

# FITNESS OF *ARABIDOPSIS THALIANA* MUTATION ACCUMULATION LINES WHOSE SPONTANEOUS MUTATIONS ARE KNOWN

Matthew T. Rutter,<sup>1</sup> Angela Roles,<sup>2</sup> Jeffrey K. Conner,<sup>3</sup> Ruth G. Shaw,<sup>4</sup> Frank H. Shaw,<sup>5</sup> Korbinian Schneeberger,<sup>6,7</sup> Stephan Ossowski,<sup>6,8</sup> Detlef Weigel,<sup>6</sup> and Charles B. Fenster<sup>9,10</sup>

<sup>1</sup>Department of Biology, College of Charleston, Charleston, South Carolina 29401

<sup>2</sup>Department of Biology, Oberlin College, Oberlin, Ohio 44074

<sup>3</sup>Kellogg Biological Station and Department of Plant Biology, Michigan State University, Hickory Corners, Michigan 49060

<sup>4</sup>Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, Minnesota 55108

<sup>5</sup>Department of Mathematics, Hamline University, St. Paul, Minnesota 55104

<sup>6</sup>Department of Molecular Biology, Max Planck Institute for Developmental Biology, Tübingen 72076, Germany

<sup>9</sup>Department of Biology, University of Maryland, College Park, Maryland 20742

<sup>10</sup>E-mail: cfenster@umd.edu

Received December 1, 2011

Accepted December 30, 2011

Data Archived: Dryad doi:10.5061/dryad.rk5fp6n9

Despite the fundamental importance of mutation to the evolutionary process, we have little knowledge of the direct consequences of specific spontaneous mutations to the fitness of the organism. Combining results of whole-genome sequencing with repeated field assays of survival and reproduction, we quantify the combined effects on fitness of spontaneous mutations identified in *Arabidopsis thaliana*. We demonstrate that the effects are beneficial, deleterious, or neutral depending on the environmental context. Some lines, bearing mutations disrupting known loci, differ strongly in fitness from the founder or premutation genotype. Those effects vary across environments, for example, a line with a major deletion spanning a transcription factor gene expressed lower fitness than the founder under most conditions but exceeded the founder's fitness in one environment. The large contribution of genotype by environment interaction ( $G \times E$ ) to mutation effects on fitness implies spatial and/or temporal variation in selection on new mutations and could contribute to the maintenance of standing genetic variation.

**KEY WORDS:** Genetic variation,  $G \times E$ , mutation, quantitative genetics.

New mutations are the original source of all genetic variation. Mutation has a central role in evolutionary theory (Fisher 1930; Kondrashov 1988; Lynch et al. 1999; Johnson and Barton 2005; Martin and Lenormand 2006) and in conservation biology (Lande

1994; Lynch et al. 1995), but empirical challenges hamper the study of its contribution to adaptation and variation in natural populations. Although effects on the fitness of known mutations singly and in pairs have been evaluated for certain viruses (Sanjuan et al. 2004a,b), the cumulative effects of known spontaneous mutations on the fitness of multicellular organisms have never been directly quantified (Bataillon 2003).

Mutation accumulation (MA) lines, which are independent lines derived from a nearly homozygous founder, can be used to estimate both the rate of mutation and the distribution of their

<sup>7</sup>Current address: Department of Plant Developmental Biology, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany

<sup>8</sup>Current address: CRG Centre for Genomic Regulation, Dr. Aiguader 88, Barcelona 08003, Spain

effects on traits and fitness (Mukai 1964; Bataillon 2000; Shaw et al. 2002). Shaw et al. (2002) and Rutter et al. (2010) used a set of MA lines derived from the reference *Columbia* accession of *Arabidopsis thaliana* to quantify the mutation rate and distribution of effects on reproductive components of fitness and on lifetime fitness (survivorship  $\times$  reproduction), respectively. More recently, Ossowski et al. (2010) reported the complete genome sequences of five of the lines for which Rutter et al. (2010) obtained estimates of lifetime fitness in the greenhouse and in one field environment. Here, we present new performance results for five additional field environments; four of the five sequenced lines are represented in all five new environments whereas the fifth sequenced line's performance was measured in three of the five new environments. We synthesize the results obtained via statistical genetics and molecular approaches, respectively, to examine the dependence of the fitness of the five MA lines in different environments on sequence differences.

