages (L3 and later) and does not score offspring production at all; survival to maturity measures survivorship from an early larval stage (L1) to adulthood and requires the production of viable offspring (VASSILIEVA *et al.* 2000).

Dominance for *w* and productivity: For both productivity and *w*, heterozygotes performed significantly better than mutant homozygotes, and the magnitude of this difference varies significantly among lines; *i.e.*, both zygosity ($F_{1,18} = 16.7$, $P < 0.001$ for productivity; $F_{1,18} = 18.0$, $P < 0.001$ for *w*) and the reference genotype × zygosity interaction ($Z = 2.1$, $P < 0.05$ for productivity; $Z = 2.0$, $P < 0.05$ for *w*) were highly significant. More specifically, contrasts show that, on average, trait estimates for heterozygotes are not significantly different from those for wild-type worms, whereas homozygote estimates are significantly lower than those for both wild-type and heterozygote worms (Table 1). In addition, heterozygote trait values are on average significantly greater than the mean of homozygote and wild-type trait values (Table 1). Thus, on average, deleterious mutations are significantly recessive in two senses: (1) heterozygotes are significantly more fit than homozygous mutants, but not significantly different from wild type, and (2) the effect of mutations in the heterozygous state is significantly smaller than would be predicted under additivity.

Our principal method of estimation of the dominance coefficient is to calculate the proportional change in mean trait value in heterozygotes *vs.* homozygotes ($\bar{h}_1$). Using this approach, we estimate $\bar{h}_1 = \sim 0.1$ for both productivity and *w* (Table 1); these estimates are not significantly different from zero, although they are significantly different from additivity. Estimates based on the regression approach yield somewhat lower estimates of $\bar{h}_2 = 0.02$ for both traits.

Looking beyond the averages, it appears that the suites of mutations represented in each individual mutant line range from overdominant (heterozygote trait value is significantly higher than that of wild type), to recessive (homozygote trait value is significantly lower than that of both heterozygote and wild type; heterozygote and wild type do not differ significantly) and partially recessive (heterozygote and homozygote both significantly lower than wild type and significantly different from each other), to dominant (heterozygote and homozygote both significantly lower than wild type but not significantly different from each other; Figures 2 and 3). To determine whether this variation in dominance is significantly different from zero, we performed a bootstrap analysis of the variance in h_1 for relative fitness (*w*) by resampling estimates of h_1 at the line level

TABLE 1

Effect estimates from mixed-model analysis

Parameter	Productivity estimate (SE)	w estimate (SE)
WT mean	214 (3.3)	1.00 (0.026)
Heterozygote mean	210 (2.5)	0.988 (0.018)
Homozygote mean	184 (2.5)	0.790 (0.019)
Heterozygote *vs.* WT	−3.5 (4.1)	−0.017 (0.032)
Homozygote *vs.* WT	−30.2 (4.1)*	−0.215 (0.032)*
Homozygote *vs.* heterozygote (dominance)	−26.7 (3.3)*	−0.198 (0.026)*
Heterozygote *vs.* (hom + WT)/2 (additivity)	12.7 (3.2)*	0.099 (0.025)*
(hom − WT)/WT ($\bar{s}$)	0.14 (0.02)	0.21 (0.03)
(het − WT)/(hom − WT) ($\overline{h_1}$)	0.12 (0.12)	0.08 (0.14)

Standard errors were computed for the specific counter and assay effects included in this experiment. Standard errors on ratio estimates ($\bar{s}$ and $\bar{h}$) were calculated via the delta method (LYNCH and WALSH 1998) and are similar to those obtained by bootstrapping. *P = 0.0001. Het, heterozygote; hom, homozygote; WT, wild type.

10,000 times. The point estimate of variance in h_1 for w was 1.89; the bootstrap 95% confidence interval was (0.347–3.71), suggesting that the variance in dominance for fitness is significantly greater than zero.

One striking manifestation of this variation in dominance is the apparent existence of overdominance for some lines. This is particularly notable for relative fitness (w), where heterozygote trait values appear to be distributed symmetrically around the wild-type trait value: 10 of 19 lines have point estimates in the overdominant range, 3 of which (E11, E13, and E25) are significant (Figure 2B). To determine whether these three overdominant lines would be expected by chance alone, we calculated the probability of getting three or more false positives in the overdominant portion of the distribution, given a per-test type I error rate of 0.05. P values should be uniformly distributed between zero and one under the null hypothesis. Since we were interested in the tail of the distribution in which heterozygote fitness was greater than wild-type fitness, we excluded the 7 lines in which heterozygote fitness was significantly less fit than the wild type in a one-tailed test. The remaining 12 P values should be uniformly distributed between 0 and 0.95, and the per-test probability of getting $P < 0.05$ is 0.05/0.95. The probability of getting x false positives is a binomially distributed random variable with (n, p) = (12, 0.05/0.95); under such a distribution, $P(x \geq 3)$ = 0.022, suggesting that there are in fact significantly more overdominant lines than would be expected by chance.

The family effect was highly significant for both productivity (Z = 5.2, $P < 0.001$) and w (Z = 7.6, $P < 0.001$), suggesting that there is substantial variation among sets of parents (composed of a single hermaphrodite mother and four to six fathers) within genotypes for productivity-based traits. A large proportion of this effect is likely due to variation in the ages of maternal worms. We have observed that eggs laid by older worms tend to be more developmentally advanced than those laid by younger worms; thus, even genetically identical hermaphrodites whose ages have been synchronized by a timed egg lay may be developmentally out of synchrony if their mothers were of different ages. If this effect carries across generations, then we may not have fully accounted for it statistically with the single-generation family effect we have fitted in our analyses.

Verification of overdominance: To test the hypothesis that the lines E11 and E13 carry overdominant mutations, we reassayed these lines. The two replicates of N2 used as controls for this reassay differed significantly from each other in productivity (T_{30} = 2.3; P = 0.03), although not in w (T_8 = 1.8; P = 0.1). This difference may have arisen from a persistent age-variation effect such as that hypothesized above or from some other persistent environmental effect, such as mild bacterial contamination (although no evidence of such contamination was observed in these lines). It is unlikely to be due to accumulated genetic differences between the replicates, because both sets of lines were frozen at the time the mutation lines were mutagenized. We dealt statistically with the difference between N2's by treating replicate, not individual, as the unit of statistical replication among N2 worms.

The deviation from additivity, as measured by the maternal treatment × paternal treatment interaction, was significant for productivity ($F_{1,7}$ = 8.6, P = 0.02), but not for relative fitness ($F_{1,4}$ = 3.9; P = 0.1), from the reassay. A more detailed analysis shows that the significant interaction effect in productivity is due largely to the fact that heterozygotes of lines E11 and E13 have significantly higher productivity than do wild-type worms and marginally significantly higher productivity than do homozygotes of those lines (Table 2). Homozygote mutants are not significantly different in productivity from wild-type worms. Although these contrasts are not significant for relative fitness, the trends are in the same direction as those for productivity (Table 2).

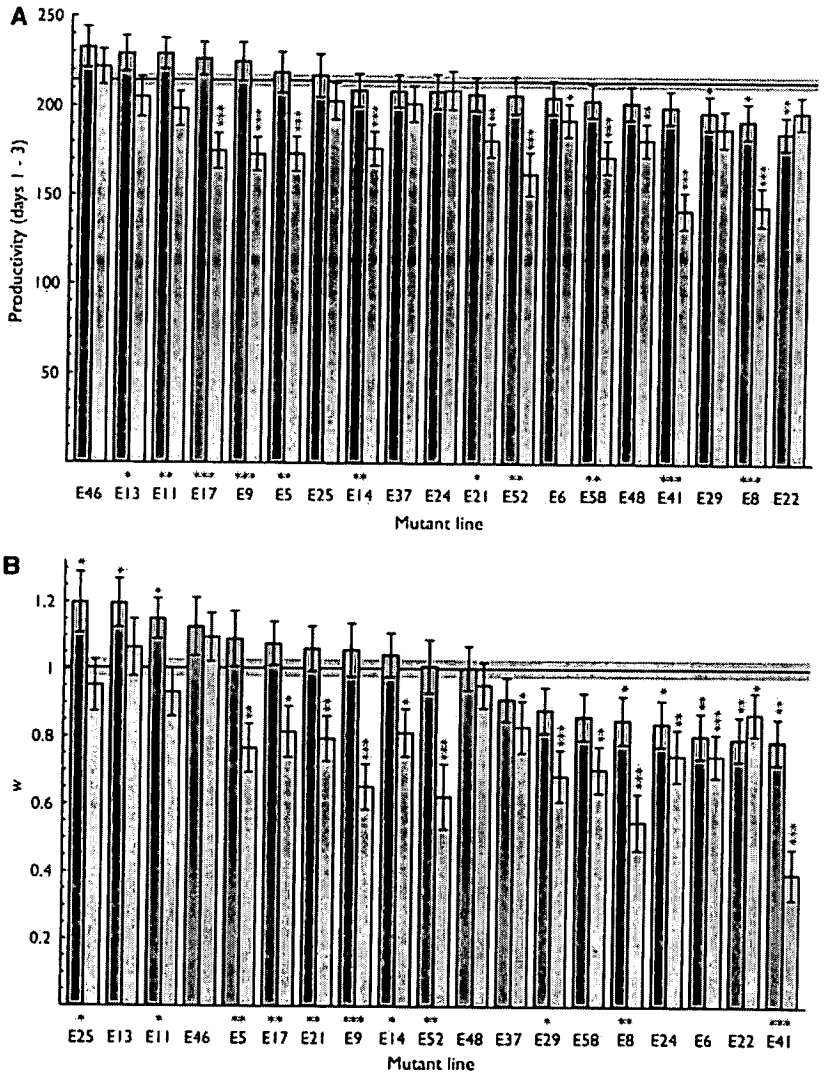


FIGURE 2.—Estimates of heterozygote (darkly shaded bars) and homozygote (lightly shaded bars) productivity (A) and relative fitness (B) ±SE by line, compared to wild type (solid line) ±SE (lightly shaded field). Asterisks above axis labels correspond to the significance of the dominance effect (het − hom); asterisks above individual bars correspond to the significance of the difference between the value of the given genotype and the wild type. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

Thus, the results from our reassay are consistent with the hypothesis that at least 2 of the 19 lines originally assayed display overdominance for at least one fitness-related trait, productivity. Although overdominance for relative fitness is not significant in the reassayed lines alone, pooling the data from the original assay and the reassay yields heterozygote relative fitnesses that are consistently greater than those of wild type and homozygote fitnesses that are either less than or no different from those of wild type (Figure 3). Although this is not a formally significant result, the trend is toward overdominance.

DISCUSSION

Our estimates of $\bar{h}_1 = \sim 0.1$ and $\bar{h}_2 = 0.02$ are broadly consistent with previous results from Drosophila and *C. elegans*, which suggest that mildly deleterious mutations are partially recessive on average. In a series of mutation-accumulation experiments on *D. melanogaster*, Mukai and coworkers made use of balancer chromosomes to protect wild-type chromosomes from selection for 30–60 generations (MUKAI *et al.* 1964, 1965, 1972; MUKAI and YAMAZAKI 1968). For estimates of dominance, chromosomes with >60% normal viability ("quasinormals") were selected, and the viabilities of these chromosomes were assayed in the homozygous and heterozygous states, alongside controls that were homozygous for wild-type or wild-type-like second chromosomes. Similar experiments were carried out by OHNISHI (1977; see reviews in SIMMONS and CROW 1977; GARCÍA-DORADO and CABALLERO 2000) and HOULE *et al.* (1997).

Estimates of $\bar{h}$ in Mukai's experiments depended on the genetic background in which fitness was estimated, whether the heterozygotes were in coupling or repulsion, and the method of calculation. Coupling heterozy-

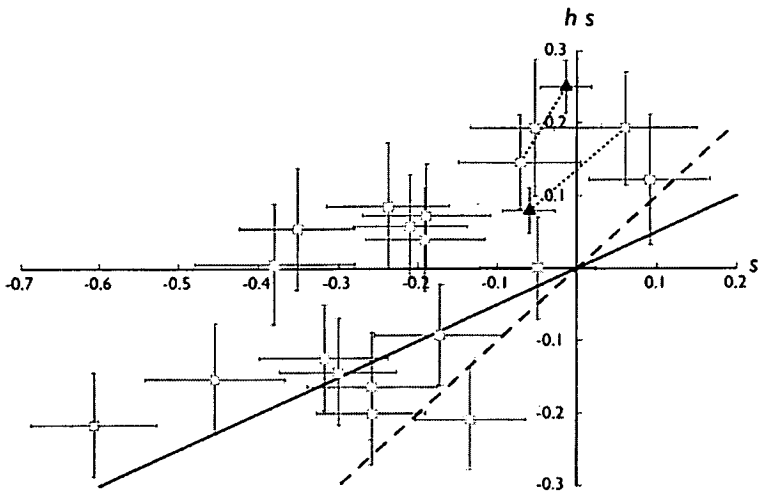


FIGURE 3.—Line estimates of heterozygous (*hs*) *vs.* homozygous (*s*) selection coefficients (±SE). Dashed line represents complete dominance; solid line represents additivity. Shaded boxes represent estimates from original assays; solid triangles represent pooled estimates of lines E11 and E13 from original experiment and reassay; dotted lines connect the two estimates of the same line.

gotes were formed by crossing mutation-accumulation (MA) chromosomes with "wild-type" chromosomes, which were either chromosomes from a healthy MA line (presumed to be close in fitness to the original chromosome; MUKAI *et al.* 1964; MUKAI 1969) or separately collected chromosomes from the same or unrelated populations (MUKAI *et al.* 1965). Repulsion heterozygotes were formed by crossing two MA chromosomes (MUKAI and YAMAZAKI 1968). In all cases, viabilities were compared to a control on the basis of high-viability MA chromosomes. Coupling heterozygotes formed by pairing with the "original" chromosome consistently yielded estimates of overdominance for both $\bar{h}_1$ and $\bar{h}_2$ ($\bar{h} = -0.32$ to -0.09; MUKAI *et al.* 1964; MUKAI 1969), whereas those formed by pairing with nonisogenic chromosomes yielded estimates that were nearly additive ($\bar{h}_2 = 0.27$–0.56; MUKAI *et al.* 1965) or nearly recessive ($\bar{h}_1 = 0.09$–0.13; SIMMONS and CROW 1977), depending on the method of calculation. Repulsion heterozygotes yielded estimates that were consistently nearly additive by either method of calculation ($\bar{h} = 0.36$–0.4; MUKAI and YAMAZAKI 1968). Ohnishi's coupling and repulsion estimates of $\bar{h}_1$ are both consistent with Mukai's repulsion results ($\bar{h}_1 = 0.40$–0.48; OHNISHI 1977); however, estimates of $\bar{h}_2$ from these experiments are much lower ($\bar{h}_2 = 0.12$–0.15; GARCÍA-DORADO and CABALLERO 2000). In a recent MA experiment in *D. melanogaster*, HOULE *et al.* (1997) estimated the mean dominance coefficient across five life-history traits (not including viability) as $\bar{h}_2 = 0.12$ (although neither the overall mean nor any individual

TABLE 2

Effect estimates from mixed-model analysis of reassayed lines

Parameter	Productivity estimate (SE)	*w* estimate (SE)
WT mean	196 (8.8)	1.00 (0.14)
Heterozygote mean	218 (6.4)	1.14 (0.06)
E11	220 (6.5)	1.22 (0.07)
E13	215 (6.5)	1.06 (0.06)
Homozygote mean	197 (6.0)	0.94 (0.06)
E11	195 (6.8)	1.00 (0.10)
E13	200 (6.9)	1.06 (0.06)
Heterozygote *vs.* WT	21.6 (8.6)**	0.15 (0.12)
E11	24.6 (8.9)**	0.23 (0.14)
E13	18.6 (8.9)*	0.06 (0.14)
Homozygote *vs.* WT	1.49 (8.6)	−0.06 (0.14)
E11	−1.29 (9.1)	0.005 (0.14)
E13	4.27 (9.2)	−0.12 (0.15)
Homozygote *vs.* heterozygote	20.1 (8.5)*	0.20 (0.12)
E11	25.9 (6.9)****	0.22 (0.08)***
E13	14.4 (7.0)**	0.18 (0.08)**

Standard errors were computed for the specific counter and assay effects included in this experiment. *$P <$ 0.07; **$P < 0.05$; ***$P < 0.01$; ****$P < 0.001$. WT, wild type.

mean was significantly different from zero); due to the unavailability of appropriate controls (HOULE *et al.* 1994), an estimate of $\bar{h}_1$ could not be calculated. It should also be noted that HOULE *et al.* (1994) included nonquasinormal chromosomes in their analysis, which might lead to a lower estimate of $\bar{h}$ (OHNISHI 1977; GARCÍA-DORADO and CABALLERO 2000).

Recently, VASSILIEVA *et al.* (2000) estimated $\bar{h}_2$ for six life-history characters in lines of *C. elegans* that had undergone 214 generations of mutation accumulation. Across all traits, the average dominance coefficient was $\bar{h}_2 = 0.38$, although the estimates from the different traits fall into two groups: the first (survival to maturity and longevity) are not significantly different from zero ($\bar{h}_2 = -0.025$), while the rest (productivity, intrinsic rate of increase, rate of convergence, and generation rate) are not significantly different from additive ($\bar{h}_2 = 0.59$). Their estimate for productivity ($\bar{h}_2 = 0.64$) differed substantially from ours ($\bar{h}_1 = 0.12$), although the variance on both of these estimates is substantial (SE = 0.18 and 0.12, respectively). This difference cannot be explained solely by the different methods of calculation: our estimate of $\bar{h}_2 = 0.02$ is even lower than our estimate of $\bar{h}_1$. The difference is also unlikely to be explained by the fact that the present experiment preferentially used fitter lines, while VASSILIEVA *et al.* (2000) used lines chosen nearly at random; if anything, high-fitness lines are expected to be biased toward higher values of $\bar{h}$ (GARCÍA-DORADO and CABALLERO 2000). Even the average number and effects of mutations are similar between our EMS lines and the MA lines of Vassilieva *et al.*: for productivity, we have estimated that these EMS lines carry ~1.5 detectable mutations on average, with an average effect of ~23% (KEIGHTLEY *et al.* 2000), while VASSILIEVA *et al.* (2000) estimate that their 214-generation lines carry ~1.6 mutations per haploid (0.015/diploid/generation × 214 generations/2) with an average homozygous effect of ~22%.

