NOTE TO USERS

The Effects of Ultraviolet-B Radiation on Mutational Parameters in *Arabidopsis thaliana*

Joanna Leigh MacKenzie

Department of Biology, McGill University, Montreal

December, 2004

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master's of Science.

Your file Votre référence
ISBN: 0-494-12494-6
Our file Notre référence
ISBN: 0-494-12494-6

Abstract

This project was designed to investigate the impact of natural levels of ultraviolet-B radiation on the genomic mutation rate in *Arabidopsis thaliana*. UV-B radiation is a known mutagen, but plants may have evolved mechanisms to cope with any genomic damage induced by routine exposure to this radiation. In an attempt to determine whether the genomic mutation rate in a plant species is elevated in the presence of UV-B, two eleven generation mutation accumulation studies were preformed. One study incorporated levels of UV-B similar to that encountered on a clear mid-summer's day, while the other was performed in the absence of this mutagen. Mutation rate estimates, obtained primarily from maximum likelihood analysis of phenotypic data, were not significantly greater than zero, both in the presence and absence of UV-B. No evidence was found to support the notion that the genomic mutation rate is increased by exposure to natural levels of UV-B.

Résumé

Ce projet de recherche vise à comprendre l'impact d'une dose naturelle de radiations ultraviolettes-B sur le taux génomique de mutations chez *Arabidopsis thaliana*. Les propriétés mutagéniques des radiations UV-B sont connues, mais il est possible que les plantes aient évolué des mécanismes permettant de supporter les dommages encourus lors d'expositions routinières à ces radiations. Deux expériences d'accumulation de mutations d'une durée d'onze générations ont été faites dans le but de déterminer si le taux génomique de mutations dans cette espèce de plante augmente en présence d'UV-B. La première expérience comporte une dose d'UV-B similaire à celle observée lors d'une journée ensoleillée d'été, alors que la deuxième est définie par l'absence de ce mutagène. Les estimés de taux de mutations, obtenus principalement à partir d'une analyse de maximum de vraisemblance des données phénotypiques, n'étaient pas significativement supérieurs à zéro en absence ou en présence d'UV-B. Il n'y a donc aucune preuve que le taux génomique de mutation est accru par l'exposition à une dose naturelle d'UV-B.

Acknowledgements

I would first and foremost like to thank my supervisor, Dr. Daniel Schoen, for allowing me to infringe upon his project and providing a helping hand and a good dose of insight along the way. Dan provided an excellent environment in which to learn from hands on experience and insightful discussion.

This project would have been much more arduous had it not been for several pairs of helping hands contributing to both the maintenance and propagation of lines, as well as the massive sowing, data collecting and harvesting in the final progeny test. In that light, sincere thanks go to Caroline Belair, Dustin Raab, Adriana Pastor, Brad Peori, Fabienne Saadé, Amy Schwartz, and Julia Sharma. Thanks as well to the Schoen lab members, Mattieu Bégin, Rod Docking, Fabienne Saadé, and Dominic Cloutier, who made life in the lab enjoyable and provided a helpful sounding board. Thanks as well to the always helpful staff of the McGill University Phytotron, including Mark Romer, Claire Cooney, Frank Scopelleti, and Glenn Orr.

I would also like to thank my supervisory committee members, Dr. Graham Bell and Dr. Rajinder Dhindsa, for helpful comments along the way. Thanks to Dr. Ruth Shaw for providing the plant material with which this project was initiated.

Funding was kindly provided by the Natural Sciences and Engineering Research Council PGS-A scholarship and a Richard H. Tomlinson Fellowship.

My acknowledgments would be far from complete if I failed to thank my family, who have provided a nurturing environment for as long as I can remember, one that now extends across hundreds of miles. Thank you Mom, Dad, Craig, Ranger, Spots, Anouk, and my piece of home away from home, Brad. You have helped me in so many ways that I can't even begin to list them. A thank-you hardly seems adequate.

Table of Contents

List of Tables

List of Figures

Introduction

The process of mutation is an important factor for the evolutionary dynamics of populations, but can also have a negative impact on the fitness of a population. Mutational parameters, namely the rate at which new mutations arise in the genome and the effects of these mutations on fitness, are often estimated using mutation accumulation (MA) experiments. The impact of ultraviolet-B radiation, a known mutagen found in the sunlight permeating natural environments, on mutational parameters has not previously been examined using MA experiments. This project was designed to investigate the effects of this important natural mutagen in a plant species.

Populations are able to adapt and evolve because of variation induced by mutation. It is this variation that provides the raw material on which natural selection acts. While beneficial mutations allow the adaptive evolution of populations, there is also a major cost to the mutational process. The majority of mutations are thought to be detrimental (e.g. Keightley et al., 1998), a premise that seems intuitively correct given the potential effect that even a single mutation may have on the proper functioning of an integrated network of genes and their products. A mutation causing just a single nucleotide change that results in an amino acid substitution in its translated product can impede the proper functioning of an enzyme; frameshift and nonsense mutations can completely eliminate a gene product. Mutations that have a severe negative effect on fitness will be removed by selection, but mutations with mildly deleterious effects are not effectively acted upon by selective forces and may thus accumulate in the genome. Even though mutations may allow for the adaptive evolution of populations in changing environments, they may also render populations less fit, and may increase the likelihood of extinction of small

populations (Lande, 1994; Lynch et al., 1995; Zeyl et al., 2001). As well, the accumulation of mutations may threaten the health of human populations that are now largely removed from selective forces due to advances in medical technology (Crow, 1997; Lynch et al., 1999). It is this dichotomy between beneficial and detrimental aspects of mutation that makes knowledge of the mutational process important to the study of evolution.

Mutations are a central feature of many evolutionary theories. For example, it has been proposed that sexual reproduction may enhance the ability of populations to purge detrimental mutations (Kondrashov, 1988). Senescence may occur as the result of a buildup of detrimental mutations in genes expressed in the post-reproductive portion of the lifecycle that are not under strong selective pressure (Medawar, 1952; Charlesworth and Hughes, 1996). Theories pertaining to the evolution of mating systems (Pamilo et al., 1987; Lande and Schemske, 1985) and inbreeding avoidance mechanisms (Charlesworth and Charlesworth, 1998) also invoke the detrimental nature of mutations. Moreover, evolutionary processes, such as selection and adaptation, rely on the existence of genetic variation within a population, variation that is supplied by a constant influx of mutations (Haldane, 1937; Kondrashov and Turelli, 1992). The occurrence of deleterious mutations is thus a crucial component of evolution.

The evaluation of many existing evolutionary hypotheses depends on knowledge of the rate at which mutations arise and their effect on fitness. To this end, experiments have been performed over the last few decades with an eye on obtaining estimates of mutational parameters. Mutation accumulation (MA) experiments have played a

prominent role in attempts to estimate mutation rates and the effects that these mutations have on the fitness of an organism (Charlesworth and Charlesworth, 1998).

The basis of estimating mutational parameters, in particular the genomic rate of mutation and mutational effects, from MA experiments lies in detecting changes in fitness that occur over several generations. Replicate lines, derived from an initially inbred individual at mutation-selection equilibrium are propagated for several generations under relaxed selection. The latter conditions are often obtained through the use of balancer chromosomes or by maintaining small effective population sizes (e.g. single seed descent). Throughout the course of an MA experiment, mutations are expected to occur independently among lines, and the fate of these mutations (fixation or loss) will be governed by random genetic drift. If the majority of mutations are deleterious, then their accumulation in the genome should lead to a decrease in the mean fitness of the assemblage of MA lines. Additionally, since mutations occur independently among the replicate lines, there should be an increase in fitness variation among the MA lines. It is from this decrease in mean fitness and increase in among line variance over time that mutation rates and mutational effects can be estimated.

Three main techniques have been developed to extract estimates of mutation rates and mutational effects from the data collected in MA experiments. The simplest and first established method of obtaining estimates of mutation rates was put forth by Bateman (1959) and employed by Mukai (1964) in his pioneering MA experiment, and is hence known as the Bateman-Mukai method. By simply measuring the per generation decrease in fitness and increase in variance among MA lines, it is possible to calculate a lower bound estimate of the mutation rate and an upper bound estimate of the average effect of

each mutation. The Bateman-Mukai method can, however, lead to inaccurate estimates when the underlying distribution of mutational effects is not well represented by an average (Keightley, 1994). The above may occur, for example, when the distribution of mutational effects is leptokurtic, where most mutations have a minor effect on fitness while a few have a larger effect.

A more complex, and perhaps more robust, method of obtaining estimates of mutational parameters using maximum likelihood was developed by Keightley (1994, 1996, 1998). This computationally intense method can be used to determine the values of the mutational parameters that are most likely to have resulted in the observed data collected from an MA experiment, based on an underlying mutational model (Keightley, 1994); the best parameter estimates will be those associated with the highest likelihood. The maximum likelihood method has a further advantage in that it can be extended to allow the estimation of the distribution of mutational effects, using, for example, the simple but flexible gamma distribution (Keightley, 1994). The inclusion of a distribution of mutational effects goes toward overcoming the problems of Bateman-Mukai estimates associated with estimating a mean mutational effect when it is not well representative of the overall distribution (Keightley, 1994). Improvements have been made to this estimation technique, including modifications that allow the incorporation of both detrimental and beneficial mutational effects (Keightley, 1994; Keightley and Ohnishi, 1998; Shaw et al., 2002), and the use of data from multiple generations (Keightley and Bataillon, 2000).

One further estimation method, the minimum distance method, was proposed by García-Dorado (1997). This technique also allows for the estimation of the distribution of

mutational effects using the gamma distribution. Using this technique, the best parameter estimates are obtained when the distance between the observed data and data theoretically predicted based on a mutational model is minimized (García-Dorado, 1997). The minimum distance method does not require estimates of the mean and variance of a control population, and may thus be useful in studies where control populations cannot easily be maintained in the absence of adaptive evolution, such as *Drosophila melanogaster* (García-Dorado, 1997). The relative merits of these three estimation techniques are currently being debated (Deng et al., 1999; García-Dorado and Gallego, 2003; Keightley, 2004), but all techniques have their own unique virtues, be it the simplicity of Bateman-Mukai estimates, the flexibility of maximum likelihood estimates, or the non-dependence of minimum distance methods on accurate measurements of a control population.

Interest in the mutational process has spawned a great deal of research, primarily utilizing MA methods. MA experiments have been employed in a wide variety of organisms over the past decades, beginning with Mukai's 1964 experiment using the widely studied fruit fly (*Drosophila melanogaster*). This initial MA study by Mukai (1964), and a subsequent follow-up experiment (Mukai et al., 1972) suggested a mutation rate of roughly one mutation per genome per generation, with each mutation having a minor average effect on fitness of only one or two percent. Evolutionary theories have since incorporated assumptions based upon this high rate of mutation and small effect of individual mutations (e.g. Kondrashov, 1988).

In attempts to corroborate Mukai's results, and lend credence to evolutionary theories based on these high estimates of mutation rates, several MA studies have

subsequently been performed. While many traits have been examined in a number of organisms, support for Mukai's high genomic rate of mutation is mixed. MA studies conducted in *Caenohabditis elegans* (Keightley and Caballero, 1997; Vassilieva and Lynch, 1999; Vassilieva et al., 2000; Azevedo et al., 2002; Estes et al., 2004) suggest a rate of mutation two orders of magnitude lower than Mukai's estimates. Estimates of the genomic rate of mutation in *Escherichia coli* (Kibota and Lynch, 1996), the yeast *Saccharomyces cerevisiae* (Korona, 1999; Zeyl and DeVisser, 2001) and the grape pest *Daktulosphaira vitifoliae* (Downie, 2003) also suggest that an estimate of one mutation per genome per generation is unrealistically high. Mutation rate estimates obtained through MA studies in *Daphnia pulex* (Lynch et al., 1998) and RNA viruses (Elena and Moya, 1999; de la Pena et al., 2000), on the other hand, are within the same range as Mukai's estimate. Studies in *Arabidopsis thaliana* (Schultz et al., 1999; Shaw et al., 2000) suggest that the rate of mutation in this plant falls between the two extremes, with estimates roughly an order of magnitude lower than Mukai's estimate, but an order of magnitude higher than estimates in organisms like *C. elegans*. Meanwhile, studies further exploring the properties of mutation in *D. melanogaster* have found a range of estimates of genomic mutation rates, some in line with the initial estimate of one mutation per genome per generation (Ohnishi, 1977; Shabalina et al., 1997) and others substantially lower (Fernández and Lopez-Fanjul, 1996; Fry et al., 1999; Fry, 2001; Chavarrías et al., 2001; Caballero et al., 2002; Charlesworth et al., 2004).

Attempts have been made to reconcile the conflicting estimates of the genomic mutation rate. Considerations have been made of the differences in genome size, gene number, and the number of cell divisions in gametes in the organisms studied (e.g. Kibota

and Lynch, 1996; Schultz et al., 1999), but have met with little success in overcoming the discrepancies. Additionally, differences in genome size and the like cannot explain the wide span of estimates obtained even within a single species, such as *D. melanogaster.* Adding to the confusion, new analyses of the data collected by Mukai (1964; Mukai et al., 1972) using maximum likelihood and minimum distance methods have called into question the validity of Mukai's results (Keightley, 1994, 1996; García-Dorado, 1997; García-Dorado et al., 1999), suggesting that the estimate of one mutation per genome per generation is indeed too high and leading to criticisms of the experimental methods employed (reviewed by Lynch et al., 1999). The reanalysis of previously collected MA data does, however, suggest that the use of different estimation methods may contribute to the variation in mutational parameter estimates (García-Dorado et al., 1999). The disparities in estimates of these important mutational parameters continue to drive further studies.

While many MA studies have been performed, there have been few carried out using plants. Plant species have indeterminate germ lines, a characteristic that may play a role in the number of mutations transmitted from parent to offspring. The formation of the germ line from somatic cells can allow intra-organismal selection to occur. In other words, the germ line may be recruited from only the healthiest cells, or those that have accumulated the fewest mutations (Klekowski and Kazarinova-Fukshansky, 1984; Otto and Orive, 1995). While intra-organismal selection may allow a reduction in the number of inherited mutations, there is another aspect of indeterminate germ lines that may counter this reduction. Germ line cells, when formed early in development, undergo few divisions in comparison to somatic cells. Deriving a germ line from somatic cells that

have undergone many rounds of DNA replication and cell division may provide many opportunities for the occurrence of mutations, and may consequently result in a germ line containing many mutations. This particular aspect of plant biology makes estimates of mutational parameters in plants particularly interesting.