## Methods

The methods of the sequence analysis of the five MA lines are presented in Ossowski et al. (2010), whereas the methods of the assessment of mutation rate and effect of mutations on fitness under the conditions of the one field experiment are presented in Rutter et al. (2010). Because results from only one of the six field trials are presented in Rutter et al. (2010), we briefly describe how the field tests were performed in all of the environments.

Using seed derived from the 25th generation of MA, incorporating sublines to account for maternal effects (five sublines per MA line), and employing a randomized block design ( $N = 14$  blocks per experiment), we planted seedlings of 100 MA lines and the premutation genotype at the two-leaf stage, approximately two weeks post germination, into a secondary successional field at Blandy Experimental Farm (BEF) (39°N, 78°W), in the northern Blue Ridge Mountains of Virginia. The *Columbia* founder premutation genotype had been stored at 4°C whereas the MA lines were advanced. At the time when sublines of the MA lines were generated for these and other assays, sublines were also produced for the premutation founder by retrieving founder generation seed from cold storage and growing them for two generations. The five MA lines from this set that were sequenced by Ossowski et al. (2010) were included in our field tests. Seedlings were planted in the Spring and Fall of 2004 and again in the Spring and Fall of 2005. The Spring and Fall plantings allow the evaluation of mutation parameters for different life-histories of *A. thaliana* populations, depending on environment. The Spring-planting environment corresponds to a spring ephemeral life-history, where plants germinate and complete the life-cycle in the spring. The Fall-planting environment

corresponds to a winter-annual life-history, with plants germinating in the fall, overwintering as rosettes, and flowering and fruiting in the ensuing spring. In separate experiments, the premutation founder and 50 MA lines, including four of the five sequenced MA lines, were planted as seeds in Fall of 2005, in replicate at BEF and the Kellogg Biological Station (KBS) (42°N, 85°W) in southern Michigan.

For all but one environment, the measure of plant fitness was lifetime fruit production, including zero values for plants that did not survive to reproduce, a commonly used measure in fitness assays and especially appropriate for a predominantly selfing annual plant, although it does not take into account variation in the survival and germination of seeds in the seedbank (seed plantings) or variation in the survival and germination of seeds (seedling plantings). In the Fall-2004 planting, fruit number was estimated from biomass, as biomass was highly correlated with fruit number in a random subset of this dataset ( $r^2 = 0.96$ ) and overall fruit numbers were extremely high. We used aster modeling (Shaw et al. 2008) to compare the fitness of the MA lines and the founder within each of the environments and to quantify genotype by environment interaction ( $G \times E$ ) of performance of the lines across environments. The analysis of  $G \times E$  was conducted with random block and subline effects. The comparison of the MA line performance with the founder line was performed without the use of random effects.

## Results and Discussion

The five sequenced *A. thaliana* MA lines provided a point mutation rate estimate of 0.60/haploid genome and generation (0.70 including indels; Ossowski et al. 2010) similar to direct sequencing studies of MA lines in *Caenorhabditis elegans* and *Drosophila melanogaster* (Denver et al. 2004; Haag-Liautard et al. 2007). The MA lines were sequenced at the 30th generation, and had on average 20 point and 3-indel mutations. Assuming mutations accumulate linearly with generations, the 25th generation MA individuals, which we assayed for fitness, had on average 17 point and 2.5-indel mutations. Because the 25th generation should harbor over 80% of mutations found at the 30th generation, our fitness estimates represent most mutations captured by sequencing. Rutter et al. (2010) estimated the diploid whole genomic mutation rate for fitness measured in the field as 0.23, similar to a previous study of a subset of the same lines when grown in the greenhouse (Shaw et al. 2002), and similar to the diploid mutation rate of nonsynonymous and indel mutations occurring in coding regions, 0.2, estimated by Ossowski et al. (2010). However, unlike Shaw et al. (2002), Rutter et al. (2010) did not detect MA line divergence in the greenhouse.