One important difference between our experiment and that of VASSILIEVA *et al.* (2000) may lie in the profile of mutation types. Our lines carry EMS-induced mutations whereas those of VASSILIEVA *et al.* (2000) carry spontaneous mutations. Although we have argued that the mutational effects of G/C → A/T transitions, which are the primary form of mutations induced by EMS, should be similar to those of substitution mutations as a whole (DAVIES *et al.* 1999), it is also likely that the proportion of insertion/deletion mutations induced by EMS differs from the spontaneous rate. This might be particularly important in light of recent findings that transposable element insertions show a trend toward higher dominance coefficient on average than the average of all other types of spontaneous mutation in Drosophila (FRY and NUZHDIN 2003). Although there is reason to think that this elevated dominance is due to specific properties of transposable elements and not to insertions in general (FRY and NUZHDIN 2003), and transposable elements are not known to be active in the strains of *C. elegans* used in our experiment or that of VASSILIEVA *et al.* (2000), the possibility that other types of insertions have similarly elevated dominance coefficients cannot be discounted. Certainly the general point that different types of mutations might have systematic differences in dominance is an important one. Thus, if EMS-induced mutations have a different spectrum of mutation types than do spontaneous mutations, EMS-induced mutations may not give a fully representative view of dominance. It is worth noting, however, that even among experiments focusing on EMS-induced mutations the variation in estimates of $\bar{h}$ has been substantial: MUKAI (1970), OHNISHI (1977), and TEMIN (1978) estimate $\bar{h}$ as ~0.03, 0.47, and 0.18, respectively, for mildly detrimental EMS-induced mutations in Drosophila; it therefore seems unlikely that all variation among estimates arises from biases in the mutation profile under EMS.

Another consideration is the possible effect of beneficial mutations on our estimates of dominance. Consider a single line homozygous for a beneficial mutation at one locus and a detrimental mutation at another. If the beneficial mutation increases the trait value by some proportion t and the detrimental mutation decreases the trait value by some proportion s, then the total trait value (assuming no epistasis) is $(1 + t)(1 - s)$. If both mutations have dominance coefficient h, the total trait value in a heterozygote is $(1 + ht)(1 - hs)$, and the apparent dominance coefficient will be $\hat{h} = (1 - (1 + ht)(1 - hs))/(1 - (1 + t)(1 - s)) = h(t - s + hst)/(t - s + st)$. Since $t - s + hst < t - s + st$ for all $0 \leq h < 1$, the apparent dominance coefficient $\hat{h}$ will be smaller than the actual dominance coefficient h. Thus, the existence of beneficial mutations may artificially decrease estimates of $\bar{h}$. In the extreme, this could lead to the spurious appearance of overdominance. This is in fact an example of associative overdominance: crosses between a parent with a mix of beneficial and detrimental mutations and a wild-type parent are equivalent to crosses between two individuals with different sets of detrimental mutations. However, such biases are likely to be important only if a large proportion of mutations are beneficial; we consider this to be unlikely.

In one sense, despite these caveats, our results fit nicely into two emerging patterns: that mildly deleterious mutations are partially recessive on average, but that there is substantial variation in the degree of dominance. Both of these patterns may have implications for evolutionary processes that are driven by deleterious mutations. If $\bar{h}$ is low, then the average strength of selection acting on newly arising mutations in diploid populations may be weak, even if the selection coefficient against homozygotes is not. This affects, for example, the rate of accumulation of deleterious mutations via Muller's ratchet in newly arising asexual populations (HAIGH 1978; GORDO and CHARLESWORTH 2000), al-

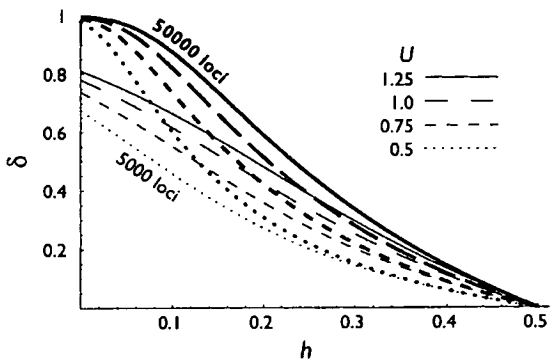


FIGURE 4.—Inbreeding depression (δ) in a population switching from random mating ($f = 0$) to $f = 0.25$, as a function of the dominance coefficient h, genomic deleterious mutation rate U, and the number of loci n contributing to inbreeding load. Thin lines, $n = 5000$ loci; boldface lines, $n = 50{,}000$ loci. Dotted lines, $U = 0.5$; short dashed lines, $U = 0.75$; long dashed lines, $U = 1.0$; solid lines, $U = 1.25$. Loci were assumed to be at mutation-selection balance and interact multiplicatively, with homozygous selection coefficient against deleterious mutations $s = 0.01$. Inbreeding depression was calculated as $\delta = 1 - (W_f/W_0)^n$, where W_0 is the mean fitness at a single locus at mutation-selection balance under random mating, and W_f is the mean fitness at that locus after inbreeding.

though it has very little effect on the probability of fixation in sexual populations (WHITLOCK and BÜRGER 2003). Low values of h may also imply extremely high values of inbreeding depression, although the magnitude of this effect depends on the genomic mutation rate U, the homozygous selection coefficient s, and the number of loci undergoing recurrent deleterious mutation (Figure 4). This has implications for the evolution of mating systems as well as for conservation biology. If inbreeding depression due to partially recessive deleterious mutations is severe, then there can be disruptive selection for selfing *vs.* outcrossing populations. Modifiers that increase selfing are associated with a decrease in fitness in the short term, but become associated with high-fitness genotypes after selfing lineages have been purged of their inbreeding load (LANDE and SCHEMSKE 1985), although over the longer term such lineages may also suffer the effects of Muller's ratchet (LYNCH *et al.* 1995b). Similarly, the decrease in genetic variation triggered by population bottlenecks causes increased inbreeding, which may in turn cause a further decrease in variation as the inbreeding load is purged from the population (HEDRICK and KALINOWSKI 2000).

The inference of substantial variability in dominance coefficients also has implications for evolutionary processes. For example, models invoking deleterious mutations often depend on the assumption that each mutation is partially recessive; however, if h varies, any given mutation may be additive ($h = 0.5$), partially dominant ($h > 0.5$), or even overdominant ($h < 0$) even if the overall estimate of $\bar{h}$ is partially recessive. Mutations with $h > 0.5$ do not contribute to inbreeding depression (Figure 4), while those with $h < 0$ may be maintained as polymorphisms by selection and contribute disproportionately.

Indeed, our results show hints of this pattern, with the majority of point estimates falling in the partially dominant and overdominant ranges, and very few in the partially recessive range (Figure 3). Although it would be a mistake to draw broad conclusions from those lines with $h > 0.5$, the repeated trend of $h < 0$ in some lines suggests that the hypothesis of overdominance is worthy of further study. During the 1950s and 1960s, a series of experiments on induced and accumulated mutations in *D. melanogaster* yielded a set of contradictory conclusions: several experiments found evidence of overdominance; but the detection of overdominance depended on the genetic background or disappeared altogether in replicated experiments (WALLACE 1957, 1963; FALK 1961; MUKAI *et al.* 1964, 1965, 1966, 1972; MUKAI and YAMAZAKI 1968; MUKAI 1969; review in LEWONTIN 1974). In addition, the controls used in these experiments might have had the opportunity to accumulate mutations of their own, leading to the possibility that the patterns observed were due to associative overdominance. Although these objections offer alternative explanations for the patterns seen, they do not directly falsify overdominance *per se*. More recent estimates of $\bar{h}$ for new mutations have failed to provide evidence for overdominance; but, importantly, they were not designed to look for it. Experiments designed and analyzed under the assumption that all mutations are partially recessive might not be expected to yield evidence of overdominance: the effects of small numbers of overdominant alleles might be swamped out by those of partially recessive alleles in lineages carrying more than a few mutations; similarly, if $\bar{h}$ is estimated across multiple lines, the presence of overdominance in individual lines may be obscured. For example, K. SZAFRANIEC and R. KORONA (Figure 4 in unpublished results) report a surprising number (~9/38) of mutation pairs in *Saccharomyces cerevisiae* that repeatably increase fitness in the heterozygous state (*i.e.*, both the average fitness of individuals carrying one of the two mutations and the fitness of individuals carrying both mutations are greater than that of the ancestral wild type), although any tests of the significance of this result would be post hoc.

If even a small proportion of loci exhibit overdominance, then the implications for biological processes are wide ranging indeed. Half a century ago, a central debate in population genetics focused on the relative importance of "balanced" *vs.* "classical" loads because of the fundamentally different mechanisms they implied for the maintenance of variation in populations (DOBZHANSKY 1955). Today, the assumption that variation and loads are due to partially recessive deleterious mutations drives models of many evolutionary processes (review in CHARLESWORTH and CHARLESWORTH 1998).

While the importance of these models is not diminished, because it is clear that most deleterious mutations are in fact partially recessive, the possibility of overdominance may restrict their generality. For example, even if sexual populations are able to purge themselves of deleterious mutations more efficiently than asexual populations (KONDRASHOV 1984; CHARLESWORTH 1990), asexuals that can fix heterozygosity at overdominant loci may still enjoy a fitness advantage over sexuals, depending on the proportion of loci that are overdominant.

Variation in dominance contributes to variation in the effect of newly arising mutations. When combined with our earlier conclusion that a large fraction of deleterious mutations have vanishingly small (but still deleterious) effects under laboratory conditions in *C. elegans* (DAVIES *et al.* 1999), the present result suggests that many newly arising deleterious mutations may have very small effects indeed. In addition, the variation in h we see across lines implies that there is substantial variation in the expected effect of new deleterious mutations. We have suggested previously that the existence of very weak selection against newly arising mutations might change the predictions of deterministic models of mutations and sex (PETERS and KEIGHTLEY 2000). Furthermore, variation in mutation effect has been shown to affect the circumstances under which stochastic processes (Muller's ratchet) are expected to operate (BUTCHER 1995). Our result reinforces the conclusion that newly arising mutations tend to be weak, but that there is great variability among effects, at least when expressed under benign conditions. If some of that variation includes overdominant mutations, then the implications for evolution are far reaching indeed.

We thank S. P. Otto, D. Promislow, and M. Pineda-Krch for extensive discussion; M. Lynch and M. Simmons for extensive comments on an earlier version of the manuscript; and J. Elrick for technical assistance. This work was funded by the Biosciences and Biotechnology Research Council (United Kingdom) and the Natural Science and Engineering Research Council (Canada).

LITERATURE CITED

BUTCHER, D., 1995 Muller's ratchet, epistasis and mutation effects. Genetics **141:** 431–437.

CABALLERO, A., P. D. KEIGHTLEY and M. TURELLI, 1997 Average dominance for polygenes: drawbacks of regression estimates. Genetics **147:** 1487–1490.

CHARLESWORTH, B., 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. Genet. Res. **55:** 199–221.

CHARLESWORTH, B., 1994 *Evolution in Age-Structured Populations.* Cambridge University Press, Cambridge, UK.

CHARLESWORTH, B., and D. CHARLESWORTH, 1998 Some evolutionary consequences of deleterious mutations. Genetica **103:** 3–19.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1991 Multilocus models of inbreeding depression with synergistic selection and partial self-fertilization. Genet. Res. **57:** 177–194.

DAVIES, E. K., A. D. PETERS and P. D. KEIGHTLEY, 1999 High frequency of cryptic deleterious mutations in *Caenorhabditis elegans.* Science **285:** 1748–1751.

DOBZHANSKY, T., 1954 Evolution as a creative process. Caryologia **30:** 435–449.

DOBZHANSKY, T., 1955 A review of some fundamental concepts and problems of population genetics. Cold Spring Harbor Symp. Quant. Biol. **20:** 1–15.

FALK, R., 1961 Are induced mutations in *Drosophila* overdominant? II. Experimental results. Genetics **46:** 737–757.

FRY, J. D., and S. V. NUZHDIN, 2003 Dominance of mutations affecting viability in *Drosophila melanogaster.* Genetics **163:** 1357–1364.

GARCÍA-DORADO, A., and A. CABALLERO, 2000 On the average coefficient of dominance of deleterious spontaneous mutations. Genetics **155:** 1991–2001.

GORDO, I., and B. CHARLESWORTH, 2000 The degeneration of asexual haploid populations and the speed of Muller's ratchet. Genetics **154:** 1379–1387.

HAIGH, J., 1978 The accumulation of deleterious genes in a population: Muller's ratchet. Theor. Popul. Biol. **14:** 251–267.

HEDRICK, P. W., and S. T. KALINOWSKI, 2000 Inbreeding depression in conservation biology. Annu. Rev. Ecol. Syst. **31:** 139–169.

HODGKIN, J., 1988 Sexual dimorphism and sex determination, pp. 243–279 in *The Nematode Caenorhabditis elegans,* edited by W. B. WOOD. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

HOULE, D., K. A. HUGHES, D. K. HOFFMASTER, J. IHARA, S. ASSIMICOPOULOS *et al.*, 1994 The effects of spontaneous mutation on quantitative traits. I. Variances and covariances of life history traits. Genetics **138:** 773–785.

HOULE, D., K. A. HUGHES, S. ASSIMICOPOULOS and B. CHARLESWORTH, 1997 The effects of spontaneous mutation on quantitative traits. II. Dominance of mutations with effects on life-history traits. Genet. Res. **70:** 27–34.

KEIGHTLEY, P. D., E. K. DAVIES, A. D. PETERS and R. G. SHAW, 2000 Properties of ethylmethane sulfonate-induced mutations affecting life-history traits in *Caenorhabditis elegans* and inferences about bivariate distributions of mutation effects. Genetics **156:** 143–154.

KONDRASHOV, A. S., 1982 Selection against harmful mutations in large sexual and asexual populations. Genet. Res. **40:** 325–332.

KONDRASHOV, A. S., 1984 Deleterious mutations as an evolutionary factor I. The advantage of recombination. Genet. Res. **44:** 199–218.

KONDRASHOV, A. S., 1988 Deleterious mutations and the evolution of sexual reproduction. Nature **336:** 435–440.

LANDE, R., and D. W. SCHEMSKE, 1985 The evolution of self fertilization and inbreeding depression in plants. I. Genetic models. Evolution **39:** 24–40.

LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change.* Columbia University Press, New York.

LITTELL, R. C., G. A. MILLIKEN, W. W. STROUP and R. D. WOLFINGER, 1996 *SAS System for Mixed Models.* SAS Institute, Cary, NC.

LYNCH, M., and B. WALSH, 1998 *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Sunderland, MA.

LYNCH, M., J. CONERY and R. BÜRGER, 1995a Mutation accumulation and the extinction of small populations. Am. Nat. **146:** 489–518.

LYNCH, M., J. CONERY and R. BÜRGER, 1995b Mutational meltdown in sexual populations. Evolution **49:** 1067–1080.

MUKAI, T., 1969 The genetic structure of natural poulations of *Drosophila melanogaster.* VI. Further studies on the optimum heterozygosity hypothesis. Genetics **61:** 479–495.

MUKAI, T., 1970 Viability mutations induced by ethyl methanesulfonate in *Drosophila melanogaster.* Genetics **65:** 335–348.

MUKAI, T., and T. YAMAZAKI, 1968 The genetic structure of natural populations of *Drosophila melanogaster.* V. Coupling-repulsion effect of spontaneous mutant polygenes controlling viability. Genetics **59:** 515–535.

MUKAI, T., S. CHIGUSA and I. YOSHIWAKA, 1964 The genetic structure of natural populations of *Drosophila melanogaster.* II. Overdominance of spontaneous mutant polygenes controlling viability in homozygous genetic background. Genetics **50:** 711–715.

MUKAI, T., S. CHIGUSA and I. YOSHIWAKA, 1965 The genetic structure of natural populations of *Drosophila melanogaster.* III. Dominance effect of spontaneous mutant polygenes controlling viability in heterozygous genetic backgrounds. Genetics **52:** 493–501.

MUKAI, T., I. YOSHIWAKA and K. SANO, 1966 The genetic structure of natural populations of *Drosophila melanogaster.* IV. Heterozygous effects of radiation-induced mutations on viability in various genetic backgrounds. Genetics **53:** 513–527.

MUKAI, T., S. I. CHIGUSA, L. E. METTLER and J. F. CROW, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster.* Genetics **72:** 335–355.

MUKAI, T., R. A. CARDELLINO, T. K. WATANABE and J. F. CROW, 1974 The genetic variance for viability and its components in a local population of *Drosophila melanogaster*. Genetics **78:** 1195–1208.

MULLER, H. J., 1964 The relation of recombination to mutational advance. Mutat. Res. **1:** 2–9.

MULLER, H. J., and R. FALK, 1961 Are induced mutations in *Drosophila* overdominant? I. Experimental design. Genetics **46:** 727–735.

OHNISHI, O., 1977 Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. III. Heterozygous effect of polygenic mutations. Genetics **87:** 547–556.

OHTA, T., 1971 Associative overdominance caused by linked detrimental mutations. Genet. Res. **18:** 277–286.

PETERS, A. D., and P. D. KEIGHTLEY, 2000 A test for epistasis among induced mutations in *Caenorhabditis elegans*. Genetics **156:** 1635–1647.

SAS INSTITUTE, 1997 *SAS/STAT Software: Changes and Enhancements Through Release 6.12*, pp. 571–701. SAS Institute, Cary, NC.

SIMMONS, M. J., and J. F. CROW, 1977 Mutations affecting fitness in *Drosophila* populations. Annu. Rev. Genet. **11:** 49–78.

SULSTON, J. E., and J. HODGKIN, 1988 Methods, pp. 587–606 in *The Nematode Caenorhabditis elegans*, edited by W. B. WOOD. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

TEMIN, R. G., 1978 Partial dominance of EMS-induced mutations affecting viability in *Drosophila melanogaster*. Genetics **89:** 315–340.