Surprisingly, all MA studies to date have ignored the potential effects of ultraviolet radiation, an important, naturally occurring mutagen. Because of this omission, the mutation rate estimates obtained so far could be underestimates of the rate of mutation in natural populations. Many organisms are routinely exposed to radiation from the sun throughout their lives, including mutagenic ultraviolet-B (UV-B) rays (280-320nm). UV-B radiation both penetrates the atmosphere of the earth and causes DNA damage (Caldwell et al., 1989). The shorter wavelengths of UV-C are filtered out by the Earth's atmosphere and the longer, less energetic wavelengths of UV-A are not thought to be mutagenic (Caldwell et al., 1989). UV-B causes damage to genetic material through a variety of mechanisms, most notably the formation of cyclobutane pyrimidine dimers (Britt, 1996). Most organisms will be exposed to the UV-B rays penetrating the earth's atmosphere throughout their lifetime; plants, however, are particularly vulnerable to UV-B radiation, as shoots constantly intercept UV-B as they collect photosynthetically-active radiation. On the other hand, plants may have adapted to UV-B exposure by evolving mechanisms to shield themselves from radiation or repair any damage that is incurred. In this study, the effects of normal summertime levels of UV-B radiation on the genomic rate of mutation and mutational effects in *Arabidopsis thaliana* are examined.

To study the effects of UV-B on mutational parameters, the plant *Arabidopsis thaliana* was chosen. *A. thaliana* is a naturally self-fertilizing annual widely used in

genetic studies. *A thaliana* is an attractive study organism, as it is easy to grow in growth chamber conditions, has a small size, and a fast rate of growth. Shaw and colleagues (2000) and Schultz and collaborators (1999) have performed MA studies using *A. thaliana*; however, neither study included UV-B radiation in the conditions under which the plants were grown. The lights used in normal growth chambers do not produce UV-B, and the glass of greenhouses filters UV-B from sunlight, so a conscious decision must be made to include UV-B in an MA experiment. Shaw et al. (2000) found no significant differences in mean fruit number or seed set before and after seventeen generations of MA, but increases in among line variance were detected over the course of their study, suggesting that mutations occurred, but had both beneficial and deleterious effects. Further analysis of the data (Shaw et al., 2002) reinforces conclusions regarding the bi-directional effects of mutations, some increasing and others decreasing the fitness related traits studied. In the MA study conducted by Schultz et al. (1999) with *A. thaliana*, an estimate of the minimum rate of mutation of 0.1 mutations per genome per generation was obtained for a composite measure of total fitness (including germination rate, fruit set and seed set). Interestingly, no significant decline in means for germination success and fruit set were detected (Schultz et al., 1999), in line with the possibility that mutations were acting to both increase and decrease fitness (Shaw et al, 2000). Given the relative scarcity of MA studies in plants, the debate surrounding the seeming occurrence of beneficial mutations over the course of the MA studies (see Discussion), and the potential impact of UV-B on mutational parameters, it is hoped that this study can provide additional insight into the mutational process.

Materials and Methods

Mutation Accumulation

A single, highly inbred individual of the Columbia ecotype of *Arabidopsis thaliana*, grown from seed kindly provided by Dr. Ruth Shaw, served as the progenitor of the MA lines used in this experiment. Inbreeding, through self-fertilization, ensures that a balance between mutation and selection has been achieved before the start of MA. Two separate sets of 120 lines were derived from the seed collected from self-fertilization of this individual. One set of lines was subjected to UV-B radiation (280-320nm) throughout the MA process, while the other was protected from UV-B, but grown in the same growth chamber in the Phytotron of McGill University. The two sets of lines were propagated by single seed descent ($N_e = 1$) for eleven generations of MA. The line development and propagation was initiated by Dr. Daniel Schoen and was completed as part of this thesis project. The experimental design is outlined in Figure 1.

To provide photosynthetically-active radiation equivalent to that of a full, cloudless mid-summer day (~2000μM/m^2/second), the growth chamber was illuminated with eight 1000-watt metal halide lamps. This radiation source is also critical to the proper functioning of normal DNA repair enzymes involved in the reparation of UV induced damage (Caldwell, 1981). Plants were grown in 7cm by 7cm square pots filled with a 1:1:1 mixture of ProMix, perlite and vermiculite. To promote synchronized germination, for a week following each planting the growth chamber was kept dark and the temperature held at 4°C. For the remainder of the growth phase, the temperature of the chamber was maintained at 22°C with 70% relative humidity. Plants were bottom watered twice a week, with an additional watering once a week with a fertilizer solution,

alternating between half strength 20-20-20 and Hoagland's solution. Plants were allowed to reach maturity before being harvested, and seeds were stored in labeled envelopes each generation. The next generation was founded by randomly planting two to four collected seeds and randomly thinning after germination to one plant per pot. Remaining seeds were then stored at 4°C. This procedure was repeated until the lines had undergone eleven generations of MA.

UV-B Exposure During Mutation Accumulation

The ultraviolet light source was set up to provide UV-B (280-320nm) equivalent to that encountered on a cloudless day in mid-summer at 45°N, or roughly 7kJ/m^2/day. To achieve this level of exposure, the chamber was equipped with a canopy of ultraviolet lights (Q-Panel Lab Products, Ultraviolet Lamp No. UV-B 313-EL) suspended at a height of 50-70cm above the plants. The UV lights were 'burned in' for a 72-hour period before the plants were placed in the chamber to stabilize the radiation output. As the power output of the lamps diminished over time, the lamps were replaced twice during the eleven generations of MA, once after generation four, and again after generation eight.

Plants exposed to UV-B radiation (henceforth referred to as UV-B exposed lines) were grown in trays placed in wooden boxes and covered with cellulose acetate film. Cellulose acetate allows the transmission of UV-A, UV-B, and photosynthetically-active radiation, but blocks the more harmful UV-C rays. Plants protected from UV-B radiation (henceforth referred to as UV-B protected lines) were grown in similar boxes, but were shielded with Mylar film, which allows penetration of photosynthetically-active radiation

and UV-A, but not UV-B or UV-C. Filters were replaced every week, as the plastics darken over time under the lights.

The aim of the UV-B treatment was to expose plants to natural, rather than artificially high, levels of UV-B radiation; thus, we tried to maintain exposures to achieve UV-B doses of no more than 7kJ/day. At the beginning of the experiment, a spectroradiometer (OL 754-O-PMT, Optronic Laboratories) was used in the growth chamber to measure the spectral irradiance of the ultraviolet lights. These measurements, combined with a DNA damage action spectrum (Setlow, 1974), were used to calibrate a Solar Light PMA 2100 Detector (dose meter) that was used each week to ensure that the proper UV-B dosage was maintained as the lights aged and the plants grew taller. Adjustment of the height of the lamps above the plant canopy and the number of hours of exposure to UV-B (between five and seven hours per day) allowed the dosage to be held at the appropriate level. At the conclusion of the eleven generations of MA, UV-B exposure measurements were again taken with a spectroradiometer (OL 754-O-PMT, Optronic Laboratories) to ensure that the concordance between the UV-B dose measured with the highly accurate, but time-consuming spectroradiometer and the quicker dose meter had been upheld. These measurements confirmed that the dose meter had indeed provided accurate exposure estimates throughout the experiment, with measurements from the two devices differing by no more than 10-15%.

Progeny Testing

After eleven generations of mutation accumulation, seeds stored at the beginning (Generation 1) and end (Generation 11) of MA, from both the UV-B exposed and UV-B

protected lines were sown in a common environment to generate seed of the same age for progeny testing (Figure 1). This procedure reduces the effects of seed storage and environmental differences that may have arisen between the start and the end of the experiment. Two chambers were used at this stage, each with one replicate of the 105 surviving lines of both the UV-B exposed and protected lines at generations 1 and 11. Many of the lines that did not survive to the end of MA were likely lost as a result of environmental variation, and most could have been recovered had time permitted. All plants were sown individually in cones 4cm in diameter and 20cm long, filled with a 1:1:1 mixture of ProMix, perlite and vermiculite. Each plant was randomly assigned a position in one of the fourteen supporting racks in each chamber, with each rack containing thirty plants positioned in a checkerboard pattern. Immediately after sowing, plants were subjected to a week in the dark at 4°C to promote synchronous germination, and were then grown at 70% humidity, 22°C, and 800μM/m^2/second of photosynthetically-active radiation (but no UV). A 14-hour photoperiod was used. Trays were randomized within each chamber every week to reduce the effects of micro-environmental variation within the chambers. At senescence, the infructescense was collected from each plant. The seeds released from these plants were used to sow the final progeny test.

Due to space limitations, we were unable to grow all lines in the progeny test and still maintain the desired level of replication. Thus, regenerated seeds from 86 lines exposed to UV-B radiation throughout mutation accumulation and 85 lines protected from UV-B radiation were used in the final progeny test. To achieve the desired level of replication, five growth chambers were used at this time. Each chamber housed one plant from each UV-B treatment-generation combination from each of the two common growth

(maternal) chambers, randomly assigned a position within the chamber (Figure 1). These plants were again grown in the tubes with a 1:1:1 mixture of ProMix, perlite and vermiculite. Each chamber contained 14 trays, each with 49 plants arranged in a checkerboard pattern (one tray per chamber held only 47 plants) for a total of 684 plants in each chamber and 3420 plants in total. Plants were again kept in the dark at 4°C for one week prior to illumination of the chamber to promote synchronous germination. Any plants that had not germinated after one week of growth under the lights were replaced with new seed. This delay in planting, however, had an observable negative effect on the growth of these replants, and it was decided to exclude them from all analyses. Tray positions were randomized each week to reduce the effects of micro-environmental variation within a chamber.

Three fitness component measurements were taken, each at a different point during the growth of the plants. The first measure was taken sixteen days after germination, and was a count of the number of leaves in the rosette of each plant that were larger than 5mm when measured from the center of the plant outwards along the petiole. After 48 days of growth, the number of flowers, buds, and seed pods on each plant were counted as a measure of the total flower number, which should reflect the total lifetime reproductive output of an individual. Plants were harvested after 56 days of growth by cutting the plant just below the rosette. Harvested plants were placed in labeled envelopes, and were then dried in a forced-air drying oven at 65°C for 48 hours (+/- 2 hours). The third measure, the dry weight of all above-ground parts, was then taken.

Statistical Analysis

Mutations, if deleterious, should act to reduce the fitness of a population. This leads to the expectation that mean trait values for the fitness related traits that were examined should be higher at the outset of the MA experiment (Generation 1) than at the conclusion of the experiment (Generation 11). Moreover, because mutations accumulate independently in the replicate lines, the variance among lines for the fitness related traits should increase from the beginning to the end of the experiment. To examine the changes in the means of the traits measured, a mixed model analysis was used (Proc Mixed, SAS Institute Version 8.2). This procedure allows both fixed and random effects to be included in the model. Line was considered as a random effect. Fixed effects were included to account for differences related to the chamber in which a plant was grown (Chamber), the chamber in which the maternal parent was grown (Maternal Chamber), the position of a plant relative to the edge of a chamber, where air flow had an effect on growth (Edge), and variation related to the people measuring the traits (Observer). To estimate the variance at the first and last MA generations, Proc VARCOMP (SAS Institute, Version 8.2) was implemented with the model described above using the restricted maximum likelihood option. A multivariate analysis was also performed (Proc GLM, SAS Institute, Version 8.2) to obtain estimates of means and variances after accounting for correlations between the measured traits.

Bateman-Mukai estimates of mutation rate and mutational effect are easily obtained from the changes in mean and among line variance. Specifically, the Bateman-Mukai method requires estimates of the per generational change in mean (ΔM) and variance (ΔV) and an estimate of the ancestral mean ($\bar{z}_o$), and yields estimates of the

minimum mutation rate per genome per generation (U_{min}), as well as the maximum mean mutational effect (s_{max}). The Bateman-Mukai formulae are (from Schultz et al., 1999):

$$U_{min} = \frac{2\triangle M^2}{\triangle V} \qquad \bar{s}_{max} = \frac{\triangle V}{\bar{z}_o \triangle M}$$

Using the estimates of the mean value of a trait at the first generation and the last generation, as obtained in Proc Mixed, ΔM was calculated. Similarly, ΔV was calculated from the Proc Varcomp estimates of among line variance.

To provide more robust estimates of the rate of mutation and average mutational effect, the maximum likelihood approach was used. For this study, a Fortran program, based upon numerical solution of the likelihood equation put forth by Keightley and Bataillon (2000) was developed to obtain estimates of the mutation rate and average mutational effect in a manner that allows the inclusion of fixed effects (see Appendix A for Fortran 77 code). Briefly, the method assumes that the number of mutations that become fixed in a line follows a Poisson distribution (mean $\lambda = Ut$, where U is the mutation rate per genome per generation and t is the time since the start of the MA experiment). Each mutation has a constant, additive effect s on the phenotype. Environmental effects are assumed to be normally distributed, with a mean of zero and variance of V_e. The ancestral population (the population at the start of MA) is assumed to have a phenotypic mean of M. If $Z_{k,t}$ represents the phenotypic value (for example, the number of flowers per plant) of line k after t generations of MA, then $Z_{k,t}$ can be expressed as:

$$Z_{k,t} = M + x \cdot s + e + c + ed + ob$$

where x is the number of new mutations that have arisen in line k over t generations of MA, and e is the environmental contribution to the observed phenotype. There were other

factors (fixed effects) in our experiment that may have contributed to the observed phenotypes, including chamber effects (*c*), edge effects (*ed*), and observer effects (*ob*). These factors were also estimated in the likelihood. Maternal effects were not found to have a large effect and were thus not included in the likelihood, but were instead estimated directly from the first generation data by comparing the trait means of these two chambers to the overall mean. Replicates (r_n) of each line were assayed at two generations, one (t_1) and eleven (t_{11}), so the likelihood for line *k* is:

$$\mathcal{L}_k = \sum_{i=0}^{1}\sum_{j=0}^{5} f(Z_{k,t_1,r_1} + i\cdot s + c + ed + ob) \times f(Z_{k,t_1,r_2} + i\cdot s + c + ed + ob) \times ... \times f(Z_{k,t_1,r_n} + i\cdot s + c + ed + ob) \times f(Z_{k,t_{11},r_1} + (i+j)\cdot s + c + ed + ob) \times f(Z_{k,t_{11},r_2} + (i+j)\cdot s + c + ed + ob) \times ... \times f(Z_{k,t_{11},r_n} + (i+j)\cdot s + c + ed + ob) \times p_{t_1}(i) \times p_{t_{11}}(j)$$

where $f(x)$ represents the Gaussian probability density function with mean *M* and variance V_e, $p(x)$ represents the Poisson probability function, *i* represents the number of mutations occurring in the first MA generation and *j* represents the number of new mutations arising between generations one and eleven. The above likelihood equation is based on the assumption that all mutations have an average negative effect; a simple change of sign in front of the *is* and *(i+j)s* terms constrains all effects to be beneficial.