Because many MA lines outperformed the premutation genotype in the field, Rutter et al. (2010) estimated a high proportion of

these mutations, 43%, as enhancing fitness, similar to estimates for components of reproduction in Shaw et al. (2002; see also MacKenzie et al. 2005). Converting the 0.23 diploid mutation rate for fitness to a haploid estimate, 0.12, and comparing with the haploid sequence level point mutation rate, we deduce that 20% (0.12/0.60) of the mutations determined by sequencing have fitness effects; 9% ( $0.43 \times 0.20$ ) of the spontaneous mutations are beneficial, 11% ( $0.57 \times 0.20$ ) are deleterious, and the rest, 80%, are neutral or nearly neutral ( $N_e s < 1$ , Wright 1931). Our estimate of the deleterious mutation rate of 11% or 0.07/haploid generation is less than the value of 0.14/haploid generation derived from sequence comparisons of *A. thaliana* and *A. lyrata* in Ossowski et al. (2010). One of many explanations for this difference is that estimates of the whole genomic deleterious rate using sequence comparison approaches assume mutations to be deleterious or neutral, whereas we detected a relatively high proportion of beneficial mutations.

Supporting the existence of an appreciable proportion of beneficial mutations, the five sequenced MA lines frequently outperformed the founder in five additional field trials under an array of conditions (Table 1). Others have recently detected high rates of beneficial mutations, up to 15% in viruses (Burch and Chao 1999, 2000; Silander et al. 2007), up to 13% in yeast (Hall et al. 2008, Hall and Joseph 2010), and a significant number in *D. melanogaster* (Azad et al. 2010), counter to the view of Keightley and Lynch (2003).

Although previous greenhouse studies varying specific aspects of environment have demonstrated modest G×E for these MA lines (Chang and Shaw 2003; Kavanaugh and Shaw 2005), we found the performance of the MA lines in the field was highly variable among environments, which varied both temporally or spatially (Table 1). Based on the performance of the five se-

quenced MA lines across the four seedling plantings in BEF, two in the Fall environment and two in the Spring environment, we found a significant interaction between MA line and experiment ( $P < 0.0025$ ). Comparing MA lines to the founder across all of the environments, for each MA line estimates of fitness were higher than that of the premutation line in at least one environment, and all MA lines (except line 119) had lower fitness than the premutation line in at least one environment (Table 1). Taken as a whole, it is clear that the effects of the mutations accumulated during the experiment varied across environments, generating a genotype by environment interaction for fitness.

Although we cannot isolate the effects of individual mutations found in any given line, and Ossowski et al. (2010) detected only the majority of all mutations, despite their sequence coverage of 90% of the genome, some general conclusions can be made regarding the mutational composition and the performance of the MA lines (Table 1 and Supporting information). Counting only mutations in UTRs, introns, and coding regions of putative genes, i.e., in or near putative genes, MA line 29 had the fewest nonsynonymous mutations (one), no deletions in coding regions, and performed most similarly to the founder across environments. This finding is consistent with expectations that nonsynonymous mutations and deletions would have the strongest fitness effects and their absence should result in little fitness change from the founder. Unexpectedly, the MA lines with the most mutations overall (line 59) and the most mutations associated with known coding loci (line 119) outperformed the founder in the greatest number of trials. However, this is consistent with our inference of a high rate of beneficial mutations (Shaw et al. 2002; Rutter et al. 2010). Because the five MA lines were a random sample of the 100 lines for which we tested performance, we expect them to

**Table 1.** Mean fruit production of the five MA lines and the founder premutation line across a range of environments. Fitness were estimated using an aster model including survival (binomial) and fruit number (Poisson). *P*-values ( $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ) represent MA-founder comparisons using separate analyses of aster for each MA-founder comparison. *P*-values were calculated by likelihood ratio tests, and validated using a parametric bootstrap. Means in bold represent a significant difference following within experiment sequential Bonferroni correction ( $P < 0.05$ ). A summary of the mutations found in each line is presented in the last column. Spring 2004 is part of a larger data set published in Rutter et al. (2010). Model abbreviations are BEF for Blandly Experimental Farm and KBS for Kellogg Biological Station.

Line	Fall '04	Fall '05	Seed BEF	Seed KBS	Spring '04	Spring '05	Mutational profile		
							Total	Total Nonsynonymous + Indels	Total Coding region
Founder	287.7	19	30.0	0.44	17.7	28.6			
29	314.7	16.4	–	–	18.6	26.3	21	6	4
49	279.7	26.3	<b>15.5*</b>	0.44	13.8*	20.1	26	6	8
59	363.4	16.4	<b>59.9***</b>	<b>1.24***</b>	19.5	24.7	30	11	7
69	374.4	8.8	25.9	0.27	<b>22.9**</b>	21.2	21	4	3
119	446.3*	33.1*	<b>48.7**</b>	0.54	19.7	31.6*	27	7	5

be a representative sample of the 100 MA lines across the different environments. Indeed, across the six field environments the lines had intermediate performance, ranging from a mean 35th–75th percentile, with greater performance indicated by a lower percentile rank.