VASSILIEVA, L. L., A. M. HOOK and M. LYNCH, 2000 The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. Evolution **54:** 1234–1246.

WALLACE, B., 1957 The effect of heterozygosity for new mutations on viability in *Drosophila melanogaster*: a preliminary report. Proc. Natl. Acad. Sci. USA **43:** 404–407.

WALLACE, B., 1963 Further data on the overdominance of induced mutations. Genetics **47:** 1027–1042.

WHITLOCK, M. C., 2002 Selection, load and inbreeding depression in a large metapopulation. Genetics **160:** 1191–1202.

WHITLOCK, M. C., and R. BÜRGER, 2003 The good, the bad, and the ugly: fixation of new mutations in small populations, in *Evolutionary Conservation Biology: The Genetics of Rarity*, edited by R. FERRIERE, U. DIECKMANN and D. COUVET. Oxford University Press, Oxford (in press).

Communicating editor: M. J. SIMMONS

Genet. Res., Camb. (2003), **82**, *pp.* 191–205. *With 4 figures.* © 2003 Cambridge University Press
DOI: 10.1017/S0016672303006499 Printed in the United Kingdom

Estimating numbers of EMS-induced mutations affecting life history traits in *Caenorhabditis elegans* in crosses between inbred sublines

DANIEL L. HALLIGAN[1], ANDREW D. PETERS[1,2] AND PETER D. KEIGHTLEY[1]*
[1] *Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK*
[2] *Department of Zoology, University of British Columbia, 6270 University Blvd, Vancouver, BC, V6T 1Z4, Canada*

(*Received 18 June 2003 and in revised form 19 September 2003*)

Summary

Inbred lines of the nematode *Caenorhabditis elegans* containing independent EMS-induced mutations were crossed to the ancestral wild-type strain (N2). Replicated inbred sublines were generated from the F1 offspring under conditions of minimal selection and, along with the N2 and mutant progenitor lines, were assayed for several fitness correlates including relative fitness (w). A modification of the Castle–Wright estimator and a maximum-likelihood (ML) method were used to estimate the numbers and effects of detectable mutations affecting these characters. The ML method allows for variation in mutational effects by fitting either one or two classes of mutational effect, and uses a Box–Cox power transformation of residual values to account for a skewed distribution of residuals. Both the Castle–Wright and the ML analyses suggest that most of the variation among sublines was due to a few (~ 1·5–2·5 on average) large-effect mutations. Under ML, a model with two classes of mutational effects, including a class with small effects, fitted better than a single mutation class model, although not significantly better. Nonetheless, given that we expect there to be many mutations induced per line, our results support the hypothesis that mutations vary widely in their effects.

1. Introduction

Several important evolutionary phenomena have been hypothesized to be consequences of recurrent deleterious mutation. These include inbreeding depression (Charlesworth & Charlesworth, 1987), the evolution of sex and recombination (Kondrashov, 1988; Charlesworth, 1990), the evolution of mating systems (Charlesworth *et al.*, 1990), ecological specialization (Kawecki *et al.*, 1997), genetic variability for quantitative traits (Bulmer, 1989), senescence (Charlesworth, 1994) and the extinction of small populations (Lande, 1994; Lynch *et al.*, 1995*b*). It has been suggested that mutation accumulation might even threaten the persistence of our own species (Muller, 1950; Kondrashov, 1995; Crow, 1997). Whether or not mutations play a role in these phenomena critically depends on parameters associated with mutations (Turelli, 1984; Caballero & Keightley, 1994), including the genomic mutation rate (U), the distribution of selection coefficients (s) and dominance coefficients (h) of new mutations.

With theory increasingly showing the potential importance of the properties of mutations, there has been a resurgence of interest in attempting to estimate U and mean s and h. Although inferring the distribution of mutation effects has received less attention (Lynch *et al.*, 1999), the distribution of effects is important for several reasons. First, there is good reason to expect that mutation effects vary substantially, because genomes contain sites that vary greatly in functional significance. Second, evaluation of some evolutionary theories, such as the time to mutational meltdown, requires knowledge of the distribution of effects (Lande, 1994, 1995; Butcher, 1995; Lynch *et al.*, 1995*a*). Third, estimates of U and mean s obtained from mutation accumulation experiments might be substantially biased if the distribution of mutation effects is not co-estimated.

Evidence for wide variation in effects of induced mutations comes from an analysis of the effects of ethyl methane sulphonate (EMS) mutagenesis in *C. elegans*

* Corresponding author. e-mail: peter.keightley@ed.ac.uk

(Davies *et al.*, 1999; Keightley *et al.*, 2000). The distribution of effects of EMS-induced mutations was evaluated by comparing an *a priori* estimate of the number of induced mutations at the molecular level with an estimate of the number of mutations detectable from fitness assays. The molecular estimate was obtained from the expected rates of EMS-induced point mutations based on experiments to measure forward mutation rates (Bejsovec & Anderson, 1988) and suppressor-induced reversion mutation rates. This yielded the prediction that approximately 45 deleterious point mutations were induced per homozygous mutant line. However, Davies and co-workers found that only 3·60 ($\pm$1·31) were detectable on the basis of fitness assays (Keightley *et al.*, 2000). It is likely, therefore, that there is a large class of mutations with undetectably small, but deleterious, effects.

In the present experiment we created inbred sublines from a random selection of the EMS-induced mutant lines produced by Davies *et al.* (1999) in an attempt to refine our estimates of the number of mutations per line. By crossing the mutant lines to an inbred wild-type line and inbreeding the offspring, we produced sublines, which are expected to contain a random selection of half of the mutations present in each mutant line. By measuring the fitness of each mutant line, the wild-type control and the individual sublines, it should be possible to estimate the number of mutations present in each mutant line. The pattern of segregation of mutations among sublines should give information about the distribution of mutation effects without having to rely on information from higher order moments. We have used a modification of the Castle–Wright estimator (Castle, 1921; Wright, 1968) and a maximum likelihood (ML) method to estimate the average number of mutations per line. The ML approach can accommodate data for which the distribution of residual data points is significantly different from the expectations of a normal distribution. The method also allows two classes of mutation effect, although it was not possible to fit a continuous distribution of mutation effects owing to the computing time required. Our results are consistent with the conclusions of Davies *et al.*, although we did not have the power to verify the existence of a large class of very small effect mutations.

2. Materials and methods

(i) *Generation of sublines and life history trait assays*

We arbitrarily chose ten of the 56 inbred EMS-induced mutant lines (E1–E9 and E11, collectively termed 'progenitor' lines (p-lines)) produced by Davies *et al.* plus one control line (N2), and thawed them from storage at −80 °C. Unless otherwise stated, worms were maintained at 20 °C on 3·5 cm MYOB agar plates seeded with *Escherichia coli* OP50 using standard techniques (Sulston & Hodgkin, 1988).

N2 males were generated by maintaining a few young N2 hermaphrodites on 6 cm agar plates at 25·5 °C. These were examined daily, and males were moved to agar plates containing several hermaphrodites of the same line and allowed to cross at 20 °C. This was repeated for three consecutive generations, after which time sufficient males had been generated to carry out the crosses described below. Male worms of the N2 strain were then randomly selected and crossed to hermaphrodites of the ten p-lines to produce offspring that were heterozygous for the mutations in each p-line. We checked that the ratio of male to hermaphrodite offspring did not significantly differ from the expected 1 : 1 using a χ^2 test with one degree of freedom. Two of ten p-lines (E1 and E7) produced too few offspring or insufficient males and so could not be included in the experiment.

For each of the eight remaining p-lines, ten F1 hermaphrodite offspring were chosen at random and moved to new plates. Each resulting subline was then inbred for a minimum of ten generations by transferring one larval hermaphrodite, chosen at random, to a new plate every generation. This minimizes selection by bottlenecking the population to one individual each generation and generates offspring that are homozygous for about half of the mutations in the original mutant line, with wild-type (N2) alleles at the rest of their loci. One backup plate was set up each generation in case the primary plate failed. If both of these plates failed, offspring from the previous generation's plates (kept at 16 °C in order to slow their growth) were used. This procedure yielded ten sublines per p-line, labelled E2.1–E2.10, E3.1–E3.10 etc. Only one subline (E4.10) was lost during the inbreeding process owing to the primary, backup and previous generation's plates failure to produce a viable worm, suggesting that the worms were subject to very little natural selection.

Daily productivity and longevity were measured contemporaneously for the control line (N2), the eight p-lines and their respective sublines over three assays. In each assay, each of three people (counters) assayed one worm for each p-line and subline, and eight worms for the control (N2) line per assay, giving a total of nine replicates for each p-line and subline, and 72 replicates for the control line. Within each assay, each counter's plates were randomized with respect to their position in the incubator and the order in which they were counted. Before each assay, replicates were maintained separately for three generations in an attempt to remove any possible maternal effects. If any replicates failed in one assay as a result of unnatural death owing to human error or worms crawling off the plate, extra replicates were added to the same counter's quota in the following assay.

Daily productivity was recorded by counting the number of offspring surviving to the L3 larval stage daily for the first 5 days of productivity. Longevity was scored by recording the day on which the parental worm failed to respond to a light touch from a platinum pick and showed any loss of turgor or visible sign of decay. Four fitness correlates were obtained from the productivity data: early productivity (days 1–2), late productivity (days 3–5), total productivity (days 1–5) and relative fitness. Relative fitness (w) is a measure related to intrinsic population growth rate and is suitable for an age-structured population. To calculate w, the intrinsic growth rate of the controls (r_c) within each assay was computed by solving Eqn 1

$$\sum_x e^{-r_c x} l_c(x) m_c(x) = 1, \tag{1}$$

where $l_c(x)$ and $m_c(x)$ are the least-square means of the proportion of worms surviving to day x and fecundity at day x, respectively, for the controls within an assay. Relative fitness was then calculated separately for each individual from Eqn 2

$$w_{ijk} = \sum_x e^{-r_{ci} x} l_{ijk}(x) m_{ijk}(x), \tag{2}$$

where r_{ci} is the average intrinsic growth rate for the control lines within an assay i, and $l_{ijk}(x)$ and $m_{ijk}(x)$ are the proportions of worms surviving to day x and fecundities at day x, respectively, for assay i, worm j of line k (Charlesworth, 1994, p. 120).

(ii) *Castle–Wright estimator of number of mutations*

The Castle–Wright estimator can be used to calculate the effective number of factors (n_e) contributing to the difference in a trait between two divergently selected inbred lines using information about the phenotypic means and variances of the two progenitor lines and their line-cross derivatives (Castle, 1921; Wright, 1968; Lande, 1981; Cockerham, 1986). We can modify this method to estimate the number of genes contributing to the fitness difference between N2 and a given p-line, assuming that all mutations are additive, unidirectional in effect and unlinked, and have equal effects. With this modification, the Castle–Wright estimator is as follows

$$\hat{n}_e = \frac{(\hat{\mu}_{N2} - \hat{\mu}_i)^2 - \hat{\sigma}^2_{\hat{\mu}_{N2}} - \hat{\sigma}^2_{\hat{\mu}_i}}{4\hat{\sigma}^2_{s_i}}, \tag{3}$$

where $\hat{\mu}_{N2}$ and $\hat{\sigma}^2_{\hat{\mu}_{N2}}$ are the observed mean and sampling variance of the trait value for N2, and $\hat{\mu}_i$ and $\hat{\sigma}^2_{\hat{\mu}_i}$ are the observed mean and sampling variance of mutant p-line i. $\hat{\sigma}^2_{s_i}$ is the segregational variance among the inbred sublines for p-line i (Lynch & Walsh, 1998). The above means and sampling variances, and the segregational variances amongst each p-line's sublines, were estimated using the MIXED procedure of SAS 6.12 (SAS Institute, 1997) for each trait. Factors included in the model were assay (1–3), counter (1–3), line (1–8), line-type (N2, p-line or subline) and subline (1–10, nested within line × line-type). Counter, assay and subline (line × line-type) were treated as random effects; all other effects were treated as fixed.

The standard error of $\hat{n}_e$ for the Castle–Wright estimator can be approximated using the delta method (Lande, 1981). Modifying this formula to use a variance estimate from sublines instead of an F1, we obtain Eqn 4.

$$\mathrm{Var}(\hat{n}_e) \simeq 4\hat{n}_e^2 \left(\frac{4\left(\hat{\sigma}^2_{\hat{\mu}_{N2}} + \hat{\sigma}^2_{\hat{\mu}_i}\right)}{(\hat{\mu}_{N2} - \hat{\mu}_i)^2} + \frac{\mathrm{Var}\left(\hat{\sigma}^2_{s_i}\right)}{\hat{\sigma}^4_{s_i}} \right). \tag{4}$$

This estimate ignores the correction factor proposed for the numerator of the Castle–Wright estimator ($\hat{\sigma}^2_{\hat{\mu}_{N2}} + \hat{\sigma}^2_{\hat{\mu}_i}$) because it has been suggested that this would unduly complicate the variance (Cockerham, 1986).

(iii) *Likelihood approach for estimating mutational parameters*

Using a ML method to estimate the number of loci contributing to the fitness difference between N2 and a given p-line has the advantage that it uses information about the distribution of fitness values amongst sublines. Similar ML approaches have been used to estimate mutational parameters in previous experiments (Keightley, 1994; Keightley & Bataillon, 2000; Keightley *et al.*, 2000; Vassilieva *et al.*, 2000); the method used here is based on Keightley & Bataillon (2000). In general, these approaches assume that mutations have additive effects on fitness that follow a given distribution and that, once these effects are removed, the residual data points are normally distributed with the same environmental variance and mean. As an extension to this method, we have relaxed the assumption of normally distributed residuals by assuming instead that the residuals are distributed normally when transformed by an unknown (but estimated) power κ (Box & Cox, 1964).

Following Box and Cox (1964), we assume that for some unknown κ, observations (y) transformed by the function

$$y^{(\kappa)} = \begin{cases} \dfrac{y^\kappa - 1}{\kappa} & (\kappa \neq 0) \\ \log y & (\kappa = 0) \end{cases}$$

satisfy the full normal theory assumptions, assuming that $y > 0$. This function is continuous at $\kappa = 0$ and is therefore preferable to simply using y^κ as the transformation (Box & Cox, 1964).

Replicates of N2 were assumed to have a mean μ and a variance V_E, and to follow a normal distribution when transformed by an unknown power (κ). The p-line and subline replicates were also assumed to have environmental variance V_E, and the number of mutations in each of the p-lines was assumed to be a Poisson random variable with mean λ. Each mutation was assumed to be unlinked from others, to have a negative effect on the trait and to fall into one of two discrete classes of effect size (s_1 and s_2), where the proportion of class 1 mutations (R) is also a parameter of the model. As a special case, we can assume that the proportion of mutations in class 1 is 1; we term this the one-class model. The model allows any number of fixed effects with any number of levels; we modelled both counter and assay as fixed effects for the experimental data.

The levels within a fixed effect were all assumed to have the same variance but different means (scaled relative to the largest level in each fixed effect). For more than one fixed effect, the total of the relevant difference between levels for each fixed effect is calculated separately for each replicate (k) of each p-line (i), and this total is labelled a_{ik} for the following equations. Because all levels are scaled relative to the largest for each fixed effect, a_{ik} can only be negative, meaning that all residuals will be positive when a_{ik} is removed, satisfying the requirement that $y>0$ for the Box–Cox transformation.

If we let X_{ik} equal the phenotypic value of p-line i replicate k then, according to the assumptions above

$$X_{ik}=\mu+x_{1i}s_1+x_{2i}s_2+a_{ik}+e_{ik}^{(\kappa)}, \tag{5}$$

where x_{1i} is the number of mutations in class 1 for p-line i and x_{2i} is the number of mutations in class 2 for p-line i, s_1 and s_2 are the effects of class 1 and class 2 mutations, respectively, (x_1+x_2) is a Poisson deviate with mean λ, and x_1 is a binomial deviate from a total of (x_1+x_2) possible mutations with a probability of success of R. $e_{ik}^{(\kappa)}$ is a transformed Gaussian deviate with mean 0 and variance V_E.

Similarly, if we let Y_{iln} equal the phenotypic value of subline l, replicate n from the p-line i, then

$$Y_{iln}=\mu+y_{1il}s_1+y_{2il}s_2+a_{iln}+e_{iln}^{(\kappa)}, \tag{6}$$

where y_1 and x_1 are binomial deviates with a total of x_1 and x_2 possible events, respectively, and probabilities of success of 0·5.

In the calculation of the likelihood for each line, the likelihood of obtaining the data for that line at every point in the probability space needs to be summed across all the possible points. In our model, there can be anywhere from 0 to an infinite number of mutations present in each p-line. Of these (j) mutations, any number m ($0\leqslant m\leqslant j$) could be in class 1; the remainder (j–m) belong to class 2. Some number p ($0\leqslant p\leqslant m$) of class 1 mutations and some number q [$0\leqslant q\leqslant(j-m)$] of class 2 mutations are present in each of the ten sublines of a given p-line. For each possible combination of subline class 1 and class 2 mutations, we need to calculate the likelihood of obtaining our subline data for the ten sublines belonging to each p-line. Each p-line and subline has some number of replicates (p-reps and s-reps).

The likelihood associated with a single line's data will therefore be

$$L(\text{line}_i)=\sum_{j=0}^{\infty}\left(p(j|\lambda)\times\sum_{m=0}^{j}\left(\text{bi}(m|j)\times\prod_{k=1}^{\text{p-reps}}f(X_{ik}-ms_1-(j-m)s_2-a_{ik})\right.\right.$$
$$\left.\left.\times\prod_{l=1}^{\text{sublines}}\sum_{p=0}^{m}\left(\text{bi}(p|m)\times\sum_{q=0}^{(j-m)}\left(\text{bi}(q|(j-m))\times\prod_{n=1}^{\text{s-reps}}f(Y_{iln}-ps_1-qs_2-a_{iln})\right)\right)\right)\right), \tag{7}$$

where $p(j|\lambda)$ denotes the (Poisson) probability that the p-line i contains j mutations given the mean λ, bi($m|j$) denotes the (binomial) probability that p-line i contains m class 1 mutations given that line i contains a total of j mutations, and the probability of each mutation being class 1 is R, bi($p|m$) is the (binomial) probability that subline i, l has p class 1 mutations (given that p-line i has m), bi($q|j$–m) is the (binomial) probability that subline i, l has q class 2 mutations (given that p-line i has j–m), and f is a transformed Gaussian probability density function, shown below (adapted from Box & Cox, 1964)

$$f(y)=\frac{1}{\sqrt{2\pi\sigma_{(\kappa)}^2}}\cdot\exp\left(-\frac{1}{2}\left(\frac{(y^{(\kappa)}-\mu_{(\kappa)})^2}{\sigma_{(\kappa)}^2}\right)\right)\cdot y^{(\kappa-1)}, \tag{8}$$

where y and $y^{(\kappa)}$ are the untransformed and transformed observations as described above. There are three parameters: $\mu_{(\kappa)}$ and $\sigma_{(\kappa)}^2$ are the mean and variance of the transformed variable, respectively, and κ is the power of the transformation.