The overall likelihood for all eighty-six (or eighty-five UV-B protected) lines studied is:

$$\mathcal{L} = \prod_{k=1}^{86} \mathcal{L}_k$$

Taking the logarithm, as likelihood values are generally very small, this then becomes:

$$\mathbf{L} = \sum_{k=1}^{86} \log(\mathcal{L}_k)$$

Estimates of *U*, *s*, *M*, *Ve*, and the various fixed effects can be substituted into the log likelihood equation. The maximum value of the log likelihood equation will be the set of estimates that best fits the gathered data. Finding this maximum is not a trivial task, given the number of parameter and fixed effect estimates required. Keightley and Bataillon (2000) employed the simplex algorithm (Nelder and Mead, 1965; Press et al., 1992) to perform this maximization. In an initial attempt to use this algorithm, however, it failed to converge consistently, and a genetic algorithm (Pikaia) was used instead.

The Pikaia algorithm uses the principles of selection and mutation to find the maximum of a function (Charbonneau, 1995; Charbonneau and Knapp, 1996), in this case, the maximum log likelihood value for the mutation model described above. Pikaia starts with sets of randomly chosen values for all parameters being estimated. Each unique set of parameter values can be thought of as an "individual" in a "population" (here, each population was composed of 1000 individuals). The log likelihood values corresponding to each set of parameters is determined, and the individuals with the highest log likelihood values then go on to preferentially contribute to the next generation (i.e. "reproduce"). There is thus selection for the best parameter sets, or those that have the highest log likelihood values. Variation is supplied through recombination and mutation of parameter values. These processes of mutation, recombination, and selection through differential reproduction are continued for a predetermined number of generations. Each run of the Pikaia algorithm spanned twenty-five generations, a time span sufficiently long to observe convergence in most cases.

For each trait measured, the maximum likelihood algorithm was run separately for the data sets corresponding to the UV-B exposed and protected lines. Multiple runs

(10,000) of the maximization algorithm were employed. This was done for two main reasons. First, the maximization algorithm begins at a different random point in the multidimensional solution space each time it is run; secondly, the algorithm converges slowly near the maximum. Thus, multiple runs allow one to more fully explore the solution space and thereby increase the chances of finding the true maximum.

An additional aspect of the program that should be noted pertains to the estimation of the average mutational effect, *s*. An initial examination of the behavior of the line means over the course of MA suggested that mutations might have had both beneficial and detrimental effects on the traits measured (Figures 2 and 3). Thus, separate estimation trials were performed, with mutational effects constrained to be positive (beneficial mutations) or negative (deleterious mutations). The maximum likelihood values obtained when effects were constrained to be beneficial or detrimental were then compared using the Akaike Information Criterion (AIC) test, to compare the fit of these two non-nested models (see Hilborn and Mangel, 1997; Burnham and Anderson, 1998). The model that best fits the data is that with the lowest AIC value. Furthermore, when the difference between two AIC values (ΔAIC) is greater than four, this suggests that one model gives a substantially better fit to the data, while a ΔAIC value greater than ten provides very strong evidence that one model better reflects the data (Burnham and Anderson, 1998).

Once the model giving the best fit to the data has been determined, it is important to determine whether the mutation rate estimate associated with that model is significantly greater than zero. To determine this, the parameter values corresponding to the top log likelihood value associated with the best model (beneficial or deleterious mutational

effects) were re-examined, however, the mutation rate estimate was set to zero (actually a low mutation rate of 10^{-10} had to be used due to constraints associated with the use of the Poisson distribution). The log likelihood values obtained with a mutation rate of zero were then compared to the maximum log likelihood with mutation rate values as initially obtained. A log likelihood ratio test was performed (G = twice the difference of the log likelihoods, with χ^2 distribution and 1 degree of freedom) to determine if the mutation rate estimate associated with the maximum likelihood was significantly greater than zero.

Results

Of the 1700 plants sown for the progeny test of the UV-B protected lines, all but 45 germinated successfully; for the UV-B exposed lines, 43 of the 1720 plants did not germinate. Plants that failed to germinate after one week of growth were replanted, however, this delay greatly influenced the growth of these plants and so they were excluded from all analyses. There were also some plants that, despite successful germination, failed to reach maturity. Seven plants died before flower number measurements were taken for the UV-B protected lines and three more perished before the harvest. Some plants also failed to reach the end of the progeny test for the UV-B exposed lines, with eleven dying before flower number was counted, and an additional two plants not surviving to the harvest. Roughly equal numbers of plants from both generations failed to germinate or reach maturity, suggesting that it was environmental conditions that contributed to the poor performance of these plants.

An examination of residuals of an analysis of the full data set suggested that there were a number of very small plants that were not well incorporated into the models. The removal of small plants that had fewer than ten flowers (less than 10% of the overall mean of roughly 100 flowers) or a dry weight of less than 0.006g (less than 5% of the mean dry weight) not only restored the normal distribution expected of residuals, but eliminated plants that accumulated highly deleterious mutations with which this study and most other MA studies are not concerned (e.g. Mukai et al., 1972 in which lines with less than 10% of the control were excluded from analysis). Furthermore, the majority of these very small plants were housed in one chamber during the progeny test, a chamber in which most plants showed an observable reduction in growth in comparison to plants housed in

the other chambers. All analyses discussed have, therefore, been performed after the removal of these very small plants from the data set.

Trait means and among line variances were calculated separately for each UV-B treatment/generation combination (Table 1). There was little variation among the number of leaves larger than 5mm measured after sixteen days of growth, thus, this measure was excluded from further analyses. Means of the measured traits were expected to be lower after MA (at Generation 11) than before (Generation 1), due to the presumed deleterious nature of mutations; however, there were no significant decreases in means for either of the traits examined, regardless of UV-B conditions (Table 1). Variance among lines was expected to increase from the beginning of MA when the lines were essentially identical, to the end of MA due to the independent occurrence of mutations within the lines. This trend was not, however, observed for either of the traits in the UV-B protected lines, with no among line variation detectable either before or after MA (Table 1). Among line variance in the UV-B exposed lines was high at the onset of MA for both flower number and dry weight (Table 1). After MA, among line variance actually decreased for dry weight in the UV-B exposed lines, whereas flower number did show the pattern expected, with an increase in among line variance over the course of MA (Table 1). The deviations from the expected decrease in trait means and increase in among line variances after MA for the traits examined raises concerns regarding the reliability of the Bateman-Mukai estimates, as none of the traits examined showed both a decrease in overall mean and an increase in among line variation over the course of MA. Estimates derived using this method were thus omitted.

A multivariate analysis was also employed to reveal patterns in means and variances after correlations between the measured traits were taken into account. The patterns were qualitatively the same as those that were seen when each trait was analyzed separately. There were, not surprisingly, relatively high correlations between the traits, with R^2 values ranging between 0.6 and 0.8.

A visual presentation of the behavior of line means shows that, in the case of both flower number (Figure 2) and dry weight (Figure 3), for both UV-B exposed and protected lines, some line means increased from the first generation to the end of the mutation accumulation stage whereas others decreased. This suggests that mutations may have acted to both increase and decrease fitness in these lines. The large amount of among line variation after only one generation of exposure to UV-B, evident in Table 1, can also be seen visually (Figures 2 and 3).

Maximum likelihood estimates of the mutation rate per genome per generation and average mutational effect on flower number are presented graphically in Figure 4. Log likelihood ratio tests, comparing the estimates with the best fit to a model with the mutation rate constrained to zero, were performed with the model (beneficial or deleterious) that provided the best fit to the data. Analysis of the UV-B protected flower number data set resulted in an estimate of the genomic mutation rate not differing significantly from zero ($G = 0.71$, $p > 0.1$), with the beneficial model providing a better fit to the data than the detrimental model ($\Delta AIC = 10.2$, Table 2). Analysis of the UV-B exposed flower number data suggests that, once again, the beneficial model provides a better fit to the data ($\Delta AIC = 4.57$, Table 3), with a rate of mutation not significantly greater than zero for this model ($G = 0.17$, $p > 0.1$). Thus, when the maximum likelihood

(Tables 2 and 3) and ANOVA results (Table 1) for the flower number analysis are taken together, they indicate that the rate of mutation is minimal in the UV-B protected and exposed lines.

The maximum likelihood values and associated mutational parameter estimates for dry weight are shown graphically in Figure 5. In the UV-B protected lines, there is evidence that the deleterious mutational effect model lends a better fit to the dry weight data than does the beneficial model ($\Delta AIC = 4.00$, Table 2). Comparison of the deleterious mutational model to a model with a mutation rate of zero suggests that the mutation rate estimate obtained is not significantly greater than zero ($G = 1.09$, $p > 0.1$). Similarly, the genomic mutation rate estimate for dry weight in the UV-B exposed lines was not significantly greater than zero ($G = 1.38$, $p > 0.1$), with the deleterious mutational model again providing a better fit to the data ($\Delta AIC = 7.25$, Table 3). As was the case for flower number, when all pieces of evidence are considered, there is not strong evidence of a substantial rate of mutation acting on dry weight in either the UV-B exposed or protected lines.

Discussion

Mutation Rate in the Absence of UV-B

We were unable to detect a significant rate of mutation in the UV-B protected MA experiment. Mean flower number and dry weight did not show a significant decline over the course of MA, nor did among line variance increase over time for either of these fitness related traits. This suggests that mutations with a measurable effect on these traits did not occur over the eleven generation span of our experiment. In accordance with this finding, maximum likelihood estimates of the genomic mutation rate did not show strong evidence of a rate of mutation significantly greater than zero for either trait examined. Other MA studies in *A. thaliana* have also failed to observe a significant amount of among line variation after slightly shorter MA periods. Schultz et al. (1999) did not find a significant increase in among line variation for the traits that were assayed before and after ten generations of MA. Similarly, Shaw et al. (2000), while finding significant among line variation after seventeen generations, failed to observe significant variation at the mid-point of their study, after eight generations of MA. Our findings in the UV-B protected MA experiment are not unique among MA studies in plants, despite suggestions that ten MA generations with roughly one hundred MA lines are sufficient for the estimation of mutational parameters (Deng et al., 1999).

Plant species, including *A. thaliana*, do not possess a separate, pre-determined germ line. Rather than developing a distinct germ line early in development, cells are recruited to form a germ line later in their life cycle. Intra-organismal selection may reduce the number of mutations passed down from parent to gamete, and could thereby reduce the observed rates of genomic mutation. If selection can act among cell lines, it is

possible that only the best cell lineages (i.e. those with the fewest deleterious mutations) will be recruited to form the germ line. Modeling efforts have lent support to the notion that selection for the best cells within an individual can act to lower the perceived genomic mutation rate (Klekowski and Kazarinova-Fukshansky, 1984; Otto and Orive, 1995). This factor was invoked in a previous study as one possible explanation for the inability to detect decreases in mean fitness after MA in *A. thaliana* (Shaw et al., 2000) and may contribute to the non-significant estimates of genomic mutation rate obtained in the present study.

Mutations that have a deleterious effect in one environment may not have the same effect in another. This is particularly seen when comparing the fitness of MA lines in stressful and non-stressful environments. Generally, mutations are thought to have a greater deleterious effect on fitness in stressful environments than when a population is experiencing optimal conditions (Kondrashov and Houle, 1994). Of the numerous examinations of the effect of environment on mutational effects, some have found larger fitness decreases in stressful environments (Kondrashov and Houle, 1994; Korona, 1999), whereas others have found that mutational effects show little environmental dependence (Fernández and Lopez-Fanjul, 1996; Fry et al., 1999; Chang and Shaw, 2003). Despite the uncertainty of the impact of stress on mutational effects, this factor may have played a role in our inability to detect mutations arising over the course of the experiment. The present study was performed under conditions that would not have caused the plants much stress, either during MA or in the following progeny test. Plants received sufficient light, water, space and nutrients to not feel the effects of competition, dehydration or nutrient starvation. Had the progeny test been performed in a more stressful environment, perhaps

any mutations that did arise over the course of the study would have had more pronounced effects on fitness.

The Effects of UV-B Exposure

Akin to our findings in the UV-B protected lines, we were also unable to detect the accumulation of mild-effect mutations in the UV-B exposed lines. Once again, from an examination in the changes of the means and among line variances of the traits examined, there was little evidence that mutations with a sufficient effect accumulated over the MA study. Maximum likelihood estimates for both flower number and dry weight in the UV-B exposed lines also do not support a significant rate of mutation. Evidence thus suggests that exposure to natural summertime levels of UV-B does not have a measurable impact on the genomic mutation rate in *A. thaliana.*

Given the unavoidable exposure to UV-B radiation as plants collect the light energy that they need to photosynthesize, mechanisms to either protect themselves from UV radiation and/or repair any damage that is incurred have likely evolved in the early evolutionary history of this group (Rozema et al., 1997). For UV-B radiation to elevate the rate of mutation to a level above that observed when this potential mutagen is absent, these protection and repair mechanisms must be overcome or overwhelmed. The results of this study suggest that such protection is, in fact, highly effective at maintaining the integrity of the genetic information inherent in DNA.

Several compounds have been shown to have properties allowing plants to reflect or absorb damaging UV radiation before it is able to harm tissues or mutate DNA. These compounds include components of cuticular waxes (Long et al., 2003), flavinoids

(Rozema et al., 1997; Britt, 1996), and hydroxycinnamates known to be present in *A. thaliana* (Landry et al., 1995). These compounds shield the lower cell layers of the plant, protecting the genetic material, and so reduce the number of UV induced mutations.