We discuss below specific mutations within particular lines in relation to the fitness profiles of those lines, acknowledging that the effects of multiple mutations within a line are confounded (Supporting information).

A mutation located in a locus encoding a gene associated with DNA replication and repair (AT4G 00070: RING/U box family) found in line 59 may account for this line having the most mutations. We do not have an explanation of why or which mutations were beneficial in the specific environmental context, with one possible exception. MA line 119 outperformed the founder line in all field experiments and had a large 600 bp deletion in a gypsy-class retrotransposon (AT3G60930). Such a mutation, occurring in a sequence that is likely to be deleterious, may be beneficial. In contrast, it is easier to hypothesize which mutations may have deleterious effects, keeping in mind that the significant  $G \times E$  was often associated with the relative performance of individual MA lines changing with respect to the founder line. Line 49 had the largest deletion, spanning three loci (AT1G77420–450), and was the most consistently poorly performing line, frequently exhibiting the lowest (4 of 6) or second lowest (1 of 6) fitness of any line (including the premutation line), congruent with the expectation that deletions of coding loci are often deleterious. This deletion in line 49 included sequences coding for a hydrolase (AT1G77420), a protein degrading enzyme (AT1G77440), and a DNA binding transcription factor associated with transcriptional regulation (AT1G77450). Which, if any, of these gene deletions have deleterious effects on fitness cannot as yet be determined, but the mutation affecting the transcription factor is intriguing in view of the conjecture that mutations in the coding regions of *trans*-regulatory loci are more likely to have deleterious pleiotropic consequences than are mutations in other loci (Stern 2000; Carroll 2008).

However, even MA line 49 performed better than the founder in one environment, despite the three-locus deletion (Table 1, Fall-2005 planting), demonstrating that effects of mutations may differ substantially among environments (Kondrashov and Houle 1994; Shabalina et al. 1997; Davies et al. 1999; Yang et al. 2001; Kulheim et al. 2002; Roles and Conner 2008).

Our joint findings indicate that mutations affect fitness, both positively and negatively, and these effects appear to differ with environment. Rutter et al. 2010 found that mutation effects on fitness are small, when scaled to environmental variation ( $h^2_m$  of the mutations in one field environment = 0.0001). Together, the environment dependence of mutation effects on fitness (spanning deleterious to beneficial) (Gillespie and Turelli, 1989) and the

relatively small effects of individual mutations on fitness may contribute to the great degree of standing genetic variation found in natural populations.

#### ACKNOWLEDGMENTS

This work was funded by the Max Planck Society to D.W., NSF grant (DEB 0108354) to J.K.C., NSF grants (DEB 9629457 and DEB 9981891) to R.G.S., and NSF grants (DEB 0315972, DEB 0307180, and DEB 0845413) to M.R. and C.F. The manuscript was improved by comments from M. Dudash, J. Kostyun, and F. Stearns. We thank C. Geyer for consultation on the analysis. The UMD evolutionary genetics group provided additional helpful input.