The overall log-likelihood is then obtained by adding the sum of the log-likelihoods across all p-lines

to the log-likelihood for N2 data. The log-likelihood for the N2 data was summed over all N2 replicates, where the likelihood for each N2 replicate is:

$$L(Z_i)=f(Z_i-a_i),$$

where Z_i is the observation for N2 replicate i and a_i denotes the total effect of any fixed effects modelled.

Approximate standard errors for all parameters were calculated by fitting a quadratic function to a profile likelihood of the parameter of interest.

To verify the calculations and functionality of the ML program, simulations were carried out using the same mutational model as in the likelihood calculation.

(iv) *Likelihood maximization*

It is necessary to search the likelihood space thoroughly to be sure that any maximum found is the true global maximum. Starting values for μ, V_E and any fixed effects were estimated from the N2 data. In order to obtain starting values for the remaining parameters, a grid search was carried out, without maximization, where the likelihood was evaluated for a combination of set values for each parameter over a broad range.

Using the most likely values obtained during the grid search, a linear search was then carried out in which a series of fixed values for λ were selected about its starting value, because this is the parameter of interest. The likelihood was maximized with respect to all other parameters, using the simplex algorithm (Nelder & Mead, 1965). The simplex was then restarted using the values for μ, V_E, λ, s and κ that gave the highest likelihood during the linear search, and the likelihood was maximized with respect to all parameters. The simplex algorithm was restarted after each maximization until there was no further increase in the likelihood.

(v) *E5.2 and E5 extra line crosses*

From the primary experiment, it was clear that line E5.2 had a significantly lower relative fitness than either of its progenitor lines (E5 and N2). Under the assumptions that all mutations are deleterious, freely recombining and show no epistasis, this result is unexpected. Possible explanations are: (1) some lines might carry mixtures of mutations with both positive and negative effects on relative fitness, in which case it would be possible for sublines to have fitnesses outside the range of their progenitor lines; (2) mutations in line E5 might interact epistatically, such that they only cause the dramatic reduction in fitness visible in line E5.2 when segregated in a line cross; (3) a new spontaneous mutation occurred during the generations of selfing that produced subline E5.2.

In order to test the hypothesis that the reduction in relative fitness in subline E5.2 was due to a new spontaneous mutation, we subjected both lines E5 and E5.2 to further line crosses. If a new large-effect mutation had occurred during the generations of selfing then we would expect to see its segregation in sublines generated from a cross between E5.2 and N2, and no evidence of it in sublines produced from a cross between E5 and N2. Alternatively, if mutations present in line E5 cancelled out each others' effects on w through epistasis or by having both positive and negative effects on w, then we would expect to see further sublines (generated from lines E5 and N2) performing outside the range of the two progenitor lines.

We generated 20 sublines from both lines E5 and E5.2 using the same experimental design as for the main experiment except that two new (independently frozen) replicates of the ancestral wild-type line (labelled N2A and N2B) were thawed. E5 and E5.2 were each crossed to the males of N2A and N2B, and ten offspring from each cross were selected randomly and selfed under minimal selection conditions for seven generations. This produced 44 different lines that were then assayed for total productivity: N2A, N2B, E5.2, E5 and 20 sublines for each of E5.2 and E5. Six replicates were set up for each of these lines, giving a total of 264 data points.

The results of the productivity assay were analysed as before using the MIXED procedure of SAS 6.12 (SAS Institute, 1997). Lines E5 and E5.2 were analysed separately, and the factors included in each model were line (N2 or E5/E5.2), line-type (wild type, p-line or subline), subline (1–20, nested within line × line-type) and N2 type (A or B). N2 type and subline (line × line-type) were treated as random effects; all other effects were treated as fixed. We also attempted to estimate the number of mutations segregating in line E5.2 and E5 by applying the Castle–Wright estimator and the ML approach discussed above.

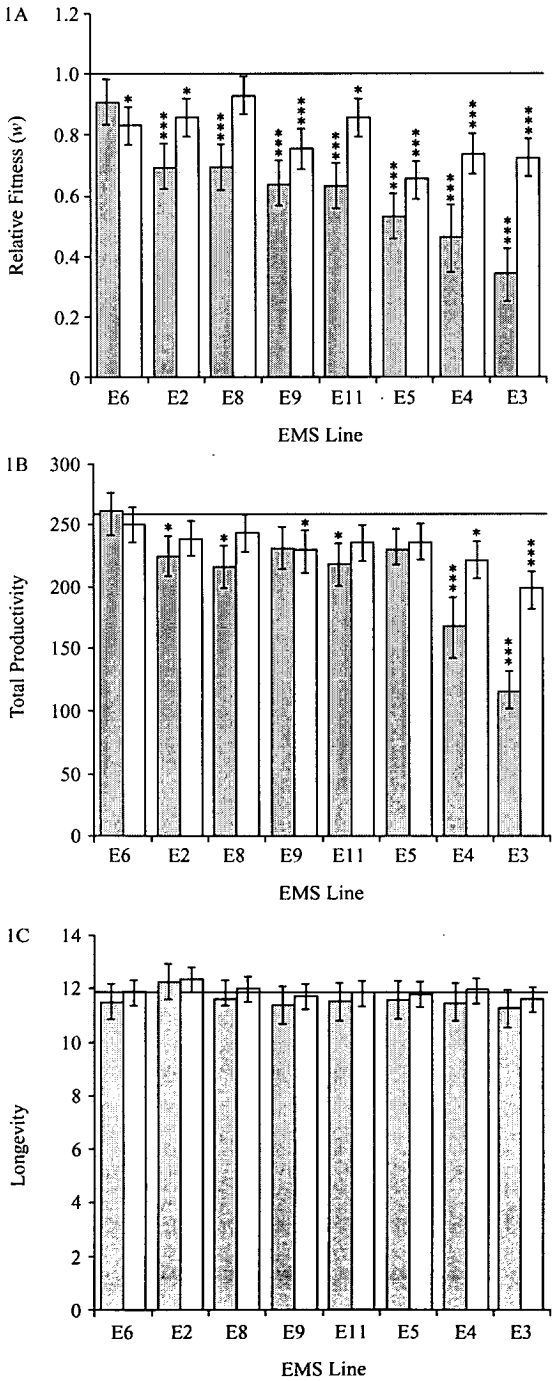
3. Results

(i) *Segregation of mutant phenotypes and Castle–Wright estimates*

A total of 830 data points were obtained from the experiment for five fitness correlates, and a total of 193,157 offspring were counted to obtain the productivity data. EMS mutagenesis has the strongest effects on early productivity, and this is reflected in a large effect on relative fitness (Table 1, Fig. 1). Mutational effects on late productivity and longevity,

Table 1. *LS-means from Proc MIXED for lines of type N2, p-line or subline. Standard errors are shown in brackets*

Trait	N2 mean	p-line mean	Subline mean
w	1·00 (0·0293)	0·611 (0·0739)	0·792 (0·0244)
Early productivity (worms)	211 (15·9)	140 (21·0)	171 (15·7)
Total productivity (worms)	258 (8·30)	208 (16·7)	231 (7·58)
Late productivity (worms)	46·7 (10·3)	70·7 (12·1)	60·1 (9·61)
Longevity (days)	11·9 (0·367)	11·6 (0·523)	11·9 (0·367)



however, are relatively weak on average (Table 1, Fig. 1). This pattern was also noted by Keightley *et al.* (2000), who hypothesized that this was due to mutations lengthening mean development time, resulting in a decrease in early reproductive output. Deleterious mutations might therefore either increase or decrease late productivity, by delaying development or by reducing total productivity. Longevity in particular appears to be a small 'mutational target', with large amounts of environmental variation. This has also been noted in previous literature, several experiments finding little evidence for strong directional effects of mutations on longevity (Keightley & Caballero, 1997; Pletcher *et al.*, 1999; Vassilieva & Lynch, 1999; Keightley *et al.*, 2000). Neither longevity nor late productivity fit a model with only negative-acting mutations, so these traits were excluded from any of the ML analyses.

The EMS-induced mutant lines tested all had lower point estimates for *w* than N2 (seven out of eight were significantly lower; $p<0{\cdot}0001$; Fig. 1A), seven out of eight had lower point estimates for total productivity (five were significant; $p<0{\cdot}05$; Fig. 1B), and, for longevity, none were significant (for all, $p>0{\cdot}5$; Fig. 1C). For *w*, the mean values of the ten sublines fell between those of their respective p-lines and the N2 for all but one of the lines studied (Fig. 1A). This was also true for all but two lines for total productivity (Fig. 1B) and all but three lines for longevity (Fig. 1C). Most individual sublines also had point estimates for *w* between their respective p-line and N2 (Fig. 2), with one major exception: line E5.2 had a significantly lower early productivity, total productivity and *w* ($p<0{\cdot}0001$) than both line E5 and N2 from which it was derived (Fig. 2D). It is shown later that this is likely to be the result of a single large-effect spontaneous mutation that occurred during the ten generations of inbreeding needed to produce sublines.

Fig. 1. Means for relative fitness (*w*), total productivity and longevity for N2 (horizontal bar) ± standard error (grey box), compared with the means of the p-lines and sublines by line ± standard error. Asterisks above the means of the p-lines and sublines correspond to the significance of the difference between the given genotype and the wild type. * $p<0{\cdot}05$; ** $p<0{\cdot}01$; *** $p<0{\cdot}0001$.

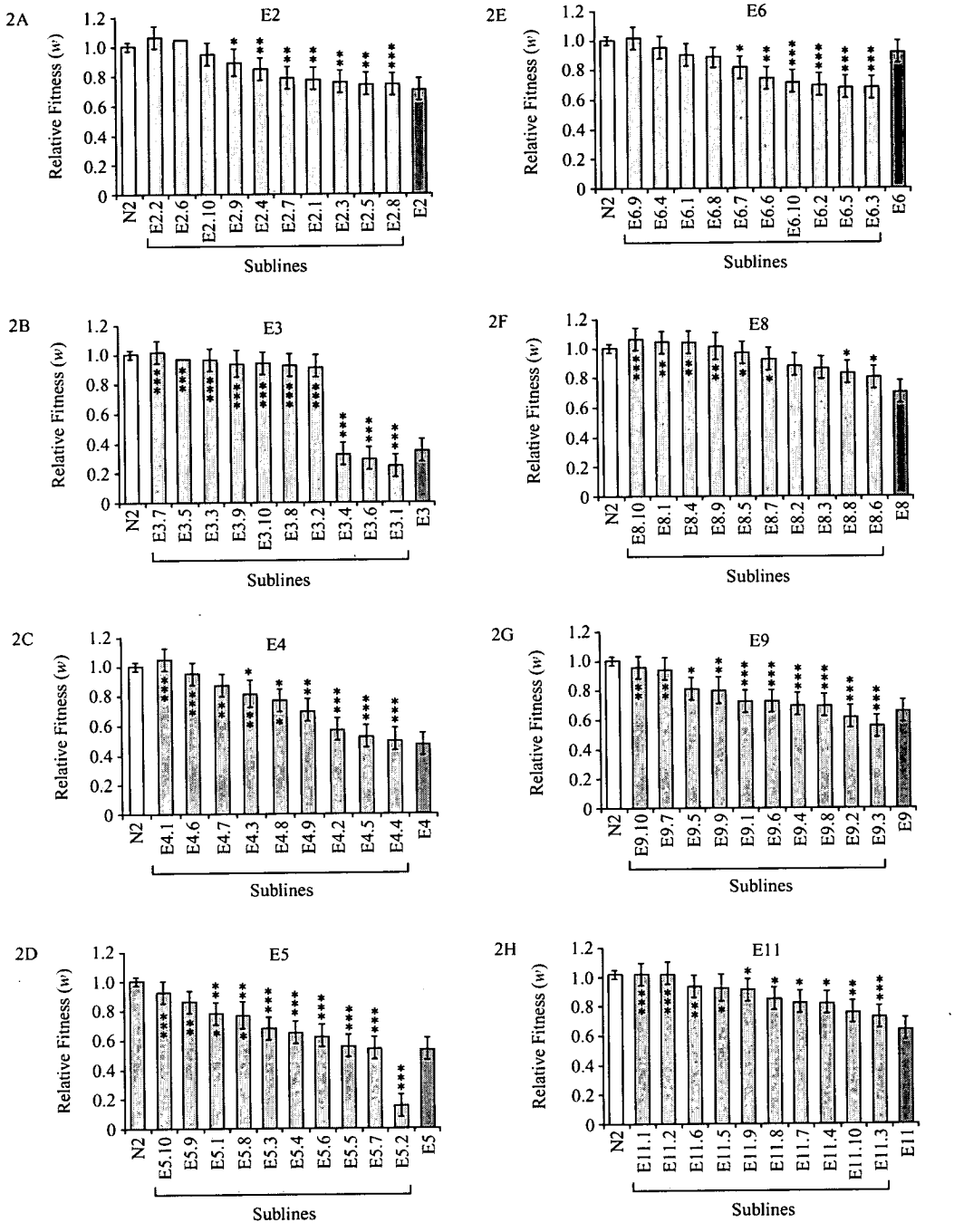


Fig. 2. Means for relative fitness by line, comparing the means for the two progenitor lines (p-line and N2) with all the sublines generated for that line (±standard errors). Asterisks above the error bars correspond to the significance of the difference between the given subline and the wild type. Asterisks below the error bars correspond to the significance of the difference between the given subline and the p-line. * $p<0.05$; ** $p<0.01$; *** $p<0.0001$.

The data for this subline were therefore excluded from the following analyses.

Several of the data points for line E4 were also excluded because many of the worms died during the assay of what were considered to be unnatural causes. Most of these deaths were a result of the worms desiccating after crawling onto the plastic edge of the agar plate. Significantly ($p<0.0001$) more worms from line E4 and its sublines (17 worms) died in this manner than from the rest of the experiment put together (only two worms). It is conceivable that line E4 contains a behavioural mutation that

Table 2. *ANOVA table for mixed-model general linear models (GLMs) of relative fitness* (w), *early productivity, total productivity, late productivity and longevity. Random effects were estimated by restricted maximum likelihood (REML) and significance was tested with* Z *scores rather than* F *statistics*

Trait	Effect	Variance	df_{num}	df_{den}	F	Z
w	Line		7	76·8	1·82	
	Line type		1	78·7	6·72*	
	Subline (line × line-type)	0·0311				5·07**
	Assay	$6·52 \times 10^{-5}$				0·28
	Counter	$9·32 \times 10^{-5}$				0·3
	Residual	0·0587				19·1**
Early productivity	Line		7	77·1	1·48	
	Line-type		1	77	5·36*	
	Subline (line × line-type)	1240				5·21**
	Assay	683				0·323
	Counter	0·928				0·912
	Residual	2110				19·4**
Total productivity	Line		7	76·3	1·65	
	Line-type		1	78	2·37	
	Subline (line × line-type)	1570				5·18**
	Assay	86·7				0·90
	Counter	0·00				–
	Residual	2570				19·1**
Late productivity	Line		7	77·5	5·46**	
	Line-type		1	80·4	2·21	
	Subline (line × line-type)	252				3·73**
	Assay	262				0·98
	Counter	11·8				0·69
	Residual	1430				19·1**
Longevity	Line		7	75·7	0·55	
	Line-type		1	80·6	0·62	
	Subline (line × line-type)	0·132				0·67
	Assay	0·360				0·92
	Counter	0·00				–
	Residual	8·76				18·8**

* $p < 0·05$.
** $p < 0·001$.

causes the worms to be more likely to die in this manner.

To estimate the variability among sublines, we performed a mixed-model analysis (SAS Institute, 1997). The effects of counter and assay on all three traits are non-significant but there is significant variation among sublines within lines for most traits (Table 2, Fig. 2). This suggests that much of the variation among sublines within a line is due to a few mutations of large effect, or that there is substantial variation in mutational effects or epistasis among mutations. The large variation in relative fitness among sublines for several p-lines can be seen in Fig. 2. For example, the sublines of line E3 (Fig. 2B) appear to have a bimodal distribution of relative fitness values, implying that there is one large-effect mutation segregating amongst them. Contrasts between p-line E3 and E3 sublines show that three of the sublines (E3.1, E3.4 and E3.6) are significantly different from N2 ($p \leqslant 0·0005$) but not E3, whereas the other seven sublines are significantly different from E3 ($p \leqslant 0·0001$) but not N2 (Fig. 2B). This pattern is most striking in line E3, although most sublines for the other p-lines show significant differences from one progenitor but not the other. Very few sublines were nonsignificantly different from either progenitor (seven out of 78, excluding subline E5.2). Similarly, very few were significantly different from both (seven out of 78, excluding subline E5.2) (although, in three of these cases, the subline performed worse than either parent). This limited evidence is suggestive of one or two major effect mutations (rather than many similarly sized small effect mutations) for most of the lines tested.