Plant species have also been shown to possess compounds that have the ability to repair the DNA damage often incurred when exposed to UV-B. Photolyases are photo-repair enzymes that can rapidly repair the most common UV-B induced DNA damage, cyclobutane pyrimidine dimers. These enzymes use light energy from UV-A and the visible spectrum to break apart the dimers (Britt, 1996). Studies also suggest that some plants, including *A. thaliana*, also possess other light-dependent enzymes that are able to repair the second most prevalent form of UV-B induced DNA damage, pyrimidine (6-4) pyrimidione photoproducts (Chen et al., 1994). As the plants used in this study were provided with a wide spectrum of light, including the blue and UV-A wavelengths needed for activation, these enzymes may have repaired most of the potentially mutagenic DNA damage before replication.

Apart from the use of light-requiring enzymes to repair DNA damage resulting from exposure to UV-B, plants may use excision repair mechanisms that remove the altered DNA and replace the removed segment with the proper nucleotide sequence. Some evidence suggests that plants may have endonucleases that recognize and cut DNA with the two major types of damage mentioned above (reviewed by Britt, 1996). These damaged areas can also be repaired by non-specific excision repair mechanisms (Britt, 1996). The possibility of both protection and repair mechanisms may, therefore, limit the mutagenic effects of natural levels of UV-B radiation in plants that have evolved these strategies as a way to cope with the necessity of exposure to an environmental mutagen.

Exposure of pollen to environmental mutagens, including UV-B occurs when pollen is released from a plant. The dehydrated state of pollen and its potential exposure to UV-B radiation as it travels from one flower to another make pollen a primary source of mutation in many plant species (Jackson, 1987). A study in *A. thaliana* suggests that heritable mutations resulting from one generation of UV exposure are more prevalent when it is the pollen grains, and not other parts of the plant, that are exposed (Whittle and Johnston, 2003). However, as *A. thaliana* self-fertilizes, pollen may not be exposed to light since it need not be released from the flower. In fact, pollen was removed and deliberately exposed to UV in the study by Whittle and Johnston (2003). Thus, in our study, the flower structure (petals and sepals) may have protected the pollen grains from UV-B radiation and the resultant mutations.

Despite the finding that exposure to naturally occurring levels of UV-B failed to cause an elevation in the genomic rate of mutation, caution should be taken when applying these results to natural systems. The effects of UV-B radiation on mutation rates may become more relevant if trends towards decreasing ozone in the atmosphere continue (Kerr and McElroy, 1993). Ozone is the primary gas responsible for the interception of UV-B radiation in the atmosphere; thus, the lower ozone levels become, the more UV-B can penetrate to the surface of the Earth (Caldwell et al., 1989). While plants may have evolved mechanisms to cope with daily encounters with this mutagen, these defense and repair mechanisms may be overwhelmed by higher levels of UV-B. Furthermore, the physiological effects of UV-B (reviewed in Stapleton, 1992) could affect the vigor of plant populations even in the absence of newly arising mutations.

There is one further point of interest in regards to the effects of UV-B radiation. A curious phenomenon was noticed in the UV-B exposed MA lines. In particular, there was a large amount of among line variance in both flower number and dry weight after only one generation of UV-B exposure, while variation among the UV-B protected lines was minimal (Figures 2 and 3). It is unlikely that this variation is the result of a single generation of MA. Rather, this variation is likely due to physiological effects of UV-B exposure, notably an observable reduction in growth rate, that varied among the seed parents used in progeny testing. Variation in growth rates arose as a result of variation in position relative to the UV-B bulbs. Some plants fell directly beneath UV-B bulbs and thus received a slightly higher dosage of UV-B and suffered a greater negative physiological response than did plants that happened to fall in the gaps between the bulbs. If this physiological response affected seed development, the plants that were directly beneath a UV-B light would have produced poorer seeds than plants that were in positions between two bulbs. While plant positions were randomized from generation to generation, a plant remained stationary throughout any given generation during MA, so there was no normalization of physiological responses. Seeds collected from the MA generation were sown in the common environment. At this stage, maternal effects, not uncommon in plants and observed in a previous MA study in *A. thaliana* (Schultz et al., 1999) would have affected the health of these plants, which would then affect the quality of seeds that were sown for the progeny test. Grand-maternal effects thus seem like a plausible cause of the large amount of variation observed in the first generation UV-B exposed lines. The among line variation present in the eleventh generation UV-B exposed plants is a result of variation due to physiological factors and genetic variation resulting

from the accumulation of mutations. Notably, there was only a small, non-significant amount of among line variation for the UV-B protected plants after one generation of MA. As these plants were shielded from UV-B, physiological responses to this stress would have been negligible, so the position of plants relative to the UV-B bulbs would not impact the physiology of the UV-B protected plants.

Beneficial Mutations

While it has been assumed that the majority of mutations have deleterious effects (e.g. Lynch et al., 1999), a recent study by Shaw and colleagues (2000) has brought this assumption into question. Shaw et al. (2000) failed to find a decrease in means for fitness related traits after seventeen generations of MA in *A. thaliana*, but did observe an increase in among line variance; they have invoked the occurrence of beneficial mutations as an explanation for this phenomenon. Further analysis of their data provides support for the theory that beneficial mutations are more prevalent than has been assumed, suggesting that as many as half of the mutations had a beneficial effect (Shaw et al., 2002). These findings have sparked debate regarding the prevalence of beneficial mutations and the impact that such mutations could have on the outcome of MA experiments (Bataillon, 2000; Keightley and Lynch, 2003; Shaw et al., 2003; Bataillon, 2003). Shaw and her colleagues (2000, 2002) reexamined much of the MA literature and found several cases in which there was not a significant decline in mean, but a trend of increasing among line variation over the course of MA, providing additional support for the assertion that newly arising mutations are not unconditionally deleterious. Further evidence in support of the prevalent occurrence of beneficial mutations comes from a study by Burch and Chao

(1999), who demonstrated that beneficial mutations can arise at a significant rate to restore the fitness of lines already containing deleterious mutations.

The present study also suggests that beneficial mutations may not be as rare as they were once thought to be, as examinations of trends over the course of MA show some lines with a higher mean at the end of MA. However, this observation is not reinforced with a corresponding increase in among line variation. Nor was a significant mutation rate found using either maximum likelihood or Bateman-Mukai estimates. This may suggest that the time frame of our study was too short to observe the accumulation of mutations, or that the statistical power of the experiment was not sufficient to detect minute changes over time.

Despite a seeming increase in trait means in some lines over time, it is possible that these increases do not act in a beneficial manner. The selective forces acting on the traits surveyed may also play a role in the mutational effects found in this study. There is evidence from an examination of line means that mutations acted in both a beneficial and detrimental manner on the traits measured. Flower number may, however, be under stabilizing selection, thus, any deviation, either above or below, optimum flower number could be viewed as detrimental. This issue was raised by Keightley and Lynch (2003), regarding the evidence of beneficial mutations acting on the number of fruits per plant examined by Shaw and colleagues (2002), a measure equivalent to flower number in the present study. While Shaw et al. (2003) defend their measure as a component of fitness, a tradeoff between fruit number and flower number may exist, implying the action of stabilizing selection to maintain an ideal balance between fitness gained through pollen and ovules (reviewed by Morgan, 1994). The other trait examined here, above-ground

dry weight of the above-ground plant parts, is more likely to be under directional selection, suggesting that increases in dry weight in some lines over the course of the MA experiment are truly the result of beneficial mutations.

The Larger Picture

This study can be placed within the larger frame of MA studies done to date. In the two other MA studies done in plant systems (Shaw et al., 2000; Schultz et al., 1999), mutation rate estimates are in the range of 0.1 mutations per genome per generation, or a new mutation arising roughly every ten generations. In comparison to mutation rate estimates that have been measured in other organisms, the mutation rate estimates derived from MA studies in plants are almost an order of magnitude lower than estimates in *Drosophila* and *Daphnia* (reviewed above), and an order of magnitude higher than estimates in *C. elegans* (reviewed above), but fall well within the range of measured mutation rates. Despite the addition of UV-B, a previously overlooked, naturally occurring mutagen, we failed to detect a significant genomic mutation rate. This suggests that ambient levels of UV-B do not elevate the rate of genomic mutation in *A. thaliana*. While the rather short duration of this experiment may have resulted in the lack of evidence of a significant rate of mutation, it is unlikely that a mutation rate as high as one mutation per genome per generation is operating in this system, even under the influence of UV-B. In simulations performed while designing the progeny test, when data sets reflecting a rate of mutation of 0.5 mutations per genome per generation with an average effect of 0.05, a significant rate of mutation was estimated using the maximum likelihood method roughly 90% of the time, despite the inclusion of high levels of environmental

variation. If the mutation rate was of this magnitude, mutations should have been detected even over the eleven generation span of our experiment, unless these mutations were of minuscule effect.

While only two published MA studies have been carried out in plants, an indirect method for estimating genomic mutation rates, proposed by Charlesworth and colleagues (1990), has been employed in an effort to quantify mutational parameters in several plant species. The method relies on estimates of inbreeding depression, which is estimated by outcrossing a population of inbred individuals and comparing the fitness of these hybrid individuals to that of the inbred individuals. At equilibrium between mutation and selection, the magnitude of inbreeding depression is a function of the deleterious mutation rate, the dominance of new mutations, and the rate of self-fertilization; thus, estimates of the deleterious mutation rates can be deduced from measures of inbreeding depression (Charlesworth et al., 1990). This method is most robust when used to examine self-fertilizing or otherwise inbred species where over-dominance is unlikely, and is, therefore, most useful in plant species. The inbreeding method has also been used outside of plant taxa to estimate the deleterious mutation rate in *Daphnia* (Deng and Lynch, 1996; 1997). A study by Johnston and Schoen (1995) employed this technique to estimate the rate of deleterious mutation in two *Amsinckia* species, obtaining estimates ranging from 0.24 to 0.87 mutations per genome per generation. Another study by Charlesworth and collaborators (1994) estimated mutation rates in *Leavenworthia* species with this method, obtaining values ranging from 0.71 to 1.68 mutations per genome per generation. These estimates suggest that mutation rates in plant species may be in the same range as the high genomic mutation rate estimated by Mukai (1964; Mukai et al., 1972); however, such

high rates are not reinforced by estimates obtained using MA methods. In spite of the addition of mutagenic UV-B radiation in this study, mutation rate estimates of this magnitude were not uncovered.

Rates of mutation can also be estimated by examining the divergence of DNA sequences in two closely related species. Sequence divergence is measured in a non-functional region, in which all mutations are expected to be neutral, and regions of DNA that code for functional products and are thus expected to be acted on by selection. Alternatively, rates of synonymous and non-synonymous mutation in protein coding regions can be compared in a similar fashion (Eyre-Walker and Keightley, 1999). Any deleterious mutations arising in the functional regions should be removed by selection, while all mutations occurring in the non-functional region will be neutral. Thus, the difference between the rate of divergence of the non-functional DNA and that of the functional DNA should be a reflection of the rate of deleterious mutation (Eyre-Walker and Keightley, 1999). Such studies have proven useful in obtaining estimates of mutation rates in species where performing MA studies is not a viable alternative, for example in humans (Eyre-Walker and Keightley, 1999). Estimates obtained using this method are generally high (Eyre-Walker and Keightley, 1999; Keightley and Eyre-Walker, 2000; Eyre-Walker et al., 2002), in accord with Mukai's early estimates (1964; Mukai et al., 1972); however, an estimate of the rate of mutation in *D. melanogaster* using this method is substantially lower (Keightley and Eyre-Walker, 1999). Barriers to the use of this method, chiefly the lack of accurate knowledge of divergence times, still remain to employing this method in plants. Comparisons of rates of mutation obtained via this method in plants to those obtained using other methods cannot be made.

Conclusions

In conclusion, we have found no evidence that exposure to UV-B radiation elevates the rate of genomic mutation in *Arabidopsis thaliana*. This suggests that plants have evolved effective mechanisms to protect themselves from UV-B and repair damage induced by UV-B. We failed to detect a significant rate of mutation in either of the MA studies. This may be attributed to the short length of the study, the use of benign environmental conditions, and cell lineage selection operating prior to the formation of the germ line. As a cautionary note, despite the finding that natural levels of UV-B failed to elevate mutation rates above the level documented in *A. thaliana* from MA studies performed in the absence of this potential mutagen (Schultz et al., 1999; Shaw et al., 2000), the possible effects of UV-B radiation should not be ignored, especially given the elevated levels of UV-B able to reach the surface of the earth as a result of holes in the protecting ozone layer.

Literature Cited

Azevedo, R. B. R., P. D. Keightley, C. Laurén-Määttä, L. L. Vassilieva, M. Lynch, and A. M. Leroi. 2002. Spontaneous mutational variation for body size in *Caenorhabditis elegans*. Genetics 162: 755-765.

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

Bataillon, T. 2003. Shaking the 'deleterious mutations' dogma? Trends in Ecology and Evolution 18: 315-317.

Bateman, A. J. 1959. The viability of near-normal irradiated chromosomes. International Journal of Radiation Biology 1: 170-180.

Britt, A. B. 1996. DNA damage and repair in plants. Annual Review of Plant Physiology and Plant Molecular Biology 47: 75-100.

Burch, C. L., and L. Chao. 1999. Evolution by small steps and rugged landscapes in the RNA virus ϕ6. Genetics 151: 921-927.

Bunham, K. P., and D. R. Anderson. 1998. Model Selection and Inference: A Practical Information-Theoretic Approach. Springer-Verlag: New York.

Caballero, A., E. Cusi, C. García, and A. García-Dorado. 2002. Accumulation of deleterious mutations: additional *Drosophila melanogaster* estimates and a simulation of the effects of selection. Evolution 56: 1150-1159.

Caldwell, M. M. 1981. Plant response to solar ultraviolet radiation. In O.L. Lange, P.S. Nobel, C.B. Osmond, and H. Ziegler (Eds). Physiological Plant Ecology I: Responses to the Physical Environment. Springer-Verlag: New York.

Caldwell, M. M., A. H. Teramura, and M. Tevini. 1989. The changing solar ultraviolet climate and the ecological consequences for higher plants. Trends in Ecology and Evolution 4: 363-367.

Chang, S. M., and R. G. Shaw. 2003. The contribution of spontaneous mutation to variation in environmental response in *Arabidopsis thaliana*: responses to nutrients. Evolution 57: 984-994.