#### LITERATURE CITED

- Azad, P., M. C. Zhang, and R. C. Woodruff. 2010. Rapid increase in viability due to new beneficial mutations in *Drosophila melanogaster*. *Genetica* 138:251–263.
- Bataillon, T. 2000. Estimation of spontaneous genome-wide mutation rate parameters: whither beneficial mutations? *Heredity* 84:497–501.
- . 2003. Shaking the ‘deleterious mutations’ dogma? *Trends Ecol. Evol.* 18:315–317.
- Burch, C. L., and L. Chao. 1999. Evolution by small steps and rugged landscapes in the RNA virus phi 6. *Genetics* 151:921–927.
- . 2000. Evolvability in an RNA virus is determined by its mutational neighbourhood. *Nature* 406:625–629.
- Carroll, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 135:25–36.
- 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.
- Davies, E. K., A. D. Peters, and P. D. Keightley. 1999. High frequency of cryptic mutations in *Caenorhabditis elegans*. *Science* 285:1748–1751.
- Denver, D. R., K. Morris, M. Lynch, and W. K. Thomas. 2004. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* 430:679–682.
- Fisher, R. A. 1930. *The genetical theory of natural selection*. Oxford Univ. Press, Oxford, United Kingdom.
- Gillespie, J. H., and M. Turelli. 1989. Genotype-environment interactions and the maintenance of polygenic variation. *Genetics* 121:129–138.
- Haag-Liautaud, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan, B. Charlesworth, and P. D. Keightley. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* 445:82–85.
- Hall, D. W., and S. B. Joseph. 2010. A high frequency of mutations across multiple fitness components in *Saccharomyces cerevisiae*. *Genetics* 185:1397–1409.
- Hall, D. W., R. Mahmoudizad, A. W. Hurd, and S. B. Joseph. 2008. Spontaneous mutations in diploid *Saccharomyces cerevisiae*: another thousand cell generations. *Genet. Res. Camb.* 90:229–241.
- Johnson T., and Barton N. 2005. Theoretical models of selection and mutation on quantitative traits. *Philos. Trans. R. Soc. B.* 360:1411–1425.
- Kavanaugh, C. M., and R. G. Shaw. 2005. The contribution of spontaneous mutation to variation in environmental responses of *Arabidopsis thaliana*: responses to light. *Evolution* 59:266–275.
- Keightley, P. D., and M. Lynch. 2003. Towards a realistic model of mutations affecting fitness. *Evolution* 57:683–685.

- 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*. *Philos. Trans. R. Soc. B.* 258:221–227.
- Kulheim, C., J. Agren, and S. Jansson. 2002. Rapid regulation of light harvesting and plant fitness in the field. *Science* 297:91–93.
- Lande, R. 1994. Risk of population extinction from fixation of new deleterious mutations. *Evolution* 48:1460–1469.
- Lynch, M., J. Connery, and R. Bürger. 1995. Mutation accumulation and the extinction of small populations. *Am. Nat.* 146:489–518.
- Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz, L. Vassilieva, and J. Willis. 1999. Perspective: spontaneous deleterious mutation. *Evolution* 53:645–663.
- Martin, G., and T. Lenormand. 2006. The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution* 60:2413–2427.
- MacKenzie, J. L., F. E. Saade, Q. H. Le, T. E. Bureau, and D. J. Schoen. 2005. Genomic mutation in lines of *Arabidopsis thaliana* exposed to ultraviolet-B radiation. *Genetics* 171:715–723.
- Mukai, T. 1964. The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* 50:1–19.
- Ossowski, S., K. Schneeberger, J. I. Lucas-Lledó, N. Warthmann, R. M. Clark, R. G. Shaw, D. Weigel, and M. Lynch. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327:92–94.
- Roles, A. J., and J. K. Conner. 2008. Fitness effects of mutation accumulation in a natural outbred population of wild radish (*Raphanus raphanistrum*): comparison of field and greenhouse environments. *Evolution* 62:1066–1075.
- Rutter, T., F. H. Shaw, and C. B. Fenster. 2010. Spontaneous mutation parameters for *Arabidopsis thaliana* measured in the wild. *Evolution* 64:1825–1835.
- Sanjuán, R., A. Moya, and S. F. Elena. 2004a. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. USA* 101:8396–8401.
- . 2004b. The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl. Acad. Sci. USA* 101:15376–15379.
- 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. *Proc. Natl. Acad. Sci. USA* 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., C. J. Geyer, S. Wagenius, H. M. Hangelbroek, and J. R. Etgerson. 2008. Unifying life history analyses for inference of fitness and population growth. *Am. Nat.* 172:E35–E47.
- Silander, O. K., O. Tenaillon, and L. Chao. 2007. Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biol.* 5:922–931.
- Stern, D. L. 2000. Perspective: evolutionary developmental biology and the problem of variation. *Evolution* 54:1079–1091.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–150.
- Yang, H. P., A. Y. Tanikawa, W. A. Van Voorhies, J. C. Silva, and A. S. Kondrashov. 2001. Whole-genome effects of ethyl methanesulfonate-induced mutation on nine quantitative traits in outbred *Drosophila melanogaster*. *Genetics* 157:1257–1265.

Associate Editor: J. Kelly

## Supporting Information

The following supporting information is available for this article:

**Table S1.** Spontaneous mutations found in five mutation accumulation lines in or near loci with putative function.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.