The Castle–Wright estimator was used to estimate the effective number of segregating factors within each mutant p-line. These estimates were then averaged over all eight p-lines to give estimates for each trait (Table 3). Estimates of the effective number of factors using the Castle–Wright estimator are quite low and, despite the large standard errors, are not substantially different from the numbers estimated by

Table 3. *Results for Castle–Wright approach for estimating gene number, for the five traits studied on a line-by-line basis. Approximate standard errors for the number of effective factors are shown in brackets after the estimate along with the average effect* (s). *Overall averages and the standard error of the average are also shown at the base of the Table. Many of the estimates, especially for those traits that were either a small mutational target or might have had their values both increased and decreased as a result of mutagenesis, were incalculable or negative. If the sampling error of the progenitor lines was large, the estimate could be negative, making the calculation of* s *impossible. If the variance amongst the sublines was estimated to be 0 then the estimate for the number of effective factors would be infinite*

	w		Early productivity		Total productivity		Late productivity		Longevity	
Line	n_e	*s*	n_e	*s*	n_e	*s*	n_e	*s*	n_e	*s*
E2	1·53 (2·78)	0·248	1·41 (3·57)	0·214	0·313 (0·832)	0·234	0·116 (0·461)	−1·28	−0·228 (2·20)	–
E3	0·769 (0·848)	0·749	0·653 (0·828)	0·753	0·536 (0·593)	0·754	−0·247 (2·97)	–	−16·5 (2390)	–
E4	1·50 (2·09)	0·440	1·23 (2·02)	0·387	1·21 (2·06)	0·318	∞ (–)	–	∞ (–)	–
E5	2·61 (3·59)	0·289	2·60 (4·39)	0·257	0·281 (0·861)	0·198	1·61 (2·90)	−0·970	−0·305 (3·54)	–
E6	0·0552 (0·196)	0·422	−0·0662 (0·337)	–	∞ (–)	–	0·215 (0·621)	−1·26	∞ (–)	–
E8	5·85 (20·0)	0·127	9·27 (46·8)	0·0905	1·18 (2·91)	0·152	−0·151 (0·761)	–	∞ (–)	–
E9	1·98 (3·10)	0·261	2·44 (5·01)	0·211	0·451 (1·57)	0·161	1·05 (2·41)	−0·801	−0·469 (4·25)	–
E11	3·56 (5·91)	0·196	2·12 (4·92)	0·191	0·661 (1·48)	0·193	−0·0161 (0·0701)	–	∞ (–)	–
Mean	2·23 (2·71)	0·341	2·46 (5·96)	–	– (–)	–	– (–)	–	– (–)	–

Davies *et al.* (1999) or Keightley *et al.* (2000). The Castle–Wright estimator assumes equal effects but, if this assumption is violated, the estimator will underestimate the number of mutations present. Any single large-effect mutation segregating amongst the sublines produced from a cross will lead to a large amount of among-subline variance, reducing the number of factors estimated. It is possible to correct for this bias if the variation of effects is known (Zeng, 1992); alternatively, a ML approach can be used that allows more than one class of mutation effect.

(ii) *Likelihood analysis*

We verified the utility of our ML approach using simulations, the results of which are shown in Table 4. Each set of parameter values in Table 4 was used to simulate 50 data sets. We then used the ML approach to estimate the parameter values from the data. Mean estimates for all parameters do not differ significantly from the simulated values. However, the estimates of some parameters appear to be noisier than others; estimates of κ have the largest standard deviations. Because the accuracy of the estimate of κ depends on the number of data points modelled, the two-class model simulations were designed to have a comparable number of data points per simulation to the experimental data. For each simulation, parameter values were estimated from 600 data points (in comparison to 830 data points for the actual experiment). Over the five sets of simulations, there is a high correlation between the simulated and average estimated values for κ ($r=0{\cdot}927$ for one class of mutational effects; $r=0{\cdot}997$ for two classes of mutational effects).

The one-class model allows one class of mutational effects and assumes additivity; in this respect, it is comparable to the Castle–Wright estimator. The number of mutations estimated for the least noisy traits are all similar, low and not substantially different from the Castle–Wright estimates but have smaller standard errors (Table 5). The two-class model allows for two classes of mutations with different effects. It was expected that including variable effects in this way would lead to higher estimates for the number of mutations with correspondingly lower average effects (Keightley, 1998). However, for the three least noisy traits, the most likely mutational model found was a few (~0·13) very-large-effect mutations (~70%) and many (~1·3) medium-effect mutations (~20%) (Table 5). The large-effect class seems to emerge as a result of the large-effect mutation segregating in line E3 (Table 3, Fig. 2B). With the one-class model, the fitness reduction associated with line E3 can only be explained away with multiple medium-effect mutations; therefore, the number of mutations estimated with the two-class model is lower (albeit not significantly) than that for the one-class model. For all three

Table 4. *Simulation results for maximum likelihood one- and two-class models. Relative fitness data was simulated according to the models described for the ML analyses. For the one-class model, two sublines per p-line were modelled for a total of 20 simulated p-lines with three replicate data points per p-line and subline. For the two-class model, more (30) p-lines with more (five) replicate data points were modelled owing to the extra number of parameters to be estimated. There were 50 replicates per parameter combination and standard deviations over the 50 replicates are shown in brackets*

One-class model

Simulated values				Estimated values			
λ	s	V_E	κ	λ	s	V_E	κ
1	0·05	0·001	1	1·10 (0·545)	0·0496 (0·00840)	0·000984 ($9{\cdot}28 \times 10^{-5}$)	0·658 (2·39)
1	0·1	0·001	1	0·979 (0·196)	0·0999 (0·00307)	0·000979 ($8{\cdot}13 \times 10^{-5}$)	1·10 (1·87)
2	0·05	0·001	1	2·02 (0·498)	0·0487 (0·00437)	0·000990 ($9{\cdot}83 \times 10^{-5}$)	1·04 (2·44)
2	0·1	0·001	2	1·98 (0·315)	0·0995 (0·00192)	0·000977 ($8{\cdot}86 \times 10^{-5}$)	1·66 (1·72)
2	0·1	0·001	2	2·01 (0·305)	0·100 (0·00188)	0·000995 ($9{\cdot}51 \times 10^{-5}$)	1·85 (1·59)

Two-class model

Simulated values						Estimated values					
λ	s_1	s_2	R	V_E	κ	λ	s_1	s_2	R	V_E	κ
1	0·05	0·02	0·4	0·0001	1	0·957 (0·169)	0·0504 (0·00192)	0·0201 (0·000956)	0·394 (0·108)	$9{\cdot}87 \times 10^{-5}$ ($5{\cdot}79 \times 10^{-6}$)	1·11 (2·57)
4	0·05	0·03	0·6	0·0001	−2	4·08 (0·390)	0·0501 (0·000532)	0·0300 (0·000661)	0·594 (0·0661)	$9{\cdot}86 \times 10^{-5}$ ($5{\cdot}40 \times 10^{-6}$)	−2·14 (2·71)
2	0·1	0·03	0·6	0·001	1	2·07 (0·379)	0·0993 (0·00388)	0·0316 (0·00735)	0·588 (0·0943)	0·000980 ($6{\cdot}38 \times 10^{-5}$)	0·876 (1·13)
1	0·05	0·03	0·4	0·001	2	1·28 (0·495)	0·0475 (0·0143)	0·0299 (0·0102)	0·354 (0·194)	0·000990 ($6{\cdot}74 \times 10^{-5}$)	2·27 (1·21)
3	0·05	0·03	0·4	0·001	−1	3·16 (1·17)	0·0534 (0·0192)	0·0317 (0·0138)	0·422 (0·263)	0·000990 ($6{\cdot}31 \times 10^{-5}$)	−0·935 (1·28)

Table 5. *ML parameter estimates for the one- and two-class models of mutation effects. Approximate standard errors are shown in brackets after the parameter estimate. The log-likelihood (loglik) associated with each parameter combination is shown in the final column. This analysis excludes data for subline E5.2 but includes all data from line E3*

Model	Trait	λ	s			μ	V_E	κ	loglik
One-class	w	1·64 (0·731)	0·229 (0·0261)			1·04 (0·0330)	0·0562 (0·00447)	1·31 (0·150)	−26·6
	Early	1·76 (0·706)	0·162 (0·0215)			241 (5·39)	1680 (129)	2·25 (0·198)	−1928·9
	Total	1·49 (0·767)	0·142 (0·0128)			278 (4·26)	1930 (158)	2·31 (0·187)	−1900·6
			s_1	s_2	R				
Two-class	w	1·41 (0·680)	0·743 (0·0589)	0·213 (0·0228)	0·0884 (0·117)	1·03 (0·0329)	1·32 (0·150)	1·32 (0·150)	−21·2
	Early	1·38 (0·663)	0·635 (0·0408)	0·163 (0·0155)	0·0905 (0·128)	241 (4·05)	2·25 (0·198)	2·25 (0·198)	−1920·9
	Total	1·81 (0·768)	0·612 (0·0496)	0·0814 (0·0240)	0·0691 (0·0955)	277 (5·75)	2·25 (0·183)	2·25 (0·183)	−1891·5

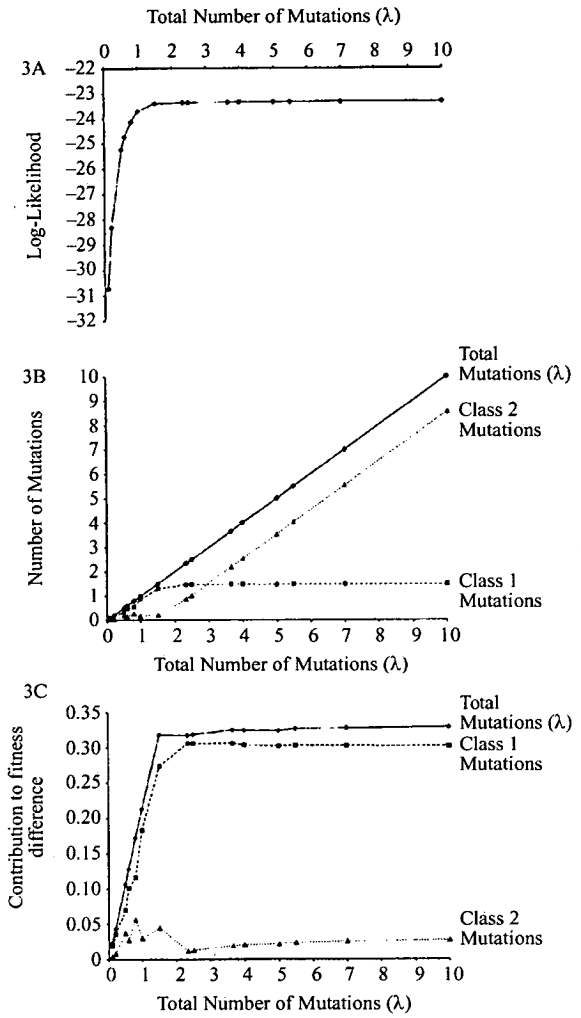


Fig. 3. Plots of total numbers of mutations against log-likelihood (A), number of mutations (B) and contribution to fitness difference of mutations (C), for class 1 mutations (squares), class 2 mutations (triangles) and class 1 + class 2 mutations (diamonds). The number of class 1 or class 2 mutations was calculated by multiplying the proportion of class 1 or class 2 mutations (R or $1-R$) by the total number of mutations. The contribution to fitness difference from class 1 or class 2 mutations is calculated by multiplying the number of class 1 or class 2 mutations by their estimated effect size.

traits studied, the two-class model fitted significantly better than the one-class model ($p<0{\cdot}0001$ in all cases).

The above analysis appears to be dominated by the single large-effect mutation in line E3. Because this might obscure patterns caused by smaller-effect mutations in the other lines, we applied the two-class ML model to our data, excluding line E3. The results of this analysis are shown in Fig. 3A–C for the trait w; similar results were found for early and total productivity. For w, there is virtually no change in log-likelihood above approximately 1·5 mutations

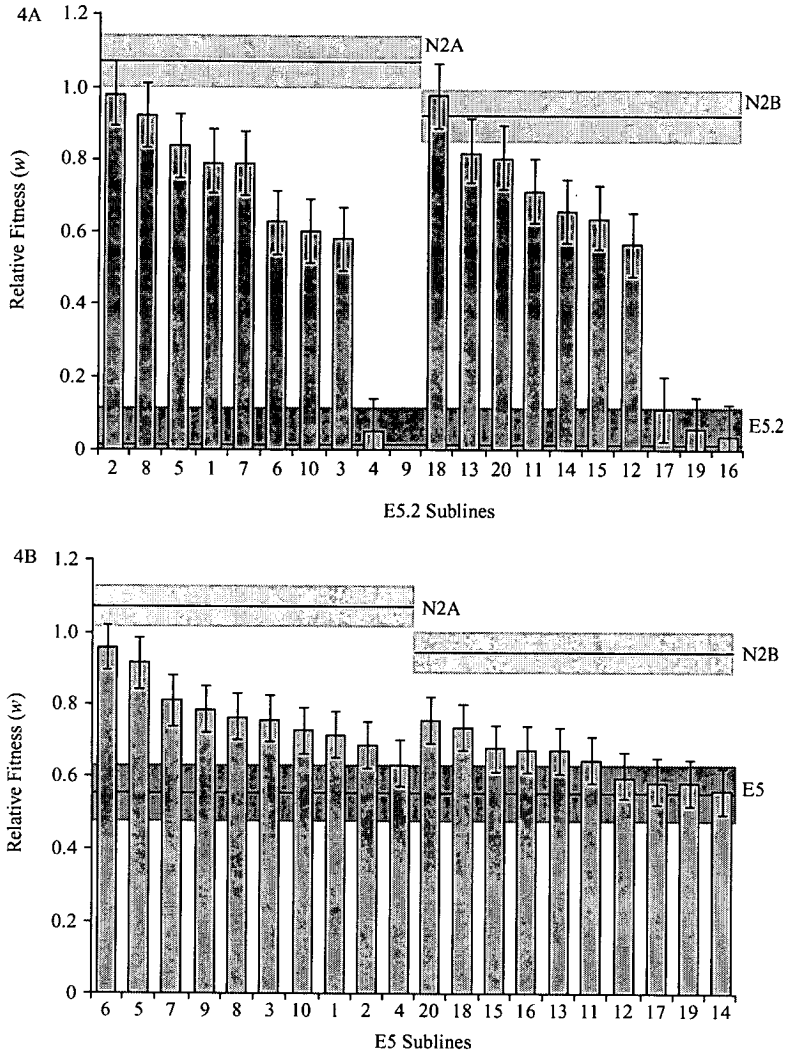


Fig. 4. Means of the sublines and progenitor lines from the analysis of lines E5 and E5.2. Progenitor lines (E5, E5.2, N2A and N2B) are shown as horizontal bars (±standard error as a grey box) above and below the sublines that they correspond to.

(Fig. 3A) (lower confidence limit of 0·487 mutations), suggesting that any number of mutations above ~1·5 is equally supported by the data. As this estimate of total mutation number increases, the number of class 1 (medium-effect) mutations in the best fitting model remains constant (at ~1·5); only the number of class 2 (small-effect) mutations increases (Fig. 3B), and these have correspondingly lower effects on fitness, such that their total contribution to the average fitness difference remains more or less constant (Fig. 3C). The only way to distinguish between a model with a few small-effect mutations (e.g. four total mutations, ~2·5 of which have very small effects of ~0·8%) and a model with many very-small-effect mutations (e.g. 20 total mutations, ~18·5 of which have very small effects of ~0·1%) is to use information about the distribution of these mutations amongst the sublines. It is unlikely, given the number of sublines used in this experiment and the level of environmental variation, that it would be possible to distinguish between these distribution patterns. For all traits, when line E3 was removed, a model with two classes of mutations is more likely than a model with one class, but not significantly so ($p < 0{\cdot}1$).

Estimates of κ, from the two-class ML model including line E3, were tested to see whether they increased the fit to normality of the N2 data after it was transformed, using a Ryan–Joiner normality test (Ryan & Joiner, 1976). Because N2 replicates were assumed to have no mutations, the residual data points could be calculated simply by removing the fixed effects estimated from the ML model. N2 data for both w and early productivity departed significantly ($p < 0{\cdot}025$) from the expectation of a normal

distribution without the transformation, but not once transformed ($p > 0{\cdot}1$). When the same tests were carried out for total productivity, the data did not significantly depart from a normal distribution, with or without the Box–Cox transformation ($p > 0{\cdot}1$). For relative fitness, a significant increase in the likelihood ($p = 0{\cdot}0285$) is obtained when κ is estimated instead of being fixed at 1. The same is true of early productivity ($p < 0{\cdot}0001$) and total productivity ($p < 0{\cdot}0001$).

(iii) *E5.2 and E5 extra line crosses*

Unexpectedly, line E5.2 had a significantly lower relative fitness than either of its progenitors, E5 and N2. To investigate this further, 20 sublines were generated from both lines E5.2 and E5 by crossing them to two freshly thawed replicates of the ancestral N2 (N2A and N2B). Subline 9 generated from line E5.2 was lost during the generations of selfing, owing to the extremely low fitness of the line. Even if this lost line is ignored, it is clear that there is one large-effect mutation present in E5.2, which is segregating amongst the sublines (Fig. 4A). Sublines 4, 16, 17 and 19, which appear to contain this mutation, are not significantly different from their progenitor line E5.2 but are all significantly different from N2 ($p \leqslant 0{\cdot}0001$). Of the remaining 15 sublines, ten are significantly different from both E5.2 and the N2 replicate from which they were generated; only five are not significantly different from line N2. This indicates that there are likely to be some other smaller-effect mutations segregating amongst the sublines of this cross.

Using the Castle–Wright estimator, we estimated that there were 2·64 mutations segregating in line E5.2 (SE 2·39) with an average effect of 0·645. Using our one-class ML model, we estimated that E5.2 contained 1·00 (SE 1·42) mutations with an average effect of 0·957. For the two-class model, we estimated that E5.2 contained 2·00 (SE 2·00) mutations, and that 0·500 of these had an effect of 0·689, whereas the remainder had a lower effect size of 0·278.

All of the extra 20 sublines produced from line E5 had fitness values that were intermediate between the two progenitor lines and there appears to be no evidence of a single large-effect mutation of the size that was observed in the original line E5.2 (Fig. 4B). Using the Castle–Wright estimator, we estimated that there were 5·88 mutations (SE 8·28) with an average effect of 0·212. Applying our one-class ML model, we found the most likely model contained 8·87 mutations, although this model was not a significantly better fit than any models with more than $\sim 0{\cdot}5$ mutations. The most likely two-class model tended towards the results from the one-class model.