Charbonneau, P. 1995. Genetic algorithms in astronomy and astrophysics. The Astrophysical Journal (Supplements) 101: 309-334.

Charbonneau, P., and B. Knapp. 1996. A User's Guide to PIKAIA 1.0, NCAR Technical note 418+IA. Boulder: National Center for Atmospheric Research.

Charlesworth, B., and D. Charlesworth. 1998. Some evolutionary consequences of deleterious mutations. Genetica 102/103: 3-19.

Charlesworth, B., and K. A. Hughes. 1996. Age-specific inbreeding depression and the components of genetic variance in relation to the evolution of senescence. Proceedings of the National Academy of Sciences of the United States of America 93: 6140-6145.

Charlesworth, B., D. Charlesworth, and M. T. Morgan. 1990. Genetic loads and estimates of mutation rates in highly inbred plant populations. Nature 347: 380-382.

Charlesworth, D., E. E. Lyons, and L. B. Litchfield. 1994. Inbreeding depression in two highly inbreeding populations of *Leavenworthia*. Proceedings of the Royal Society of London Series B. 258: 209-214.

Charlesworth, B., H. Borthwick, C. Bartolomé and Patricia Pignatelli. 2004. Estimates of the genomic mutation rate for detrimental alleles in *Drosophila melanogaster*.

Genetics 167: 815-826.

Chavarrías D., C. López-Fanjul, and A. García-Dorado. 2001. The rate of mutation and the homozygous and heterozygous mutational effects for competitive viability: a long-term experiment with *Drosophila melanogaster*. Genetics 158: 681-693.

Chen, J. J., D. L. Mitchell, and A. B. Britt. 1994. A light dependent pathway for the elimination of UV-induced pyrimidine (6-4) pyrimidinone photoproducts in *Arabidopsis*. The Plant Cell 6: 1311-1317.

Crow, J. F. 1997. The high spontaneous mutation rate: is it a health risk? Proceedings of the National Academy of Sciences of the United States of America 94: 8380-8386

de la Pena, M., S. F. Elena, and A. Moya. 2000. Effect of deleterious mutation-accumulation on the fitness of RNA bacteriophage MS2. Evolution 54: 686-691.

Deng, H. W., J. Li, and J. L. Li. 1999. On the experimental design and data analysis of mutation accumulation experiments. Genetical Research 73: 147-164.

Deng, H. W., and M. Lynch. 1996. Estimation of deleterious-mutation parameters in natural populations. Genetics 144: 349-360.

Deng, H. W., and M. Lynch. 1997. Inbreeding depression and inferred deleterious-mutation parameters in *Daphnia*. Genetics 147: 147-155.

Downie, D. A. 2003. Effects of short-term spontaneous mutation accumulation for life history traits in grape phylloxera, *Daktulosphaira vitifoliae*. Genetica 119: 237-251.

Elena, S. F., and A. Moya. 1999. Rate of deleterious mutation and the distribution of its effects on fitness in vesicular stomatitis virus. Journal of Evolutionary Biology 12: 1078-1088.

Estes, S., P. C. Phillips, D. R. Denver, W. K. Thomas, and M. Lynch. 2004. Mutation accumulation in populations of varying size: the distribution of mutational effects for fitness correlates in C*aenorhabditis elegans*. Genetics 166: 1269-1279.

Eyre-Walker, A., and P. D. Keightley. 1999. High genomic deleterious mutation rates in hominds. Nature 397: 344- 347.

Eyre-Walker, A., P. D. Keightley, N. G. C. Smith, and D. Gaffney. 2002. Quantifying the slightly deleterious mutation model of molecular evolution. Molecular Biology and Evolution 19: 2142-2149.

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

Fry, J. D. 2001. Rapid mutational declines of viability in *Drosophila*. Genetical Research, 77: 53-60.

Fry, J. D., P. D. Keightley, S. L. Heinsohn, and S. V. Nuzhdin. 1999. New estimates of the rates and effects of mildly deleterious mutation in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America 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., and A. Gallego. 2003. Comparing analysis methods for mutation-accumulation data: a simulation study. Genetics 164: 807-819.

García-Dorado, A., C. López-Fanjul, and A. Caballero. 1999. Properties of spontaneous mutations affecting quantitative traits. Genetical Research 74: 341-350.

Haldane, J. B. S. 1937. The effect of variation on fitness. American Naturalist 71: 337-349.

Hilborn, R., and M. Mangel. 1997. The Ecological Detective: Confronting Models with Data. Princeton University Press: Princeton. pp. 159-160.

Jackson, J. F. 1987. DNA repair in pollen: a review. Mutation Research 181: 17-29.

Johnston, M. O., and D. J. Schoen. 1995. Mutation rates and dominance levels of genes affecting total fitness in two angiosperm species. Science 267: 226-229.

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. 2004. Comparing analysis methods for mutation-accumulation data. Genetics 167: 551-553.

Keightley, P. D., and T. M. Bataillon. 2000. Multigenerational maximum likelihood analysis applied to mutation-accumulation experiments in *Caenorhabditis elegans*. Genetics 154: 1193-1201.

Keightley, P. D., and A. Caballero. 1997. Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. Proceedings of the National Academy of Sciences of the United States of America 94: 3823-3827.

Keightley, P. D., and A. Eyre-Walker. 1999. Terumi Mukai and the riddle of deleterious mutation rates. Genetics 153: 515-523.

Keightley, P. D., and A. Eyre-Walker. 2000. Deleterious mutations and the evolution of sex. Science 290: 331-333.

Keightley, P. D., and M. Lynch. 2003. Toward a realistic model of mutations affecting fitness. Evolution 57: 683-685.

Keightley, P. D. and O. Ohnishi. 1998. EMS-induced polygenic mutation rates for nine quantitative characters in *Drosophila melanogaster*. Genetics 148: 753-766.

Keightley, P. D., A. Caballero, and A. García-Dorado. 1998. Population genetics: surviving under mutation pressure. Current Biology 8: R235-R237.

Kerr, J. B., and C. T. McElroy. 1993. Evidence for large upward trend of ultraviolet-B radiation linked to ozone depletion. Science 262: 1032-1034.

Kibota, T. T., and M. Lynch. 1996. Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature 381: 694-696.

Klekowski, E. J., and N. Kazarinova-Fukshansky. 1984. Shoot apical meristems and mutation: selective loss of disadvantageous cell genotypes. American Journal of Botany 71: 28-34.

Kondrashov, A. S. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336: 435-440.

Kondrashov, A. S., and D. Houle. 1994. Genotype-environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. Proceedings of the Royal Society of London, Series B 258: 221-227.

Kondrashov, A. S., and M. Turelli. 1992. Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. Genetics 132: 603-618.

Korona, R. 1999. Genetic load of the yeast *Saccharomyces cerevisiae* under diverse environmental conditions. Evolution 53: 1966-1971.

Lande, R. 1994. Risk of population extinction from fixation of new deleterious mutations. Evolution 48: 1460-1469.

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

Landry, L. G., C. C. S. Chapple, and R. L. Last. 1995. *Arabidopsis* mutants lacking phenolic sunscreens exhibit enhanced ultraviolet-B injury and oxidative damage. Plant Physiology 109: 1159-1166.

Long, L. M., H. P. Patel, W. C. Cory, and A. E. Stapleton. 2003. The mazie epicuticular wax layer provides UV protection. Functional Plant Biology 30: 75-81.

Lynch, M., J. Conery, and R. Bürger. 1995. Mutation accumulation and the extinction of small populations. American Naturalist 146: 489-518.

Lynch, M., L. Latta, J. Hicks, and M. Giorgianni. 1998. Mutation, selection, and the maintenance of life-history variation in a natural population. Evolution 52: 727-733.

Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz, L. Vassilieva, and J. Willis. 1999. Spontaneous deleterious mutation. Evolution 53: 645-663.

Medawar, P. B. 1952. An Unsolved Problem of Biology. H.K. Lewis: London.

Morgan, M. T. 1994. Models of sexual selection in hermaphrodites, especially plants. American Naturalist 144: S100-S125.

Mukai, T. 1964. The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rates of polygenes controlling viability. Genetics 50: 1-19.

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.

Nelder, J. A., and R. Mead. 1965. A simplex method for function minimization. Computer Journal 7: 308-313.

Ohnishi, O. 1977. Spontaneous and ethyl methanesulphonate-induced mutations controlling viability in *Drosophila melanogaster*. II. Homozygous effect of polygenic mutations. Genetics 87: 529-545.

Otto, S. P., and M. E. Orive. 1995. Evolutionary consequences of mutation and selection within an individual. Genetics 141: 1173- 1187.

Pamilo, P., M. Nei, and W. H. Li. 1987. Accumulation of mutations in sexual and asexual populations. Genetical Research 49: 135-146.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes in Fortran 77. Cambridge University Press, Cambridge.

Rozema, J., J. van de Staaij, L. O. Björn, and M. Caldwell. 1997. UV-B as an environmental factor in plant life: stress and regulation. Trends in Ecology and Evolution 12: 22-28.

Schultz, S. T., M. Lynch, and J. H. Willis. 1999. Spontaneous deleterious mutation in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America 96: 11393-11398.

Setlow, R. B. 1974. The wavelengths in sunlight effective in producing skin cancer: a theoretical analysis. Proceedings of the National Academy of Sciences of the United States of America 71: 3363-3366.

Shabalina, S. A., L. Y. Yampolsky, and A. S. Kondrashov. 1997. Rapid decline of fitness in panmictic populations of *Drosophila melanogaster* maintained under relaxed natural selection. Proceedings of the National Academy of Sciences of the United States of America 94: 13034-13039.

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

Shaw, R. G., D. L. Byers, and E. Darmo. 2000. Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. Genetics 155: 369-378.

Shaw, R. G., F. H. Shaw, and C. Geyer. 2003. What fraction of mutations reduces fitness? A reply to Keightley and Lynch. Evolution 57: 686-689.

Stapleton, A. E. 1992. Ultraviolet radiation and plants: burning questions. The Plant Cell 4: 1353-1358.

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

Vassilieva, L. L, A. M. Hook, and M. Lynch. 2000. The fitness effects of spontaneous mutations in *Caenorhabditis elegans*. Evolution 54: 1234-1246.

Whittle, C. A., and M. O. Johnston. 2003. Male-biased transmission of deleterious mutations to the progeny in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America 100: 4055-4059.

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

Zeyl, C., M. Mizesko, and J. A. G. M. DeVisser. 2001. Mutational meltdown in laboratory yeast populations. Evolution 55: 909-917.

Table 1 – Means and among line variances for dry weight and flower number before (Generation 1) and after (Generation 11) MA.

	Mean Before MA	Mean After MA	p-value[1]	Variance Before MA	Variance After MA
UV-B Protected					
Flower Number	101.44	101.81	0.8124	0	0
Dry Weight	0.1917	0.1919	0.9410	0	0
UV-B Exposed					
Flower Number	100.78	100.37	0.8078	39.68221*	79.45073**
Dry Weight	0.1929	0.1877	0.0503	0.0000909*	0.00004228

[1]For comparison of means before and after MA

* Significantly greater than zero at the 95% level

**Significantly greater than zero at the 99% level.

Table 2 - Estimates of mutation rate and average mutational effect associated with the highest log likelihood obtained in 10,000 runs for the UV-B protected MA lines of *Arabidopsis thaliana.*

Trait	Estimated Mutation Rate	Estimated Average Mutational Effect	Maximum Log Likelihood	AIC Value
Beneficial Mutation Model				
Flower Number	0.121	0.056	-7935.38	7967.38*
Dry Weight	0.194	0.043	-8800.07	8822.07
Deleterious Mutation Model				
Flower Number	0.230	0.009	-7945.58	7977.58
Dry Weight	0.073	0.014	-8796.07	8818.07*

* Model providing better fit to the data by Akaike Information Criterion (AIC).

Table 3 - Estimates of mutation rate and average mutational effect associated with the highest log likelihood obtained in 10,000 runs for the UV-B exposed MA lines of *Arabidopsis thaliana.*

Trait	Estimated Mutation Rate	Estimated Average Mutational Effect	Maximum Log Likelihood	AIC Value
Beneficial Mutation Model				
Flower Number	0.131	0.005	-8148.69	8180.69*
Dry Weight	0.001	0.026	-8950.20	8972.29
Deleterious Mutation Model				
Flower Number	0.096	0.013	-8153.26	8185.26
Dry Weight	0.104	0.007	-8942.95	8964.95*

* Model providing better fit to the data by Akaike Information Criterion.

Figure Legends

Figure 1 – Experimental design of the mutation accumulation experiment and subsequent progeny test. Two sets of lines (UV-B Protected lines shown in the top half of the figure, UV-B Exposed lines shown in the bottom half) were derived from a common ancestor and propagated by single seed descent for eleven mutation accumulation generations. Seed collected from the first and last MA generations of both the UV-B exposed and protected lines were then grown together for one generation in two chambers for the common growth stage. Seeds collected from a subset of the lines grown together were then sown in five chambers for the progeny test. Measures of leaf number, flower number and dry weight were taken on these plants.

Figure 2 – Mean flower number at the beginning and end of MA for each line assayed in the progeny test. A) UV-B protected lines B) UV-B exposed lines.

Figure 3 – Mean dry weight at the beginning and end of MA for each line assayed in the progeny test. A) UV-B protected lines B) UV-B exposed lines.

Figure 4 – Distribution of likelihood values associated with mutational parameter estimates for flower number. The top 100 estimates of 10,000 are shown. U is the mutation rate per genome per generation, s is the average absolute mutational effect. Plots A & B are the results for the UV-B Protected lines, when effects are assumed to be beneficial or detrimental, respectively. Similarly, Plots C & D are the results for the UV-B Exposed lines.

Figure 5 – Distribution of likelihood values associated with mutational parameter estimates for dry weight. The top 100 estimates of 10,000 are shown. *U* is the mutation rate per genome per generation, *s* is the average absolute mutational effect. Plots A & B are the results for the UV-B Protected lines, when effects are assumed to be beneficial or detrimental, respectively. Similarly, Plots C & D are the results for the UV-B Exposed lines.