We have established that E5.2 contains a single large-effect mutation but we were unable to detect this mutation in the progenitor line E5, suggesting that the mutation occurred spontaneously during the generations of inbreeding that produced line E5.2. Alternatively, it is possible, although unlikely, that the mutation causing the reduction in fitness is present in line E5 but that another tightly linked mutation masked its effects. These mutations might then have been separated after a recombination event during the period of inbreeding that led to line E5.2 but none of the other 29 sublines.

4. Discussion

Davies *et al.* (1999) compared the number of EMS-induced mutations detectable from fitness assays to the number estimated to have been induced in the DNA. They estimated that they had induced an average of at least 45 amino-acid-changing mutations that would be deleterious under natural conditions per homozygous line they studied. However, Davies *et al.* (1999) were able to detect only 1·6 (SE 0·21) (assuming equal effects) or 2·5 (assuming a γ distribution of effects) mutations affecting productivity. Subsequently, Keightley *et al.* (2000) found that only 3·60 ($\pm$ 1·31) mutations could be detected per line on average with effects on relative fitness. The aim of the present experiment was to estimate more accurately the number of induced mutations per EMS-induced mutant line, by producing sublines for a random selection of the mutant lines. The use of sublines allows large-effect mutations to segregate and it should therefore be possible to determine whether the fitness difference between a wild-type line and a single EMS-induced mutant line is primarily due to few or many mutations with correspondingly large or small effects on fitness. This information in turn can then be used to draw inferences about the distribution of mutation effects.

We used a modification of the Castle–Wright estimator (Castle, 1921) to estimate the number of mutations segregating per line and their average effect. With this approach, we estimated that there were 2·23 mutations on average affecting relative fitness (SE 2·71) and 2·46 on average affecting early productivity (SE 5·96). We also developed a maximum-likelihood approach to estimate the number of mutations, which can allow for variable mutation effects, modelled as two classes of effects. Under the assumption of two mutation classes, ML estimates of mutation numbers are lower than either the Castle–Wright or ML estimates under a one mutational class model. This surprising result seems to be a consequence of the segregation of a single large-effect mutation in one line (E3), which is modelled as several medium-effect (15–20%) mutations under the one-class model but as a single large-effect mutation under the two-class model. When line E3 was removed from the analysis,

it was found that the most likely two-class model consisted of approximately 1·5 medium-effect (~20%) mutations plus several smaller-effect mutations affecting w. However, it proved impossible to determine the number and corresponding effect size of these smaller-effect mutations, despite the extra power afforded by producing sublines. Our data are therefore consistent with both a model with several small effect mutations (~3 mutations with an effect size of ~1%) and a model with many very small effect mutations (>20 mutations with an effect size <0·2%). Distinguishing between these models would clearly require very much more data.

Our estimates of mutation number are dependent on how we treat variability in effects of mutations. If it is assumed that all mutations have the same effect then it is possible to obtain a concrete estimate of their number, but this is not possible if we assume that there are two classes of effects. Unfortunately, we were unable to test the fit of a γ or other continuous distribution because of the limits of computing power, but such an analysis might provide greater support for a leptokurtic distribution of mutation effects than the two-class model.

There are at least three possible explanations for the difference between the numbers of mutations estimated to have been induced and the number of mutations detected at the phenotypic level. If the estimate of at least 45 deleterious mutations induced per p-line is correct then our results suggest that the distribution of mutation effects is highly leptokurtic and that a large class of mutations have undetectable effects in laboratory assays. This is consistent with several other direct and indirect estimates of the shape of the distribution of mutation effects. For example, transposable elements provide an opportunity to control the number of mutational events at the DNA level, and experiments using these have provided estimates of the distribution of mutation effects. Analysis of the effects of P-element insertions in *Drosophila melanogaster* on metabolic parameters (Clark *et al.*, 1995) and bristle numbers (Lyman *et al.*, 1996) suggest that mutations with the smallest effects are the most frequent. Similarly, there is direct evidence from Tn10 insertions in *E. coli* for an L-shaped distribution of mutational effects (Elena *et al.*, 1998; Elena & Lenski, 1997). A second possibility is that each p-line carries many fewer than 45 deleterious mutations on average, because our estimates of the number carried are indirect. A possible way to resolve this would be to estimate the number of mutational events at the DNA sequence level directly (Denver *et al.*, 2000). Finally, it is possible that assaying fitness under standard lab conditions would not reveal every large effect deleterious mutation and that assaying fitness under a range of environments could reveal many more potentially large-effect deleterious mutations.

If the distribution of mutation effects is L-shaped and the vast majority of deleterious spontaneous mutations have nearly neutral (but still deleterious) effects on fitness then this could have implications for several areas of evolutionary theory. For example, mildly detrimental mutations on the border of neutrality are the most damaging to population viability if the effective population size is larger than a few individuals (Lande, 1994). Second, mutations of very small effect are undetectable in the vast majority of fitness assays, leading to underestimates of the mutation rate, which has implications for our understanding of the evolution of sex. It is thought that that the diploid mutation rate must be above one per generation for sexual reproduction to be maintained by deleterious mutations alone (Kondrashov, 1988, 1995). Many estimates of the mutation rate from mutation accumulation experiments that do not account for variability in the effects of mutations fall well below this value (see reviews by Drake *et al.*, 1998; Keightley & Eyre-Walker, 1999; Lynch *et al.*, 1999). However, these might be substantial underestimates if the degree of variation in mutation effects is high.

We thank B. Charlesworth, J. Christians and two anonymous referees for helpful comments and suggestions, and J. Elrick for technical assistance.

References

Bejsovec, A. & Anderson, P. (1988). Myosin heavy-chain mutations that disrupt *Caenorhabditis elegans* thick filament assembly. *Genes and Development* **2**, 1307–1317.

Box, G. E. P. & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B* **26**, 211–252.

Bulmer, M. G. (1989). Maintenance of genetic variability by mutation-selection balance: a child's guide through the jungle. *Genome* **31**, 761–767.

Butcher, D. (1995). Muller's ratchet, epistasis and mutation effects. *Genetics* **141**, 431–437.

Caballero, A. & Keightley, P. D. (1994). A pleiotropic nonadditive model of variation in quantitative traits. *Genetics* **138**, 883–900.

Castle, W. E. (1921). An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. *Proceedings of the National Academy of Sciences of the USA* **81**, 6904–6907.

Charlesworth, B. (1990). Mutation–selection balance and the evolutionary advantage of sex and recombination. *Genetical Research* **55**, 199–221.

Charlesworth, B. (1994). *Evolution in Age-structured Populations*. Cambridge, UK: Cambridge University Press.

Charlesworth, D. & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics* **18**, 237–268.

Charlesworth, D., Morgan, M. T. & Charlesworth, B. (1990). Inbreeding depression, genetic load, and the evolution of outcrossing rates in a multilocus system with no linkage. *Evolution* **44**, 1469–1489.

Clark, A. G., Wang, L. & Hulleberg, T. (1995). Spontaneous mutation rate of modifiers of metabolism in *Drosophila*. *Genetics* **139**, 767–779.

Cockerham, C. C. (1986). Modifications in estimating the number of genes for a quantitative character. *Genetics* **114**, 659–664.

Crow, J. F. (1997). The high spontaneous mutation rate: is it a health risk? *Proceedings of the National Academy of Sciences of the USA* **94**, 8380–8386.

Davies, E. K., Peters, A. D. & Keightley, P. D. (1999). High frequency of cryptic deleterious mutations in *Caenorhabditis elegans*. *Science* **285**, 1748–1751.

Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L. & Thomas, W. K. (2000). High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**, 2342–2344.

Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* **148**, 1667–1686.

Elena, S. F. & Lenski, R. E. (1997). Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–397.

Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A. & Lenski, R. E. (1998). Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Contemporary Issues in Genetics and Evolution* **7**, 349–358.

Kawecki, T. J., Barton, N. H. & Fry, J. D. (1997). Mutational collapse of fitness in marginal habitats and the evolution of ecological specialisation. *Journal of Evolutionary Biology* **10**, 407–430.

Keightley, P. D. (1994). The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**, 1315–1322.

Keightley, P. D. (1998). Inference of genome-wide mutation rates and distributions of mutation effects for fitness traits: a simulation study. *Genetics* **150**, 1283–1293.

Keightley, P. D. & Caballero, A. (1997). Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the USA* **94**, 3823–3827.

Keightley, P. D. & Eyre-Walker, A. (1999). Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* **153**, 515–523.

Keightley, P. D. & Bataillon, T. M. (2000). Multigeneration maximum-likelihood analysis applied to mutation–accumulation experiments in *Caenorhabditis elegans*. *Genetics* **154**, 1193–1201.

Keightley, P. D., Davies, E. K., Peters, A. D. & Shaw, R. G. (2000). Properties of ethylmethane sulfonate-induced mutations affecting life-history traits in *Caenorhabditis elegans* and inferences about bivariate distributions of mutation effects. *Genetics* **156**, 143–154.

Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440.

Kondrashov, A. S. (1995). Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *Journal of Theoretical Biology* **175**, 583–594.

Lande, R. (1981). The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**, 541–553.

Lande, R. (1994). Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**, 1460–1469.

Lande, R. (1995). Mutation and conservation. *Conservation Biology* **9**, 782–791.

Lyman, R. F., Lawrence, F., Nuzhdin, S. V. & Mackay, T. F. C. (1996). Effects of single P-element insertions on bristle number and viability in *Drosophila melanogaster*. *Genetics* **143**, 277–292.

Lynch, M. & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer Associates.

Lynch, M., Conery, J. & Borger, R. (1995*a*). Mutational meltdowns in sexual populations. *Evolution* **49**, 1067–1080.

Lynch, M., Conery, J. & Buerger, R. (1995*b*). Mutation accumulation and the extinction of small populations. *American Naturalist* **146**, 489–518.

Lynch, M., Blanchard, J., Houle, D., Kibota, T., Schultz, S., Vassilieva, L. & Willis, J. (1999). Perspective: spontaneous deleterious mutation. *Evolution* **53**, 645–663.

Muller, H. J. (1950). Our load of mutations. *American Journal of Human Genetics* **2**, 111–176.

Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *Computer Journal* **7**, 308–313.

Pletcher, S. D., Houle, D. & Curtsinger, J. W. (1999). The evolution of age-specific mortality rates in *Drosophila melanogaster*: genetic divergence among unselected lines. *Genetics* **153**, 813–823.

Ryan, T. A. & Joiner, B. L. (1976). *Normal probability plots and tests for normality*. Pennsylvania: The Pennsylvania State University.

SAS Institute (1997). The MIXED procedure. In *SAS/STAT® Software: Changes and Enhancements Through Release 6.12*, pp. 571–701. Cary, NC: SAS Institute.

Sulston, J. & Hodgkin, J. (1988). *The Nematode* Caenorhabditis elegans, pp. 587–606. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Turelli, M. (1984). Heritable genetic variation via mutation–selection balance: Lerch's zeta meets the abdominal bristle. *Theoretical Population Biology* **25**, 138–193.

Vassilieva, L. L. & Lynch, M. (1999). The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* **151**, 119–129.

Vassilieva, L. L., Hook, A. M. & Lynch, M. (2000). The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. *Evolution* **54**, 1234–1246.

Wright, S. (1968). *Evolution and the genetics of populations. I. Genetic and biometric foundations*. Chicago, IL: University of Chicago Press.

Zeng, Z. B. (1992). Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. *Genetics* **131**, 987–1001.

Patterns of Evolutionary Constraints in Intronic and Intergenic DNA of *Drosophila*

Daniel L. Halligan,[1] Adam Eyre-Walker,[2] Peter Andolfatto,[1,3] and Peter D. Keightley[1,4]

[1]*University of Edinburgh, School of Biological Sciences, Edinburgh EH9 3JT, UK;* [2]*University of Sussex, Centre for the Study of Evolution and School of Biological Sciences, Brighton BN1 9QG, UK*

We develop methods to infer levels of evolutionary constraints in the genome by comparing rates of nucleotide substitution in noncoding DNA with rates predicted from rates of synonymous site evolution in adjacent genes or other putatively neutrally evolving sites, while accounting for differences in base composition. We apply the methods to estimate levels of constraint in noncoding DNA of *Drosophila*. In introns, constraint (the estimated fraction of mutations that are selectively eliminated) is absolute at the 5′ and 3′ splice junction dinucleotides, and averages 72% in base pairs 3–6 at the 5′-end. Constraint at the 5′ base pairs 3–6 is significantly lower in the lineage leading to *Drosophila melanogaster* than in *Drosophila simulans*, a finding that agrees with other features of genome evolution in *Drosophila* and indicates that the effect of selection on intron function has been weaker in the *melanogaster* lineage. Elsewhere in intron sequences, the rate of nucleotide substitution is significantly higher than at synonymous sites. By using intronic sites outside splice control regions as a putative neutrally evolving standard, constraint in the 500 bp of intergenic DNA upstream and downstream regions of protein-coding genes averages ~44%. Although the estimated level of constraint in intergenic regions close to genes is only about one-half of that of amino acid sites, selection against single-nucleotide mutations in intergenic DNA makes a substantial contribution to the mutation load in *Drosophila*.

[Supplemental material is available online at www.genome.org. The sequence data from this study have been submitted to GenBank under accession nos. AY459538–AY459582.]

Understanding the functional significance of intronic and intergenic noncoding DNA sequences is one of the major challenges in genomics research at present. If functional elements of the genome are close to adaptive optima owing to past directional selection, these sequences are expected to show evidence of purifying selection. This manifests itself as a lower rate of between-species nucleotide substitution when comparisons are made with evolutionary rates in neutrally evolving DNA segments having similar base composition and mutation rates. The level of functional conservation in the genome is important in determining the genome-wide mutation load due to the selective elimination of deleterious mutations (Kondrashov 1995), and this affects several important evolutionary issues (Charlesworth and Charlesworth 1998). Although it is well established that most protein-coding sequences are strongly constrained, that is, that most amino acid altering mutations are deleterious and become selectively eliminated (e.g., Li 1997), functional conservation in noncoding DNA has been much less well studied and is subject to controversy. Although some introns contain regulatory elements, several comparative studies suggest that introns evolve largely free from selective constraints (Gilbert 1978; Li and Graur 1991; Li 1997). However, recent genome-wide interspecific comparisons imply that intron sequences are subject to significant evolutionary pressures (Jareborg et al. 1999; Shabalina and Kondrashov 1999; Bergman and Kreitman 2001). In comparisons involving mammals, the issue of relative rates of substitution is complicated by the presence of methylated CG dinucleotides, which have greatly elevated mutation rates, and whose frequency varies between coding and noncoding DNA and between different categories of noncoding DNA (Chen and Li 2001; Hellmann et al. 2003; Subramanian and Kumar 2003). In intergenic DNA, genome-wide interspecific comparisons (Jareborg et al. 1999; Shabalina and Kondrashov 1999; Bergman and Kreitman 2001; Shabalina et al. 2001), and comparisons of known or putative regulatory elements (Ludwig and Kreitman 1995; Glazko et al. 2003) have also revealed substantial constraints, but the overall level of conservation and the distribution of conserved elements in intergenic regions of the genome is still largely unknown.

Present methods to quantify functional constraints in DNA sequences mostly depend on comparative genomics approaches. They relate to a method for inferring the genome-wide deleterious mutation rate based on sequence divergence (Kondrashov and Crow 1993). Shabalina and Kondrashov (1999) proposed that the proportion of bases that are subject to strong purifying selection can be quantified by comparing the genomes of distantly related species. It is assumed that homologous segments lacking similarity are saturated with nucleotide and/or indel substitutions, and are evolving free from functional constraint, whereas segments showing similarity ("hits") are under strong functional constraint. Constraint is quantified as the fraction of conserved nucleotides in the hits, which is assumed to comprise bases under strong purifying selection. Potential difficulties with the approach are variation across the genome in the mutation rate, which could make nonfunctional elements appear functional (Clark 2001), and obtaining the correct (or most probable) sequence alignment. If the DNA sequence alignment method is heuristic, and, for example, genuine similarities are missed, then functional elements could appear nonfunctional.

A second general approach for quantifying evolutionary constraint also uses comparisons between DNA segments of re-

[3]Present address: Department of Zoology, University of Toronto, Toronto, ON M5S 3G5, Canada.
[4]Corresponding author.
E-MAIL Peter.Keightley@ed.ac.uk; FAX 44 131 650 5443.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.1329204.

lated species, but uses sequences from species showing lower levels of divergence that are far from saturation. It is based on comparing the rate of evolution of a putative functional segment of noncoding DNA with the rate of evolution of a DNA segment or a category of nucleotide sites that is assumed to be evolving free from constraint (a neutral segment), that has the same mutation rate, and can therefore act as a standard. Constraint is the factor by which evolution is slowed down in the functional segment (Kimura 1983). Nucleotides are assumed to fall into two classes in the functional sequence: neutral, which evolve at the same rate as the neutral sequence; or strongly constrained, in which mutations are eliminated unconditionally by natural selection. The neutral segment should be adjacent to the functional segment, thereby making the assumption of equality of mutation rates defensible. A close to ideal situation would be to compare the rate of evolution of a pseudogene (assumed to completely lack function) to that of an adjoining noncoding DNA segment. Unfortunately, because in many taxa, including *Drosophila*, pseudogenes are uncommon (Petrov et al. 1996), an alternative category of neutrally evolving sequences is needed. A candidate for such a category is synonymous sites of genes, because changes in these do not lead to change in the amino acid sequence. In many taxa, including *Drosophila*, however, there is evidence of past selection acting on synonymous codon usage (Shields et al. 1988), and this could retard rates of evolution at synonymous sites (Li 1997, Chapter 7).