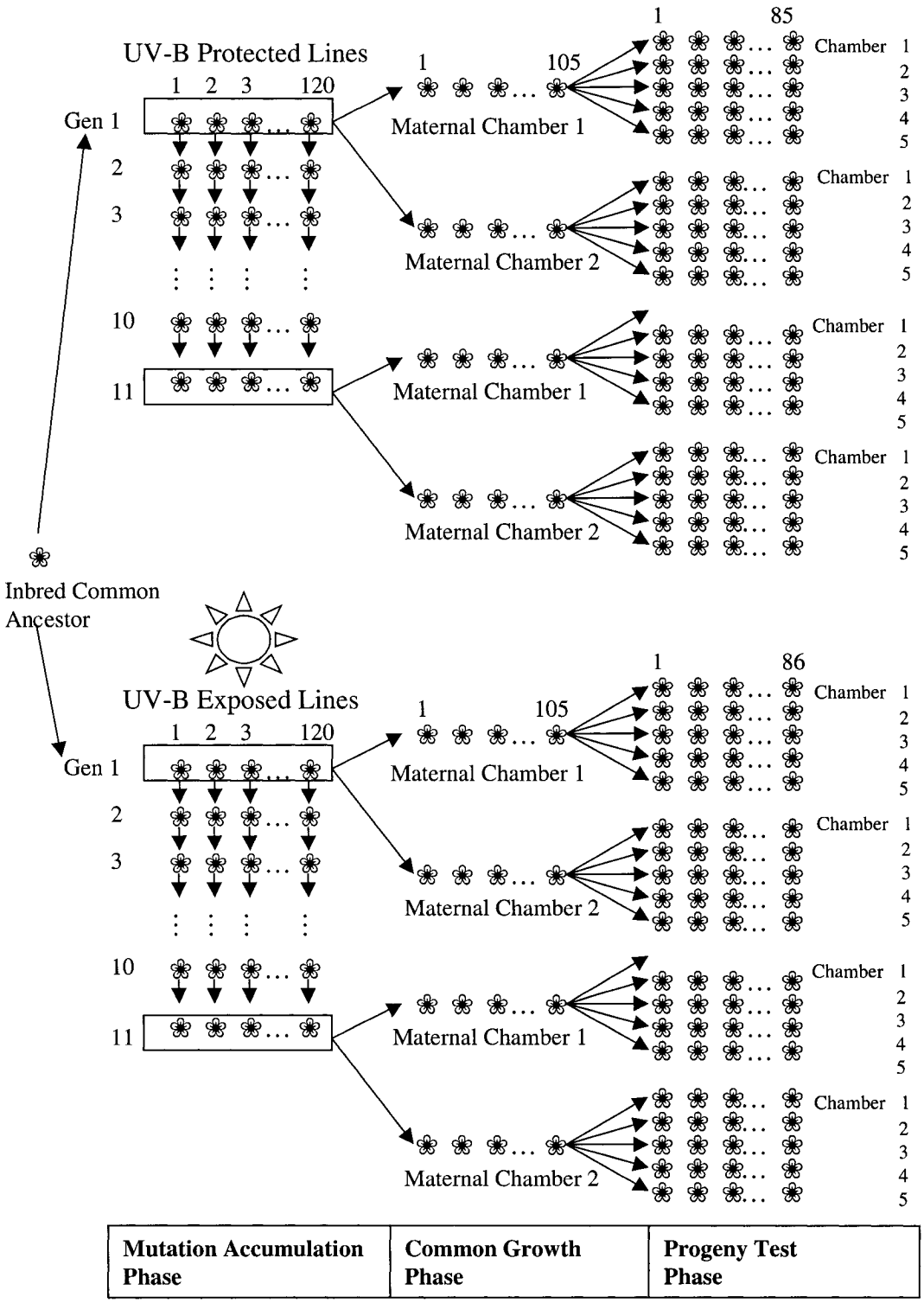
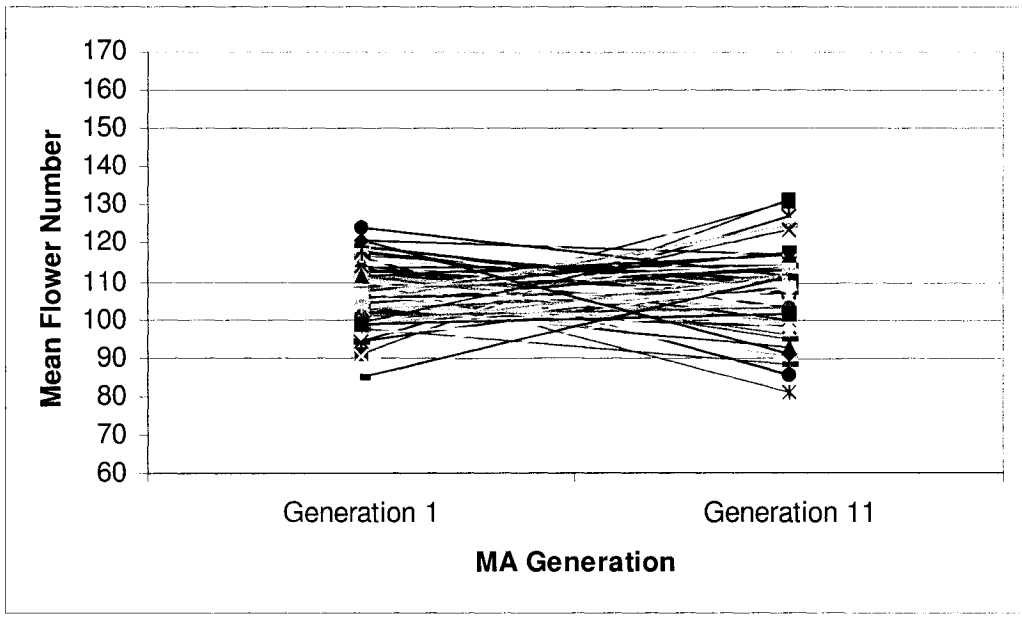
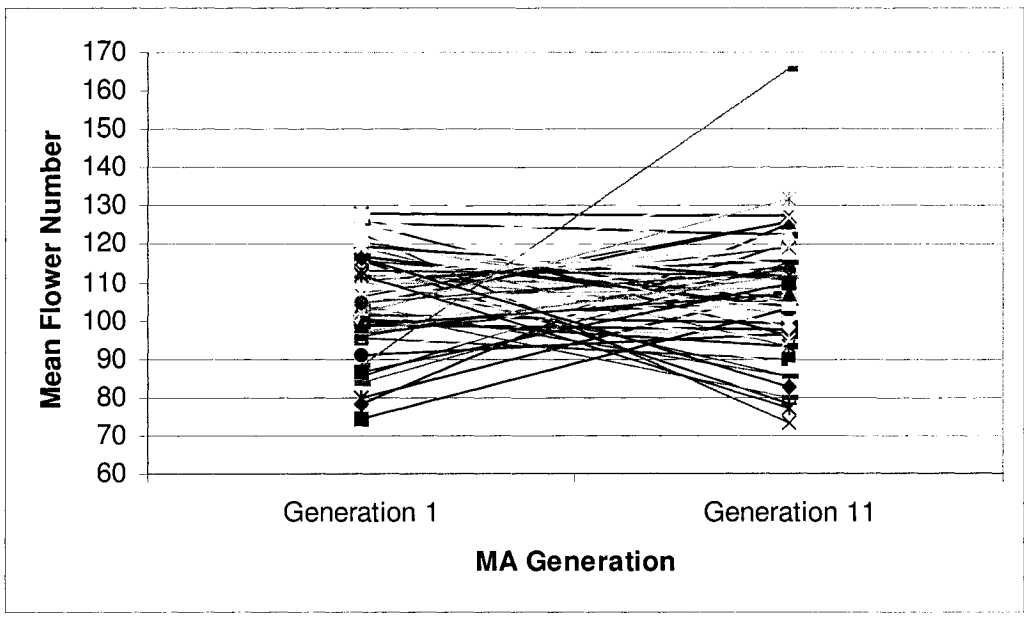

UV-B Protected Lines
1 2 3 120
Gen 1
2
3
10
11
1 105
Maternal Chamber 1
Maternal Chamber 2
1 85
Chamber 1
2
3
4
5
Inbred Common Ancestor
UV-B Exposed Lines
1 2 3 120
Gen 1
2
3
10
11
1 105
Maternal Chamber 1
Maternal Chamber 2
1 86
Chamber 1
2
3
4
5
Mutation Accumulation Phase
Common Growth Phase
Progeny Test Phase

Figure 1

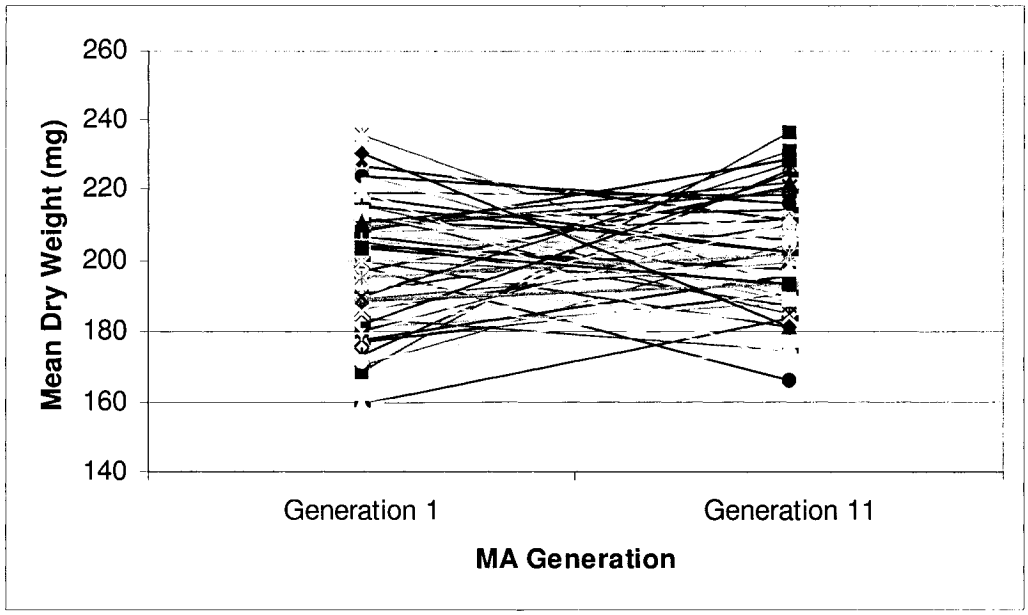


A) UV-B Protected

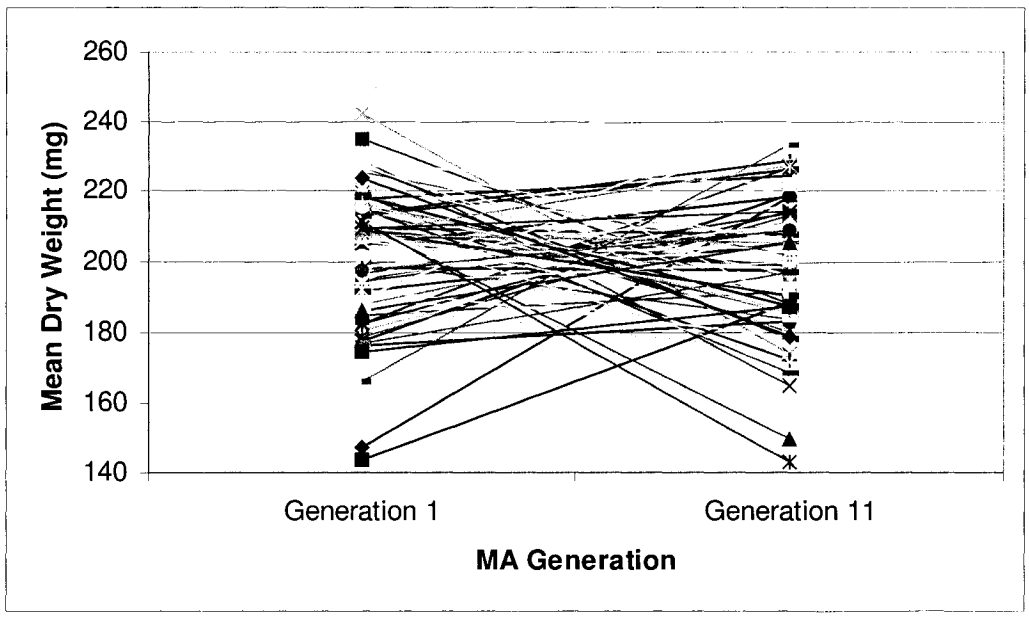


B) UV-B Exposed

Figure 2

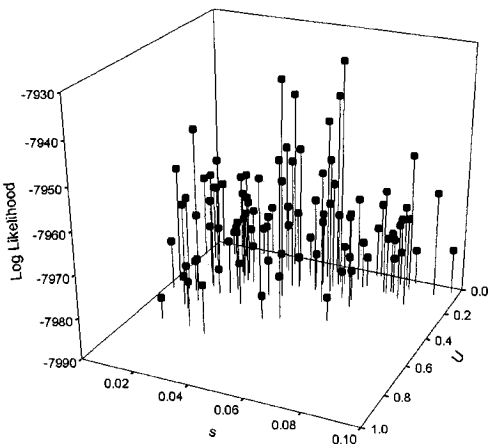


A) UV-B Protected

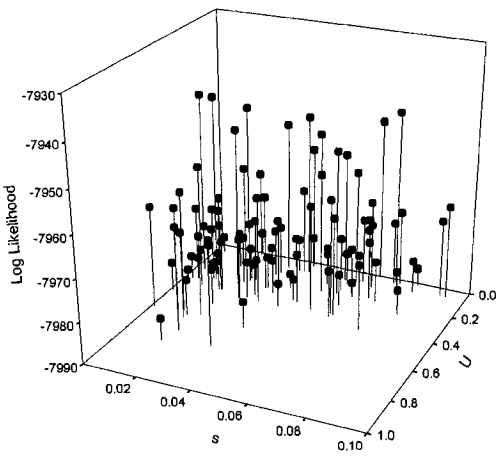


B) UV-B Exposed

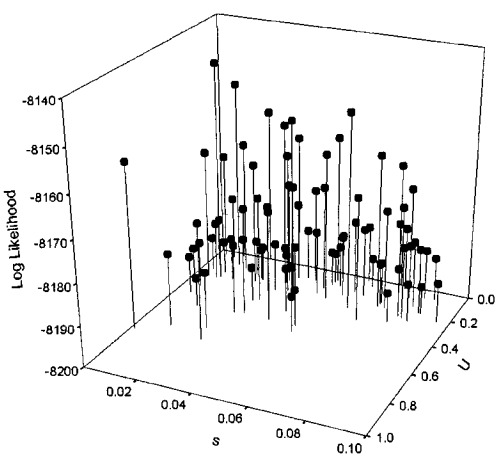
Figure 3



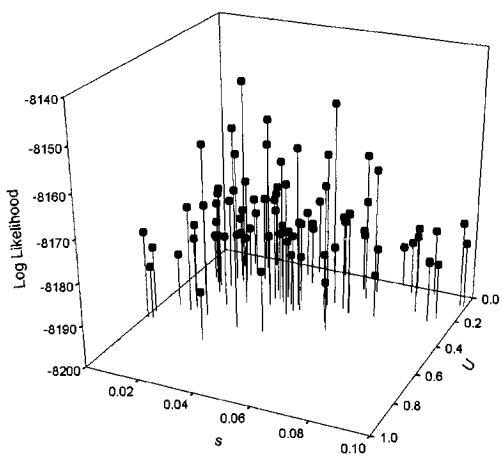
A) UV-B Protected
Beneficial Effects



B) UV-B Protected
Detrimental Effects

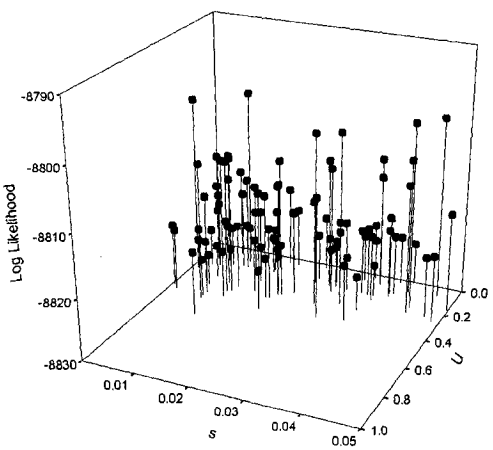


C) UV-B Exposed
Beneficial Effects

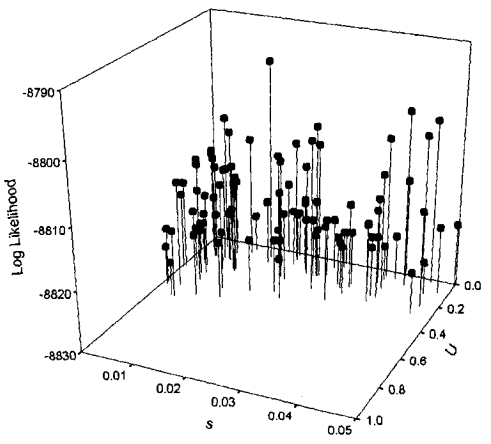


D) UV-B Exposed
Detrimental Effects

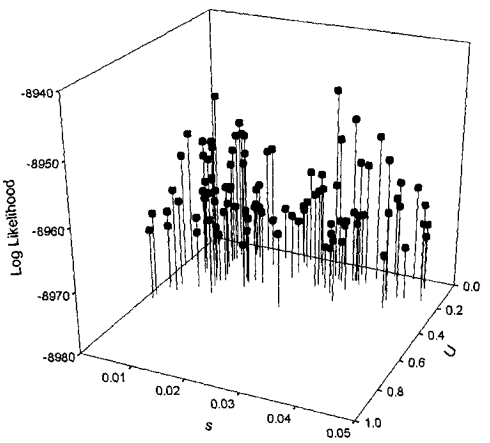
Figure 4



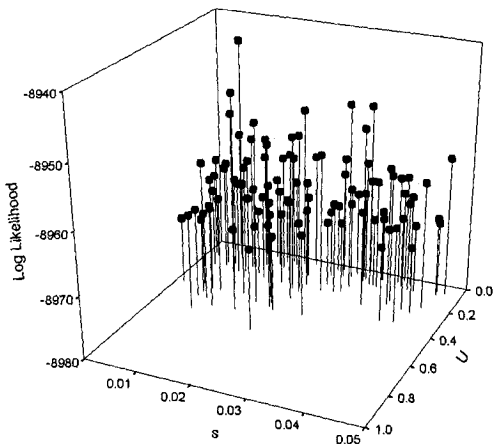
A) UV-B Protected
Beneficial Effects



B) UV-B Protected
Detrimental Effects



C) UV-B Exposed
Beneficial Effects



D) UV-B Exposed
Detrimental Effects

Figure 5

Appendix A - Fortran 77 Code for Maximum Likelihood Program

```
C*********************************************************************C
C*********************************************************************C
C*    Program ArabidopsisMaximumLikelihood--                         *C
C*    Determines the likelihood of a set of parameter estimates      *C
C*     using the mutational model put forth by Keightley and         *C
   C*     Bataillon (2000)                                        *C
C*    Finds the best parameter set (highest log likelihood) using    *C
C*    the genetic algorithm Pikaia                                   *C
C*********************************************************************C
C*********************************************************************C

      Program ArabidopsisMaximumLikelihood

c     PROGRAM REQUIREMENTS
c     Requires coded data file
c     Requires an output file initialized with 0
c     Requires a file containing random numbers
c     Requires the IMSL Library

c     EXPLANATION OF PROGRAM VARIABLES
c     NOTE: These variables will be found throughout the various
c           subroutines and functions composing this program
c     numobs=number of observations in the file. Must be known
c     numfix=number of fixed effects
c     Levels=array containing the number of levels of each fixed effect
c     NegPos=array containing the sign of each fixed effect estimate
c     Info=array containing data
c     LogLikelihood=log likelihood value associated with a set of
c           parameter estimates

c     VARIABLE DECLARATIONS
      Integer numobs,numfix,Levels(10),NegPos(10,10)
      Parameter(numobs=1590)
      Real LogLikelihood,Info(numobs,10)
      External LogLikelihood
      Implicit none


      Call GetInfo(Info,numobs,NumFix,Levels)
      Call CalcFixedEffects(Info,numobs,NumFix,Levels,NegPos)
      Call RunPikaia(LogLikelihood,Info,numobs,NegPos)


      End
```

```
C*********************************************************************C
*     Subroutine GetInfo--                                            *
*     Reads data from a text file ('DataFile.txt)into an array        *
*     Sets the number of fixed effects                                *
C*********************************************************************C
      Subroutine GetInfo(Info,numobs,NumFix,Levels)

c     VARIABLE DECLARATIONS
      Integer Numobs,NumFix,Levels(10)
      Real Info(numobs,10)
```

```
c     Set the number of fixed effects and the number of levels of each
      Numfix=4
      Levels(1)=5  !Number of chambers in assay
      Levels(2)=2  !Levels of edge effect (on edge or not)
      Levels(3)=5  !Number of flower counters
      Levels(4)=2  !Number of maternal chambers

c     Read data from a text file into an array
      Open (unit=5,name='DataFile.txt',status='old')
      Do 20 i=1,numobs
        Read(5,*) Info(i,1),Info(i,2),Info(i,3),Info(i,4),Info(i,5),
     &     Info(i,6),Info(i,7),Info(i,8),Info(i,9),Info(i,10)
   20 Continue

      Close (Unit=5)

      Return
      End


C*********************************************************************C
*     Subroutine CalcFixedEffects--                                   *
*     Calculates fixed effect estimates from Generation 1 data        *
C*********************************************************************C
      Subroutine CalcFixedEffects(Info,numobs,NumFix,Levels,Negpos)

c     EXPLANATION OF SUBROUTINE VARIABLES
c     GenCol= column number of Info array containing the generation
c           of each plant
c     ChamCol=column number of Info array containing the chamber each
c           plant was grown in during the progeny test
c     EdgeCol=column number of Info array containing an indicator of
c           whether a given plant was grown on the edge of a chamber
c     DatCol= column number of Info array containing the number of
c           flowers on each plant
c     MomCol= column number of Info array containing the chamber
c           in which each plants mother was grown in
c     Tot= total number of generation 1 plants in the file
c     Sum= array containing the sum of flower numbers of all plants
c           with a given level of each fixed effect
c     Count=array containing the number of plants with a given level
c           of each fixed effect
c     Effect=array containing the estimate of each fixed effect
c     Mean= mean flower number of generation 1 plants
c     Ve= variance in flower number among generation 1 plants
c     i,j,k = counters

c     VARIABLE DECLARATIONS
      Integer numobs,NumFix,Levels(10),Count(10,10),GenCol,ChamCol,
     &    EdgeCol,Tot,DatCol,CountCol,MomCol,i,j,k,Negpos(10,10)
     &    Real Sum(10,10),Effect(10,10),Info(numobs,10),Mean,Ve
      Implicit none

c     Set all constants, such as the generations surveyed,
c     and where the appropriate data is stored
      ChamCol=1
      GenCol=3
      MomCol=4
      EdgeCol=5
      DatCol=7
      CountCol=8
      Mean=0
      Tot=0
      Ve=0
```

```
c     Initialize array for storage of fixed effect estimates
      Do 40 i=1,NumFix
        Do 30 j=1,Levels(i)
          Count(i,j)=0
          Sum(i,j)=0
          Effect(i,j)=0
          NegPos(i,j)=0
   30   Continue
   40 Continue

c     For each fixed effect, select observations from the first
c      generation then add observation to a running sum, that will then
c      be used to calculate a deviation from the mean
      Do 70 i=1,Numfix
        Do 60 j=1,Levels(i)
          Do 50 k=1,numobs
            If (int(Info(k,GenCol)).eq.1) then
                If (int(Info(k,ChamCol)).eq.j) then
                  Count(1,j)=1+Count(1,j)
                  Sum(1,j)=Sum(1,j)+Info(k,DatCol)
                End If

                If (int(Info(k,EdgeCol)).eq.j) then
                  Count(2,j)=Count(2,j)+1
                  Sum(2,j)=Sum(2,j)+Info(k,DatCol)
                End If

                If (int(Info(k,CountCol)).eq.j) then
                  Count(3,j)=Count(3,j)+1
                  Sum(3,j)=Sum(3,j)+Info(k,DatCol)
                End If

                If (int(Info(k,MomCol)).eq.j) then
                  Count(4,j)=Count(4,j)+1
                  Sum(4,j)=Sum(4,j)+Info(k,DatCol)
                End If
            End If
   50     Continue
   60   Continue
   70 Continue

c     Calculate the overall trait mean in the first generation
      Do 85 i=1,numobs
        If (int(Info(i,GenCol)).eq.1) then
          Mean=Mean+Info(i,DatCol)
          Tot=Tot+1
        End If
   85 Continue
      Mean=Mean/Tot
      Write(*,*) 'Mean=',Mean,'Tot=',Tot


c     Calculate the variation in the trait at the first generation
      Do 86 i=1,numobs
        If (int(Info(i,GenCol)).eq.1) then
          Ve=Ve+((Info(i,DatCol)-Mean)**2)
        EndIf
   86 Continue
      Ve=Ve/(Tot-1)
      Write(*,*) 'Ve=',Ve

c     Calculate fixed effect estimates by subtracting the mean trait
c       value of observations with a given fixed effect
```

```
c       from the overall mean.
c     Determine the sign of the effect for use in likelihood
      Do 90 i=1,NumFix
        Do 80 j=1,Levels(i)
          Effect(i,j)=(Sum(i,j)/Count(i,j))-Mean
          If (Effect(i,j).gt.0) then
            NegPos(i,j)=1
          Else
            NegPos(i,j)=-1
          End If
          Write(*,*) 'Fixed Effect',i,'Level',j,'=',Effect(i,j),
     &       'Sign=', NegPos(i,j)
   80   Continue
   90 Continue

c     Apply relevant fixed effect estimate from observations with
c      given fixed effects (if not being estimated in the ML)
      Do 95 i=1,numobs
        Info(i,DatCol)=Info(i,DatCol)+Effect(4,Info(i,MomCol))
   95 Continue

         Return
         End
 C*********************************************************************C
 *     Subroutine Run Pikaia--                                         *
 *     Initializes all necessary variables for Pikaia                  *
 *       Calls the Pikaia subroutine                                   *
 *     Prints Pikaia ML estimates to the screen and appends estimates  *
 *       to a file                                                     *
 C*********************************************************************C
      Subroutine RunPikaia(LogLikelihood,Info,numobs,Negpos)

c     SUBROUTINE REQUIREMENTS
c     Requires a file containing random numbers greater than 0, with 1
c     as the first entry (Random.txt)
c     Requires an output file initialized with a 0 (Output.txt)
c     Requires IMSL library for RNUN and RNSET functions

c     EXPLANATION OF SUBROUTINE PARAMETERS
c     i,j,n= counters
c     iseed= random number seed to initialize Pikaia
c     num=
c     numrun= number of times the program has been run
c     seed= vector containing random numbers read from a file
c     n= the number of parameters to be estimated in the likelihood
c     LogLikelihood= the value of the log of the likelihood of a given
c           run of Pikaia
c     ctrl(12)= vector containing control parameters required by Pikaia
c           (see Pikaia sunbroutine)


c     xb= array containing the values of all parameters associated
c           with the maximum log likelihood from a given run of
c           Pikaia
c     fb= maximum log likelihood value obtained by a given run of
c           Pikaia
c     Est= array containing all maximum log likelihood values and
c           associated parameter values from the total number of
c           runs done to date

c     VARIABLE DECLARATIONS
      Integer n,numobs,i,j,status,num,Negpos(10,10),iseed,numrun,
     &   seed(10000)
      Parameter (n=16)
```

```
      Real LogLikelihood,Info(numobs,10),ctrl(12),xb(n),fb
      Real Est(10000,17)
      External LogLikelihood,RNUN,RNSET
      Implicit none

      write(*,*) 'RunPikaia is running'
c     Initialize Pikaia's Control vector
      Do 10 j=1,12
        ctrl(j)=-1  !Sets all control vector parameters to default
   10 Continue
      ctrl(1)=1000 !Number of individuals in Pikaia population
      ctrl(2)=25  !Number of generations in Pikaia


c     Draw a random number to seed Pikaia
      Open (Unit=8,name='Random.txt',status='old')
      Read(8,*) numrun
      Do 15 i=1,10000
        read(8,*) seed(i)
        If (i.eq.numrun) iseed=seed(i)
   15 Continue
      Close (unit=8)
      Open (unit=8,name='Random.txt',status='old')
      Write(8,*) numrun+1
      Do 16 i=1,10000
        Write(8,*) seed(i)
   16 Continue
      Close (unit=8)

      write(*,*) 'ISeed=',iseed
      Call Rninit(iseed)

c     Call the Pikaia subroutine
      Call Pikaia(LogLikelihood,n,ctrl,xb,fb,status,Info,numobs,
     &  Negpos)

c     Print the results to the screen, adjusting Pikaia estimates
c      between 0 and 1 back to the relevant scale
      write(*,*) ' status: ',status
      write(*,*) 'Estimate of u:',xb(1)
      write(*,*) 'Estimate of s:',xb(2)*10
      write(*,*) 'Estimate of ancestral mean:',xb(3)*1000
      write(*,*) 'Estimate of error variance:',xb(4)*10000
      write(*,*) 'Estimate of Chamber Effect:'
      write(*,*) '   Level 1:',Negpos(1,1)*xb(5)*30
      write(*,*) '   Level 2:',Negpos(1,2)*xb(6)*30
      write(*,*) '   Level 3:',Negpos(1,3)*xb(7)*30
      write(*,*) '   Level 4:',Negpos(1,4)*xb(8)*30
      write(*,*) '   Level 5:',Negpos(1,5)*xb(9)*30
      write(*,*) 'Estimate of Counter Effect:'
      write(*,*) '   Counter 1:',Negpos(3,1)*xb(10)*50