In this paper, our initial approach is to use synonymous sites of *Drosophila* protein-coding genes as a standard for estimation of constraint in adjacent noncoding DNA. Features of the dynamics of synonymous site evolution indicate that selection at synonymous sites is weak or, in some cases, absent in *Drosophila*. In the lineage leading to *Drosophila melanogaster* from its common ancestor with *Drosophila simulans*, there has been a surge in the rate of preferred to unpreferred synonymous substitutions (Akashi 1995, 1996; McVean and Vieira 2001). This surge in the rate of substitution indicates a genome-wide relaxation of selection at synonymous sites, possibly because of demographic changes that have changed the efficiency of natural selection. In *D. melanogaster*, a population genetics analysis of the pattern of synonymous site divergence indicates that selection has been relaxed to the point of being completely absent (McVean and Vieira 2001). Further evidence for low levels of selection presently acting on *D. melanogaster* synonymous codon usage comes from an analysis of the frequency spectrum of segregating synonymous sites (Akashi 1999). A weakening of selection to approximately one-half of that in the ancestral species is estimated to have occurred in the *D. simulans* lineage (McVean and Vieira 2001). Furthermore, weak selection of the magnitude thought to be acting on synonymous codon usage in *Drosophila* (Akashi 1995, 1996) is predicted to have only a small effect on substitution rates (Eyre-Walker and Bulmer 1995). Recently, an apparent excess of preferred to unpreferred synonymous site substitutions has been reported in the *Xdh* gene of several *Drosophila* species (Begun and Whitley 2002). Possible explanations for this observation are an evolutionary shift in base composition towards A/T nucleotides in many *Drosophila* lineages (the explanation favored by Begun and Whitley 2002; see also Duret et al. 2002), a general weakening of selection in *Drosophila* lineages, or an artifact of parsimony if nucleotide mutation rates are sufficiently variable.

There are two additional difficulties in interpreting comparisons between synonymous site or other putatively neutral site divergence and nucleotide divergence in noncoding DNA. First, differences in the rate of substitution can be induced by differences in base composition; this stems from variation in average mutation rates between different kinds of nucleotides. We address this by comparing expected and observed numbers of substitutions; expected numbers in a noncoding segment are predicted on the basis of substitution rates at synonymous or other putatively neutral sites of adjacent genes, after the compositional effect has been accounted for.

A second potential problem in analyzing evolutionary rates in noncoding DNA concerns inference of the correct sequence alignment. Consider two alternative plausible alignments of a pair of sequences containing at least one gap:

Alignment 1	Alignment 2
Three substitutions	One substitution
ATGCATGCG	ATGCATGCG
AT--CAGCA	AT-CA-GCA

If alignment 1 were taken as the true alignment, the fraction of nucleotide differences ($k = 3/7$) would be radically different from taking alignment 2 ($k = 1/7$). The uncertainty is due to the unknown pattern of indels (gaps) between the sequences. In general, putative alignments containing too many gaps relative to the true alignment tend to have too few nucleotide substitutions or vice versa, and the bias can be serious. A solution to this problem has been proposed by Thorne et al. (1991, 1992), who developed an algorithm to compute probabilities of alternative alignments according to explicit models of indel evolution. Here, we use a Monte Carlo approach, MCALIGN, to tackle the problem of aligning noncoding DNA (P.D. Keightley and T. Johnson, unpubl.). Noncoding DNA sequences are aligned according to a model of indel evolution that is parameterized by relative rates of indels and nucleotide substitutions in noncoding DNA of closely related *Drosophila* species, and the most probable alignment is used in the subsequent analysis.

RESULTS

Indel Frequencies in Noncoding DNA Between *D. simulans* and *D. sechellia* and Parameterization of Alignment Models

To investigate the frequency distribution of indels, and to parameterize models of indel evolution suitable for aligning *Drosophila* noncoding DNA, we compiled intronic and intergenic DNA sequences from homologous loci of *D. simulans* and *D. sechellia* (Table 1). These species were chosen because they are part of the *melanogaster* subgroup, and are sufficiently closely related as to make alignment of noncoding DNA by standard heuristic methods virtually unambiguous. The frequency distribution of indel length in three DNA categories is shown in Figure 1 (two long intronic indels of length 29 and 37 are omitted to aid clarity). Distributions of indel length are not dissimilar to geometric distributions, as has been suggested previously (Gu and Li 1995). Numbers of substitutions do not differ significantly between DNA categories: A likelihood ratio test for heterogeneity among indel rates relative to substitution rates is nonsignificant (Table 1;

Table 1. Rates of Nucleotide Substitution (k) and Relative Rates of Indels (θ) in Noncoding DNA of *D. simulans* and *D. sechellia*

DNA category	No. of loci	Total no. bp	Substitutions	k	Indels	θ
Intronic	24	6302	193	0.0306	44	0.228
5′ intergenic	15	3094	85	0.0275	9	0.106
3′ intergenic	14	3159	101	0.0320	18	0.178

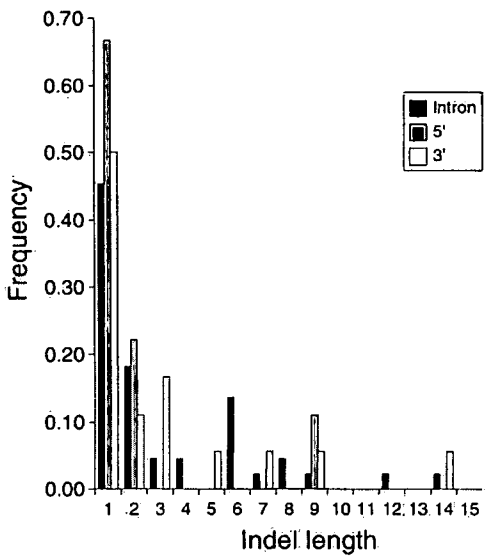


Figure 1 Frequency distribution of indel length in intronic and intergenic segments upstream (5') and downstream (3') from the start or stop codon, in DNA sequences of *Drosophila sechellia* and *Drosphila simulans*. Two intronic indels of length 29 and 37 have been omitted to aid clarity.

2 ln likelihood ratio $\approx \chi^2_2 = 4.5$; $P = 0.11$), although there is a suggestion that the number of indels in 5'-intergenic regions is lower in relation to the number of substitutions than in intronic DNA (Table 1; 2 ln likelihood ratio $\approx \chi^2_1 = 4.4$; $P = 0.04$, uncorrected for multiple tests).

Evolutionary Conservation in Intronic DNA Sequences of *Drosophila*

We computed estimates of the level of constraint in *D. melanogaster/simulans* intron sequences by the two lineage approach, using synonymous sites as the putatively neutral standard, as described in Methods, under the assumption that the equilibrium G + C content (f_e) is equal to the G + C content of intronic sequences in our data set (0.37). Separate estimates were made for complete intron sequences, and for intron sequences stripped of putative 5'- and 3'-splice control sequences (Table 2). Estimates of constraint in intron sequences are negative in sequences either including ($P = 0.1$) or excluding ($P = 0.007$) splice control sequences. Negative estimates of constraint imply that fourfold sites evolve more slowly than intronic sites (particularly those sites that are outside splice control regions), after differences in base composition have been accounted for. We investigated the slightly higher constraint in sequences including putative splice control sequences by calculating constraints for groups of bases at the 5'- and 3'-ends of intron sequences (Table 3). In the sequences analyzed, conservation is absolute for the 5'-splice-junction GT dinucleotide, and constraint is also strong for base pairs 3–6 at the 5'-end; all these bases are important in delimiting the 5'-end of introns (Sharp 1994). With the exception of the invariant 3'-splice-junction AT dinucleotide, constraint is nonsignificant at the 3'-end, a somewhat surprising result, considering that the consensus for nucleotides 5–16 is a run of pyrimidines showing moderate conservation across eukaryotic lineages (Sharp 1994). Constraint in intronic splice sequences, calculated using non-splice-control intronic sequences as the putatively neutral standard, gives similar results to those shown in Table 3 (i.e., 5' bp 3–6: $C = 0.715$, SEM = 0.0484; 3' bp 3–16: $C = 0.0517$, SEM = 0.0746).

We also calculated constraint specific to the *D. melanogaster* and *D. simulans* lineages by the three-lineage approach described in Methods, using *Drosophila yakuba* sequences as the outgroup. The results are consistent for those obtained for the two-sequence method, and suggest weak constraint in intron sequences lacking putative splice sites in both species. The level of constraint is significantly lower for bases 3–6 of the 5'-end in *melanogaster* than *simulans* ($P = 0.048$; bootstrap analysis, Table 4), lending support to the idea of a lower intensity of selection in the lineage leading to *D. melanogaster*. The constraint difference at the 3'-end (Table 4) is nonsignificant ($P > 0.8$).

Evolutionary Conservation of Intergenic DNA Sequences of *Drosophila*

We computed constraint by the two-lineage method in 5'- and 3'-intergenic DNA sequences of *D. melanogaster* and *D. simulans* using synonymous sites as the putatively neutral standard (Table 2). The estimated levels of constraint contrast with the results for intronic DNA: There is moderate to strong positive constraint in much of the 1000 bp of intergenic DNA analyzed, implying the action of purifying selection. The average levels of constraint in segments of up to 500 bp upstream and downstream of genes are 0.174 (SEM = 0.058) and 0.256 (SEM = 0.135), respectively. The corresponding values for constraint computed using intronic nucleotides excluding splice sites as the putatively neutral standard are 0.373 (SEM = 0.078) and 0.522 (SEM = 0.082). These

Table 2. Observed and Expected Numbers of Nucleotide Substitutions Along With Estimates for Constraint in Noncoding DNA Sequences of Three Categories, Calculated by the Two Lineage Approach Using Four-Fold Sites of Homologous Genes from *D. simulans* and *D. melanogaster* as a Standard

DNA category	Data set	Number of loci	Base pairs per locus	Substitutions per locus: Observed (SEM)	Substitutions per locus: Expected (SEM)	Constraint (SEM)
Intronic	Complete	91	228	16.23 (1.69)	14.86 (1.56)	−0.094 (0.059)
Intronic	Splice sequences omitted	91	190	13.72 (1.63)	11.75 (1.45)	−0.17 (0.069)
5' intergenic	bp 1–100	77	99	5.43 (0.36)	8.29 (0.46)	0.34 (0.053)
	bp 101–200	73	98	5.58 (0.45)	7.97 (0.46)	0.30 (0.056)
	bp 201–300	69	98	6.51 (0.53)	7.82 (0.49)	0.16 (0.079)
	bp 301–500	66	188	13.37 (1.06)	13.86 (0.86)	0.03 (0.088)
3' intergenic	bp 1–100	42	95	5.78 (0.69)	7.29 (0.54)	0.21 (0.086)
	bp 101–200	31	88	5.35 (0.77)	6.58 (0.83)	0.17 (0.14)
	bp 201–300	21	92	5.91 (0.79)	6.65 (1.06)	0.093 (0.17)
	bp 301–500	18	150	7.10 (1.30)	11.60 (2.13)	0.36 (0.19)

Table 3. Estimates of Constraint in Runs of Nucleotides Close to the 5′- or 3′-Ends of Introns, Calculated Using the Two Lineage Method, Applied to Sequences from *D. melanogaster* and *D. simulans*

Intron end	Base pairs	Substitutions: Observed (SEM)	Substitutions: Expected (SEM)	Constraint (SEM)
5′	1–2	0 (0.0)	0.294 (0.021)	1.00 (0.00)
	3–4	0.145 (0.037)	0.277 (0.021)	0.47 (0.13)
	5–6	0.055 (0.024)	0.295 (0.022)	0.82 (0.077)
3′	1–2	0 (0.0)	0.292 (0.021)	1.00 (0.00)
	3–16	2.248 (0.202)	1.912 (0.142)	−0.178 (0.098)

findings agree with Duret and Mouchirod (2000), who found a negative correlation between gene expression level and substitution rate in 5′- and 3′-untranslated regions of mammals, indicating the operation of purifying selection.

DISCUSSION

In contrast to intron sequences in mammals and several other taxa (International Human Genome Sequencing Consortium 2001; Mouse Genome Sequencing Consortium 2002), intron sequences tend to be rather short in *Drosophila*, with a peak length of only ~60 bp (Adams et al. 2000), and several studies have revealed precursor mRNA secondary structure in intronic sequences (Stephan and Kirby 1993; Kirby et al. 1995; Leicht et al. 1995). We therefore expected that constraints would be easily observed in *Drosophila*, if introns commonly contain gene expression control sequences.

The analysis did not bear this expectation out. The results are consistent with somewhat faster evolution at most intronic sites than fourfold sites, which themselves are thought to be under weak selection. Under the assumption of a nonequilibrium model of sequence evolution, our analysis indicates that intronic sequences outside splice control sequences evolve ~17% faster than fourfold sites of adjacent genes. Similar findings have recently been reported in primates (Chen and Li 2001; Hellmann et al. 2003) and rodents (Keightley and Gaffney 2003), although a different study in primates did not reveal the pattern (Subramanian and Kumar 2003). We examined the robustness of the result by changing f_e, the equilibrium GC content. Values of f_e below 0.53 give higher rates of evolution at intronic sites than fourfold sites, and for values below 0.41 the difference in rates (as measured by constraint) is significant at $P < 0.05$. In introns, moderate to strong constraint was only detected between *melanogaster* and *simulans* at the dinucleotides at exon–intron boundaries and at 5′ nucleotides 3–6; these latter nucleotides also show a notable difference in constraint between the two lineages, possibly brought about by a difference in the long-term effectiveness of selection between the species (Aquadro et al. 1988; Akashi 1995; Moriyama and Powell 1996; Andolfatto 2001; Eyre-Walker et al. 2002). This pattern of constraint close to intron boundaries implies that mutations at these sites are slightly deleterious (Ohta 1992), and is therefore indirect evidence that the remaining sequences are genuinely evolving free from selective constraints. There is little difference between expected and observed numbers of substitutions between *melanogaster* and *simulans* (Table 4), an observation that is also consistent with models of weak selection, because selection coefficients of the order of the reciprocal of N_e are predicted to have little influence on substitution rates. The present findings are in broad agreement with McVean and Vieira (2001), who found that predicted rates of substitution were similar to observed rates in *Drosophila*. Our results concord with observations of the density of nucleotide polymorphisms in human introns as a function of distance from the 5′- or 3′-end (F.A. Kondrashov pers. comm.); there was no evidence for selection operating beyond about nucleotide 10 from the 5′- or 3′-end.

Our results contrast with recent estimates of the levels of constraint in introns and intergenic DNA in *Drosophila* (Bergman and Kreitman 2001) and *Caenorhabditis* (Shabalina and Kondrashov 1999), in which constraint was calculated from the fraction of conserved nucleotides in alignable blocks of DNA between distantly related species. Surprisingly, frequencies of conserved blocks in introns and intergenic DNA were similar to each other (of the order of 20%). However, variability in the mutation rate from region to region (Clark 2001) could give the false impression of evolutionary conservation in a segment that is evolving at the neutral rate. Furthermore, alignment of noncoding DNA is problematical with widely diverged species. Any noncoding DNA alignment that is not based on a model of indel evolution is likely to be biased (Thorne et al. 1991), and estimates of numbers of nucleotide substitutions may either be too high or too low depending on whether the alignment algorithm inserts too few or too many indels. Estimates of the proportion of conserved blocks in noncoding regions between mouse and human (Jareborg et al. 1999) are also likely to be susceptible to such biases.

The data in Table 3 indicate that the number of constrained nucleotides per intron is ~4.1. If there are 41,000 introns in the *Drosophila* genome (Adams et al. 2000), the predicted number of constrained nucleotides in introns is therefore only 0.17 Mb. The level of constraint at amino acid sites of *Drosophila* genes has been estimated to be ~84% (Eyre-Walker et al. 2002), implying that the total number of constrained amino acid sites in the *Drosophila* genome is ~16 Mb (~14,000 protein-coding genes, comprising an average of 591 codons [Adams et al. 2000], and about three-quarters of sites in coding DNA lead to an amino acid change if mutated). The number of constrained nucleotides in introns is therefore relatively small in relation to the protein-coding segment of the genome. However, the number of constrained nucleotides in intergenic DNA could potentially be of the same order as in coding DNA. For example, if we assume the average constraint values calculated relative to intronic sequences, we

Table 4. Estimates of the Level of Constraint in Introns of 38 Loci From *D. melanogaster* and *D. simulans*, Computed Using the Three-Lineage Method

Lineage	Data set	bp Per locus	Substitutions: Observed (SEM)	Substitutions: Expected (SEM)	Constraint (SEM)
melanogaster	Complete	204	7.05 (1.07)	7.17 (1.15)	0.010 (0.11)
simulans			6.49 (1.04)	7.82 (1.59)	0.15 (0.14)
melanogaster	Splice sequences omitted	161	5.87 (0.99)	5.45 (0.96)	−0.087 (0.13)
simulans			5.39 (0.95)	5.90 (1.36)	0.065 (0.15)
melanogaster	5′ bases 3–6	7.8	0.184 (0.074)	0.326 (0.045)	0.44 (0.21)
simulans			0.052 (0.036)	0.370 (0.057)	0.86 (0.10)
melanogaster	3′ bases 3–16	27.3	1.003 (0.200)	1.112 (0.196)	0.084 (0.19)
simulans			1.034 (0.200)	1.205 (0.201)	0.12 (0.21)

obtain 14,000 genes × 1000 bp × 0.44 = 6.2 Mb. This is a minimum estimate, whose value could be much larger if there are appreciable functional constraints deep in the intergenic DNA.

METHODS

Data

Homologous gene sequences (partial or complete) from *D. simulans* and *D. melanogaster*, and, where available *D. yakuba*, were compiled from GenBank. Genes were selected if they contained at least one intron, or at least 60 bp of intergenic DNA upstream or downstream for the start or stop codon. Coding sequences were aligned using CLUSTAL (Thompson et al. 1994) and corrected manually. Noncoding sequences were aligned using MCALIGN (P.D. Keightley and T. Johnson, unpubl.), a procedure that attempts to find the most probable alignment according to specific models of indel evolution (see below). The parameters of the alignment model were derived from data on relative rates of indels and nucleotide substitutions between *D. simulans* and *D. sechellia* (Table 1), and the frequency distribution of indel lengths (Fig. 1). Noncoding DNA alignments of *D. simulans* and *D. sechellia* are virtually unambiguous, by alignment with any standard heuristic alignment method. The most probable alignment of the *D. simulans/melanogaster/yakuba* sequences were used in subsequent analysis. Intergenic DNA was categorized either as 5′ or 3′. In cases in which genes are so close together in the genome that this categorization was ambiguous, stretches of DNA were arbitrarily assigned to the 5′ category, although they could have been considered to belong to the 3′-segment of an adjacent gene. Intergenic DNA includes any DNA that is 5′ or 3′ from the start or stop codon, and therefore contains elements of transcribed untranslated DNA.