      write(*,*) '   Counter 2:',NegPos(3,2)*xb(11)*50
      write(*,*) '   Counter 3:',Negpos(3,3)*xb(12)*50
      write(*,*) '   Counter 4:',Negpos(3,4)*xb(13)*50
      write(*,*) '   Counter 5:',Negpos(3,5)*xb(14)*50
      write(*,*) 'Estimate of Edge Effect:'
      write(*,*) '   Edge:    ', Negpos(2,1)*xb(15)*30
      write(*,*) '   Not Edge:', Negpos(2,2)*xb(16)*30
      write(*,*) 'Maximum likelihood value:',fb
      write(*,20) ctrl
   20 format(   '    ctrl: ',6f9.5/10x,6f9.5)

c     Open file to store estimates
      Open(unit=7,name='Output.txt',status='old')
```

```
      Read(7,*) Num  !Number of estimate sets in the file
      write(*,*) 'Num=',Num
      If (Num.gt.0) then
        Do 21 i=1,Num !Read estimates from the file into an array
          read(7,*) Est(i,1),Est(i,2),Est(i,3),Est(i,4),Est(i,5),
     &Est(i,6),Est(i,7),Est(i,8),Est(i,9),Est(i,10),Est(i,11),
     &      Est(i,12),Est(i,13),Est(i,14),Est(i,15),Est(i,16),Est(i,17)
   21   Continue
      End If
      Close(unit=7)

c     Add new estimates to array containing previous estimates
      j=Num+1
      Est(j,1)=xb(1)
      Est(j,2)=xb(2)*10
      Est(j,3)=xb(3)*1000
      Est(j,4)=xb(4)*10000
      Est(j,5)=Negpos(1,1)*xb(5)*30
      Est(j,6)=Negpos(1,2)*xb(6)*30
      Est(j,7)=Negpos(1,3)*xb(7)*30
      Est(j,8)=Negpos(1,4)*xb(8)*30
      Est(j,9)=Negpos(1,5)*xb(9)*30
      Est(j,10)=Negpos(3,1)*xb(10)*50
      Est(j,11)=Negpos(3,2)*xb(11)*50
      Est(j,12)=Negpos(3,3)*xb(12)*50
      Est(j,13)=Negpos(3,4)*xb(13)*50
      Est(j,14)=Negpos(3,5)*xb(14)*50
      Est(j,15)=Negpos(2,1)*xb(15)*30
      Est(j,16)=Negpos(2,2)*xb(16)*30
      Est(j,17)=fb


c     Write the number of estimate sets & all estimates back into the
c      file
      Open(unit=7,name='Output.txt',status='old')
      Write(7,*) Num+1
      Do 24 j=1,Num+1
        Write(7,23) Est(j,1),Est(j,2),Est(j,3),Est(j,4),Est(j,5),
     &    Est(j,6),Est(j,7),Est(j,8),Est(j,9),Est(j,10),Est(j,11),
     &    Est(j,12),Est(j,13),Est(j,14),Est(j,15),Est(j,16),Est(j,17)
   23   Format(F15.4,F15.4,F15.4,F15.4,F15.4,F15.4,F15.4,F15.4,F15.4,
     &    F15.4,F15.4,F15.4,F15.4,F15.4,F15.4,F15.4,F15.4,F15.4)
   24 Continue

      Close(unit=7)

      Return
      End
```

```
C*********************************************************************C
*    Function LogLikelihood--                                         *
*     Calculates the likelihood for each family and the data set as   *
*     a whole. These likelihood values are then relayed to Pikaia     *
*      subroutine that attempts to maximize the loglikelihood   *
C*********************************************************************C
      Function LogLikelihood(n,x,Info,numobs,Negpos)
```

```
c     FUNCTION REQUIREMENTS
c     Requires IMSL Library for POIPR function

c     EXPLANATION OF FUNCTION VARIABLES
c     family= line number/family number assigned to a data point
c     count= number of individuals of a given family in the data
c     FamCount= number of individuals of a given family in the data
c     row= row of Info array
c     GenCol= column number of Info array containing the generation
c           of each plant
c     ChamCol=column number of Info array containing the chamber each
c           plant was grown in during the progeny test
c     EdgeCol=column number of Info array containing an indicator of
c           whether a given plant was grown on the edge of a chamber
c     DatCol= column number of Info array containing the number of
c           flowers on each plant
c     FamCol= column number of Info array containing the line/family
c           number of each plant
c     CounterCol= column number of Info array containing a number
c           representing the person counting the plant's flowers
c     Chamber= vector containing adjusted chamber effect estimates
c           from Pikaia
c     Counter= vector containing adjusted counter effect estimates
c           from Pikaia
c     Edge= vector containing adjusted edge effect estimates from
c           Pikaia
c     MaxFam= highest line/family number in the data set
c     FirstGen= the first MA generation assayed
c     LastGen= the last MA generation assayed
c     M= estimate of ancestral mean
c     s= estimate of average mutational effect
c     u= estimate of the mutation rate per genome per generation
c     Ve= estimate of variance
c     Exponent= contains the exponential component of the normal
c           probability function
c     Z= vector containing data for a family, adjusted for relevant
c           fixed effects
c     FamGen= vector containing generation number for each individual
c           of a given family
c     x= vector containing the parameter values generated in Pikaia,
c           to determine their likelihood
c     pi= parameter assigned the value of pi
c     NormalProb= contains the normal probability portion of the
c           mutational model
c     Lambdai= parameter of the Poisson distribution for FirstGen
c           generations of MA
c     Lambdak= parameter of the Poisson distribution for LastGen
c           generations of MA
c     Product= product of the normal and Poisson probabilities for
c           each plant
c     FamLikeli= the likelihood of each family
c     LogLike= the running total of the sum of the log likelihoods of
c           each family
c     LogLikelihood= total log likelihood of the data set given the
c           parameters supplied by Pikaia

c     VARIABLE DECLARATIONS
      Integer numobs,family,count,FamCount,row,n,GenCol,DatCol,FamCol,
     &   MaxFam,FirstGen,LastGen,i,k,ChamCol,CounterCol,EdgeCol,
     &   Negpos(10,10)
      Real M,s,u,Ve,Exponent,Info(numobs,10),Z(5000),FamGen(5000),
     &   x(n),pi,NormalProb,LogLike,POIPR,Lambdai,Lambdak,LogLikelihood,
     &   Chamber(5),Counter(5),Edge(2)
```

```
      DoublePrecision Product,FamLikeli
      Parameter (pi=3.141592654)
      External POIPR
      Implicit none


c     Set all constants, such as the generations surveyed,
c      and where the appropriate data is stored
      FirstGen=1
      LastGen=11
      ChamCol=1
      FamCol=2
      GenCol=3
      EdgeCol=5
      DatCol=7
      CounterCol=8
      MaxFam=Info(numobs,FamCol)

c     Set parameters for likelihood based on Pikaia supplied parameters
c      Pikaia parameters range from zero to one, so they must be
c      adjusted to the relevant scale
      u=x(1)
      s=x(2)*10
      M=x(3)*1000
      Ve=x(4)*10000
      Chamber(1)=negpos(1,1)*(x(5)*30)
      Chamber(2)=negpos(1,2)*(x(6)*30)
      Chamber(3)=negpos(1,3)*(x(7)*30)
      Chamber(4)=negpos(1,4)*(x(8)*30)
      Chamber(5)=negpos(1,5)*(x(9)*30)
      Counter(1)=negpos(3,1)*(x(10)*50)
      Counter(2)=negpos(3,2)*(x(11)*50)
      Counter(3)=negpos(3,3)*(x(12)*50)
      Counter(4)=negpos(3,4)*(x(13)*50)
      Counter(5)=negpos(3,5)*(x(14)*50)
      Edge(1)=negpos(2,1)*(x(15)*30)
      Edge(2)=negpos(2,2)*(x(16)*30)
      LogLike=0  !Initialize LogLikelihood value

c     If mutation rate estimate is less than zero (non-sensical),
c      set the likelihood to a very low value...
      If (u.le.0) then
        LogLikelihood=-100000000
      Else
c     But if the mutation rate estimate is greater than zero,
c      Calculate the likelihood for each line
        Do 600 family=1,MaxFam
         count=0
         FamLikeli=0
         Product=1

c     Apply all necessary fixed effect estimates,
c      by adding effect estimate to the observation
          Do 100 row=1,numobs
            If (INT(Info(row,FamCol)).eq.family) then
              count=count+1
             Z(count)=Info(row,DatCol)
             Do 50 i=1,5
               If (Info(row,ChamCol).eq.i)then
                 Z(count)=Z(count)+Chamber(i)
               End If
               If (Info(row,CounterCol).eq.i) then
                 Z(count)=Z(count)+Counter(i)
               EndIf
```

```
              If (Info(row,EdgeCol).eq.i) then
                Z(count)=Z(count)+Edge(i)
              EndIf
   50       Continue
            FamGen(count)=(INT(Info(row,GenCol)))
          End If
  100   Continue

c     Calculate Lambda values for the Poission probability(Lambda=u/2*t)
Lambdai=FirstGen*(u/2)
        Lambdak=(LastGen-FirstGen)*(u/2)

c     Now calculate the likelihood, using formula based on Keightley &
c      Batallion
        Do 500 i=0,1
          Do 300 k=0,5
            Do 200 FamCount=1,count
              If (FamGen(FamCount).eq.FirstGen) then
                Exponent=((Z(FamCount)+(i*s)-M)**2)/(2*Ve)
              Else If (FamGen(FamCount).eq.LastGen) then
                Exponent=((Z(FamCount)+((i+k)*s)-M)**2)/(2*Ve)
              Else
                Write(*,*)'ERROR - invalid generation number'
              End If
              NormalProb=(1/(sqrt(2.*pi*Ve))*exp(-Exponent))
              Product=Product*NormalProb
  200       Continue
            Product=Product*POIPR(i,Lambdai)*POIPR(k,Lambdak)
            FamLikeli=FamLikeli+Product
            Product=1
  300     Continue
  500   Continue

c     Add LogLikelihood for a family to the total
        LogLike=LogLike+log(FamLikeli)
  600   Continue

c     Return LogLikelihood value to Pikaia
       LogLikelihood=LogLike
      End If

      Return
      End

c*********************************************************************c
      subroutine Pikaia(LogLikelihood,n,ctrl,x,f,status,Info,numobs,
      Negpos)
c=====================================================================c
c     Optimization (maximization) of user-supplied "fitness" function
c     ff  over n-dimensional parameter space  x  using a basic genetic
c     algorithm method.
c
c     Paul Charbonneau & Barry Knapp
c     High Altitude Observatory
c     National Center for Atmospheric Research
c     Boulder CO 80307-3000
c     <paulchar@hao.ucar.edu>
c     <knapp@hao.ucar.edu>
c
c     Version 1.0   [ 1995 December 01 ]
c
c     Genetic algorithms are heuristic search techniques that
c     incorporate in a computational setting, the biological notion
c     of evolution by means of natural selection.  This subroutine
```

```
c     implements the three basic operations of selection, crossover,
c     and mutation, operating on "genotypes" encoded as strings.
c
c     See References:
c
c        Charbonneau, Paul.  "Genetic Algorithms in Astronomy and
c           Astrophysics."  Astrophysical J. (Supplement), vol 101,
c           in press (December 1995).
c
c        Goldberg, David E.  Genetic Algorithms in Search, Optimization,
c           & Machine Learning.  Addison-Wesley, 1989.
c
c        Davis, Lawrence, ed.  Handbook of Genetic Algorithms.
c           Van Nostrand Reinhold, 1991.
c
c     For code, please visit:
c     http://www.hao.ucar.edu/public/research/si/pikaia/pikaia.html

c     USES: ff, urand, setctl, report, rnkpop, select, encode, decode,
c           cross, mutate, genrep, stdrep, newpop, adjmut
      implicit none

c     Input:
      integer   n,numobs,Negpos(10,10)
      real      LogLikelihood,Info(numobs,10)
      external  LogLikelihood
c
c      o Integer  n  is the parameter space dimension, i.e., the number
c        of adjustable parameters.
c
c      o Function  ff  is a user-supplied scalar function of n vari-
c        ables, which must have the calling sequence f = ff(n,x), where
c        x is a real parameter array of length n.  This function must
c        be written so as to bound all parameters to the interval [0,1];
c        that is, the user must determine a priori bounds for the para-
c        meter space, and ff must use these bounds to perform the appro-
c        priate scalings to recover true parameter values in the
c        a priori ranges.
c
c        By convention, ff should return higher values for more optimal
c        parameter values (i.e., individuals which are more "fit").
c        For example, in fitting a function through data points, ff
c        could return the inverse of chi**2.
c
c        In most cases initialization code will have to be written
c        (either in a driver or in a separate subroutine) which loads
c        in data values and communicates with ff via one or more
c        labeled
c        common blocks.  An example exercise driver and fitness
c        function
c        are provided in the accompanying file, xpkaia.f.
```