Introns were either analyzed as complete sequences or partial sequences after removal of putative splice control sequences. The base pairs removed were 1–6 at the 5′-end and 1–16 at the 3′-end. The exact limits of the control sequences are somewhat arbitrary (Sharp 1994).

Lists of loci are shown in Supplemental Tables 1 and 2, and aligned sequences are available from PDK's Web site.

Sequencing of Additional *Drosophila simulans* Intergenic Sequences

We obtained additional intergenic DNA sequences from *Drosophila simulans* by sequencing the 5′-flanking regions of genes for which the orthologous coding sequences were available for both *simulans* and *melanogaster* on GenBank. Genes for which there was only a short length of available coding sequence in *simulans* were excluded (we used an arbitrary cutoff of 200 bp), and we did not sequence upstream DNA from previously sequenced *simulans* loci. Primers for sequencing were designed (using Primer Premier 5.00; Premier Biosoft International) to be ~650 to 700 bp apart, based on the *melanogaster* sequence. Upstream primers were usually designed from the noncoding *melanogaster* sequence (where possible an upstream coding sequence was used), and downstream primers were designed using the *simulans* coding sequence.

Genomic DNA for PCR reactions was prepared (Gentra Systems, Research Triangle Park) from a single partially inbred male *Drosophila simulans* fly collected in Aswan, Egypt in 2001. A single male fly was used as a source of DNA in all cases to reduce sequencing problems associated with heterogeneity in template DNA. A combination of standard PCR and asymmetric PCR (Miller et al. 2003) was used to amplify the appropriate section of DNA. If the primers failed to amplify the appropriate section of DNA, the primers were redesigned. If the appropriate DNA segment still could not be amplified after the primers had been redesigned three times, investigation of the gene was terminated. In 18 cases out of 63, we could not obtain sufficient amplification of the appropriate section of DNA.

Purified PCR products were sequenced on both strands using an ABI prism BigDye terminator cycle sequencing kit (Applied Biosystems) and run on an Applied Biosystems 3730 DNA Analyzer (Applied Biosystems). Sequences from each strand for each gene were then assembled using Sequencher 3.0 software (Gene Codes), and alignments were checked manually. The GenBank accession numbers are AY459538–AY459582.

Likelihood Ratio Test for Variation in Rates of Indels Relative to Nucleotide Substitutions

The test was constructed under the assumption that sequences are sufficiently closely related such that multiple hits can be ignored, and that the number of indels is linearly related to the number of nucleotide substitutions. Assume that there are n categories of noncoding DNA (say, $n = 3$ with 1 = intronic, 2 = 5′-intergenic, and 3 = 3′-intergenic). Let k_i be a parameter for the fraction of nucleotide differences between sequences of category i, and $\theta_i k_i$ be a compound parameter for the fraction of indels differentiating sequences in category i. Under the assumption of independent binomially distributed nucleotide substitution and indel numbers, the likelihood of observing n_i substitutions and g_i indels is

$$L_i = k_i^{n_i}(1-k_i)^{l_i-n_i}(\theta_i k_i)^{g_i}(1-\theta_i k_i)^{m_i}, \quad (1)$$

where l_i is the number of base pairs in the sequence (excluding indels) and m_i is the number of sites not occupied by an indel. The likelihood of the observations of three categories of DNA is $L = L_1 \times L_2 \times L_3$, and the two models are compared according to $\theta_1 \neq \theta_2 \neq \theta_3$ (full model), and $\theta_1 = \theta_2 = \theta_3$ (reduced model). The likelihood with respect to k and θ was maximized numerically.

Alignment of Noncoding DNA Sequences According to Models of Indel Evolution

Alignment was carried out by a procedure MCALIGN, available at PDK's Web pages. The procedure uses a Monte Carlo algorithm to search for the most probable alignment of a pair of sequences or of three sequences that includes an outgroup, based on a model of indel evolution. The parameters of the model are θ, the rate of indels relative to nucleotide substitutions, and a vector parameter w specifying the frequency distribution of indel lengths. Because θ is a parameter of the model, estimated alignments containing large (small) numbers of nucleotide differences tend to have large (small) numbers of indels, a pattern supported by mouse–human sequence alignments (Hardison et al. 2003). In aligning pairs of sequences, the Jukes-Cantor method is used to correct for multiple nucleotide substitutions. For three sequences, parsimony is used to assign substitutions and indel events to the ingroup or the outgroup, and the probability of the alignment is the product of probabilities for the ingroup and outgroup.

The model parameters θ and w come from external data, in the present case from alignments of noncoding DNA of *D. simulans* and *D. sechellia*. Values of θ from Table 1 were used to parameterize three alternative models, for aligning intronic, 3′-intergenic, or 5′-intergenic DNA. The vector parameter w was assumed to be the same for each model, and was derived from the frequency distribution of indels in introns, after some smoothing of the distribution had been applied.

Two-Lineage Approach to Compute Constraint in Noncoding DNA

Following distance-based methods for calculating constraint in coding DNA (Eyre-Walker and Keightley 1999), the present method uses rates of substitution at fourfold sites or other putative neutral sites of a gene to predict expected numbers of substitutions in an adjacent noncoding DNA segment, such as an intron or flanking sequence, assuming equal rates of mutation in the sequences. The method takes into account differences in base composition. The expected numbers of substitutions (E) are compared with the observed numbers (O) to calculate constraint (C). For example, if $E = O$, the constraint in the noncoding segment is

zero; if $O = 0$, constraint takes the value of 1. The method is only applicable to closely related species for which multiple hits can be safely ignored.

In a pairwise comparison, it is not possible to determine the direction of a particular substitution (i.e., whether a C↔T difference is caused by a C→T or a T→C substitution). However, it is possible to infer the proportion of changes that are in a particular direction if we assume or know the equilibrium base composition. Let us group Gs and Cs together, and As and Ts together. Let f_e be the equilibrium G+C content of the sequence; this is the G+C content that the sequence will eventually reach, and let z be a mutation rate parameter such that the rate at which A or T sites change to G or C is zf_e, and the rate at which G or C sites change to A or T is $z(1 - f_e)$. We can then use the present and equilibrium G+C content to infer the proportion of observed AT↔GC differences that go in a particular direction (this category of differences involves the following pairwise differences: A↔G, A↔C, T↔G, and T↔C). However, with only two species, we cannot infer whether an observed G↔C difference is caused by a G-to-C mutation or C-to-G mutation (this would require a parsimony approach). Similarly, we cannot assign polarity to any observed A↔T differences. We can therefore only calculate four different rates ($i = 1 \ldots 4$), two pairwise rates (A↔T and G↔C) and two directional rates (AT→GC and GC→AT). If we consider evolution over a fairly short period of time so that the G+C content does not change dramatically (or not at all if the sequence is at equilibrium), then the numbers (X) of AT→GC mutations and GC→AT mutations are

$$X_{(AT\rightarrow GC)} = f_e zN(1 - f_a) \quad (2)$$

$$X_{(GC\rightarrow AT)} = (1 - f_e)zNf_a \quad (3)$$

where f_a is the G+C content of the sequence being considered at the separation of the two species being considered (in practical terms we can assume that this is equal to the present G+C content if the time of divergence is small). Under these assumptions, an equilibrium is reached when the number of mutations in one direction equals the number of mutations in the reverse direction, that is, when $f_e zN(1 - f_a) = (1 - f_e)zNf_a$, or when $f_a = f_e$.

The total number of AT↔GC differences that have occurred ($X_{AT\leftrightarrow GC}$) can be written as the sum of equations 2 and 3, and rearranging this for z:

$$z = \frac{X_{(AT\leftrightarrow GC)}}{((1 - f_e)f_a + f_e(1 - f_a))N} \quad (4)$$

By substituting equation 4 into equations 2 and 3, we can remove the unknown parameter z and express the two estimates of the number of directional mutations in terms of f_e, f_a, and the number of pairwise AT↔GC differences. If we then divide by the number of sites at which each of these types of mutation could occur [$N(1 - f_a)$ and Nf_a, respectively], we can obtain an estimate of the per site rate of AT→GC and GC→AT mutations.

$$K_{(AT\rightarrow GC)} = \frac{X_{(AT\leftrightarrow GC)}f_e}{((1 - f_e)f_a + f_e(1 - f_a))N} \quad (5)$$

$$K_{(GC\rightarrow AT)} = \frac{X_{(AT\leftrightarrow GC)}(1 - f_e)}{((1 - f_e)f_a + f_e(1 - f_a))N} \quad (6)$$

The predicted (expected) number of substitutions in the noncoding DNA segment is,

$$E = \sum_{i=1}^{4} K_i M_i, \quad (7)$$

where M_i is the number of sites in the noncoding segment corresponding to rates of type i. The observed number of differences in the segment, O, is the number of nucleotide differences in the noncoding segment. Constraint for a segment is given by $C = 1 - O/E$, or, for several segments it is

$$C = 1 - \sum O_j / E_j, \quad (8)$$

where the summation is carried out over segments. Standard errors of O, E, and C are calculated by bootstrapping the data, by gene (Eyre-Walker and Keightley 1999).

Three-Lineage Approach to Compute Constraint in Noncoding DNA

The basis of the approach is to calculate rates for all possible kinds of nucleotide substitutions at fourfold and twofold degenerate synonymous sites of a gene, and to use these rates to calculate expected numbers of substitutions in an adjacent noncoding DNA segment. The numbers of substitutions at synonymous and noncoding sites are estimated using parsimony. The rates for the 12 possible kinds of synonymous substitution in one of the branches, $K_{A\rightarrow T}$, $K_{A\rightarrow C}$, $K_{A\rightarrow G}$, $K_{T\rightarrow A}$, and so on, are computed by taking weighted averages of the fraction of differences at fourfold, and, where applicable, twofold degenerate sites. Under the assumption of neutral evolution at synonymous sites and equal mutation rates in the coding and noncoding DNA, the expected number of substitutions in the noncoding segment associated with gene i is, therefore,

$$E_i = N_A(K_{A\rightarrow T} + K_{A\rightarrow C} + K_{A\rightarrow G}) + N_T(K_{T\rightarrow A} + \ldots) + N_C(K_{C\rightarrow A} + \ldots) + N_G(K_{G\rightarrow A} + \ldots), \quad (9)$$

where N_A, N_T, N_C, and N_G, are the numbers of A, T, C, and G nucleotides, respectively, in the noncoding sequence. In cases in which all three base pairs differ, averages of $K_{A\rightarrow T} + K_{A\rightarrow C}$, and so on, are calculated, weighted by the probabilities of alternative ancestral states, on the assumption that the relative lengths of the branch from the *melanogaster/simulans* common ancestor (a) to *simulans* and *melanogaster* are 0.2, and the relative length of the branch from a to *yakuba* is 0.6. This model gives probabilities for the ancestral state being the *melanogaster* or *simulans* base of 0.462, and for the *yakuba* base of 0.0769. O_i is the observed number of substitutions in the aligned noncoding sequence associated with gene i. The average constraint is calculated by equation 8.

ACKNOWLEDGMENTS

We thank Brian Charlesworth, Toby Johnson, Penny Haddrill, and two reviewers for helpful comments and suggestions.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287:** 2185–2195.

Akashi, H. 1995. Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA. *Genetics* **139:** 1067–1076.

———. 1996. Molecular evolution between *Drosophila melanogaster* and *D. simulans*: Reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* **144:** 1297–1307.

———. 1999. Inferring the fitness effects of DNA mutations from polymorphism and divergence data: Statistical power to detect directional selection under stationarity and free recombination. *Genetics* **151:** 221–238.

Andolfatto, P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18:** 279–290.

Aquadro, C.F., Lado, K.M., and Noon, W.A. 1988. The *rosy* region of *Drosophila melanogaster* and *Drosophila simulans* .1. contrasting levels of naturally-occurring DNA restriction map variation and divergence. *Genetics* **119:** 875–888.

Begun, D.J. and Whitley, P. 2002. Molecular population genetics of *Xdh* and the evolution of base composition in *Drosophila*. *Genetics* **162:** 1725–1735.

Bergman, C.M. and Kreitman, M. 2001. Analysis of conserved noncoding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res.* **11:** 1335–1345.

Charlesworth, B. and Charlesworth, D. 1998. Some evolutionary consequences of deleterious mutations. *Genetica* **103:** 3–19.
Chen, F.-C. and Li, W.-H. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees.*Am. J. Hum. Genet.* **68:** 444–456.
Clark, A.G. 2001. The search for meaning in noncoding DNA. *Genome Res.* **11:** 1319–1320.
Duret, L. and Mouchiroud, D. 2000. Determinants of substitution rates in mammalian genes: Expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.* **17:** 68–74.
Duret, L., Semon, M., Piganeau, G., Mouchiroud, D., and Galtier, N. 2002. Vanishing GC-rich isochores in mammalian genomes. *Genetics* **162:** 1837–1847.
Eyre-Walker, A. and Bulmer, M. 1995 . Synonymous substitution rates in enterobacteria. *Genetics* **140:** 1407–1412.
Eyre-Walker, A. and Keightley, P.D. 1999. High genomic deleterious mutation rates in hominids. *Nature* **397:** 344–347.
Eyre-Walker, A., Keightley, P.D., Smith, N.G.C., and Gaffney, D. 2002. Quantifying the slightly deleterious model of molecular evolution. *Mol. Biol. Evol.* **19:** 2142–2149.
Gilbert, W. 1978. Why genes in pieces. *Nature* **271:** 501.
Glazko, G.V., Koonin, E.V., Rogozin, I.B., and Shabalina, S.A. 2003. A significant fraction of conserved noncoding DNA in human and mouse consists of predicted matrix attachment regions. *Trends Genet.* **19:** 119–124.
Gu, X. and Li, W-H. 1995. The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. *J. Mol. Evol.* **40:** 464–473.
Hardison, R.C., Roskin, K.M., Yang, S., Diekhans, M., Kent, W.J., Weber, R., Elnitski, L., Li, J., O'Connor, M., Kolbe, D., et al. 2003. Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res.* **13:** 13–26.
Hellmann, I., Zollner, S., Enard W., Ebersberger, I., Nickel, B., and Pääbo, S. 2003. Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res.* **13:** 831–837.
International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.
Jareborg, N., Birney, E., and Durbin, R. 1999. Comparative analysis of noncoding regions of 77 orthologous mouse and human gene pairs. *Genome Res.* **9:** 815–824.
Keightley, P.D. and Gaffney, D.J. 2003. Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents. *Proc. Natl. Acad. Sci.* **100:** 13402–13406.
Kimura, M. 1983. *The neutral theory of molecular evolution.* Cambridge University Press, Cambridge, UK.
Kirby, D.A., Muse, S.V., and Stephan, W. 1995. Maintenance of pre-mRNA secondary structure by epistatic selection. *Proc. Natl. Acad. Sci.* **92:** 9047–9051.
Kondrashov, A.S. 1995. Contamination of the genome by very slightly deleterious mutations. Why have we not died 100 times over? *J. Theor. Biol.* **175:** 583–594.
Kondrashov, A.S. and Crow, J.F. 1993. A molecular approach to estimating the human deleterious mutation rate. *Hum. Mutat.* **2:** 229–234.
Leicht, B.G., Muse, S.V., Hanczyc, M., and Clark, A.G. 1995. Constraints on intron evolution in the gene encoding the myosin alkali light-chain in *Drosophila*. *Genetics* **139:** 299–308.
Li, W.-H. 1997. *Molecular evolution.* Sinauer, Sunderland, MA.
Li, W.-H. and Graur, D. 1991. *Fundamentals of molecular evolution.* Sinauer, Sunderland, MA.
Ludwig, M.Z. and Kreitman, M. 1995. Evolutionary dynamics of the enhancer region of *even-skipped* in *Drosophila*. *Mol. Biol. Evol.* **12:** 1002–1011.
McVean, G.A.T. and Vieira, J. 2001. Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. *Genetics* **157:** 245–257.
Miller, M.D., Duan, S., Lovins, E.G., Kloss, E.F., and Kwok, P. 2003. Efficient high-throughput resequencing of genomic DNA. *Genome Res.* **13:** 717–720.
Moriyama, E.N. and Powell, J.R. 1996. Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13:** 261–277.
Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520–562.
Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Ann. Rev. Ecol. Syst.* **23:** 263–286.
Petrov, D.A., Lozovskaya, E.R., and Hartl, D.L. 1996. High intrinsic rate of DNA loss in *Drosophila*. *Nature* **384:** 346–349.
Shabalina, S.A. and Kondrashov, A.S. 1999. Pattern of selective constraint in *C. elegans* and *C. briggsae* genomes. *Genet. Res.* **74:** 23–30.
Shabalina, S.A., Ogurtsov, A.Y., Kondrashov, V.A., and Kondrashov, A.S. 2001. Selective constraint in intergenic regions of human and mouse genomes. *Trends Genet.* **17:** 373–376.
Sharp, P.A. 1994. Split genes and RNA splicing. *Cell* **77:** 805–815.
Shields, D.C., Sharp, P.M., Higgins, D.G., and Wright, F. 1988. Silent sites in *Drosophila* genes are not neutral—Evidence of selection among synonymous codons. *Mol. Biol. Evol.* **5:** 704–716.
Stephan, W. and Kirby, D.A. 1993. RNA folding in *Drosophila* shows a distance effect for compensatory fitness interactions. *Genetics* **135:** 97–103.
Subramanian, S. and Kumar, S. 2003. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res.* **13:** 838–844.
Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W—Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.
Thorne, J.L., Kishino, H., and Felsenstein, J. 1991. An evolutionary model for maximum likelihood alignment of DNA sequences. *J. Mol. Evol.* **33:** 114–124.
———. 1992. Inching toward reality—An improved likelihood model of sequence evolution. *J. Mol. Evol.* **34:** 3–16.

WEB SITE REFERENCES

http://homepages.ed.ac.uk/eang33/; executables, source code, and user instructions for MCALIGN.
http://homepages.ed.ac.uk/eang33/; aligned sequence data files.

Received March 11, 2003; accepted in revised form December 1, 2003.