doi:10.1111/j.1558-5646.2008.00423.x

HOW REPEATABLE IS ADAPTIVE EVOLUTION? THE ROLE OF GEOGRAPHICAL ORIGIN AND FOUNDER EFFECTS IN LABORATORY ADAPTATION

Pedro Simões,[1,2] Josiane Santos,[1,3] Inês Fragata,[1,4] Laurence D. Mueller,[5,6] Michael R. Rose,[5,7] and Margarida Matos[1,8]

[1]*Centro de Biologia Ambiental, Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal*

[2]*E-mail: pmsimoes@fc.ul.pt*

[3]*E-mail: jmssantos@fc.ul.pt*

[4]*E-mail: if@fc.ul.pt*

[5]*Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697*

[6]*E-mail: ldmuelle@uci.edu*

[7]*E-mail: mrrose@uci.edu*

[8]*E-mail: mmatos@fc.ul.pt*

Received June 20, 2007
Accepted April 28, 2008

The importance of contingency versus predictability in evolution has been a long-standing issue, particularly the interaction between genetic background, founder effects, and selection. Here we address experimentally the effects of genetic background and founder events on the repeatability of laboratory adaptation in *Drosophila subobscura* populations for several functional traits. We found disparate starting points for adaptation among laboratory populations derived from independently sampled wild populations for all traits. With respect to the subsequent evolutionary rate during laboratory adaptation, starvation resistance varied considerably among foundations such that the outcome of laboratory evolution is rather unpredictable for this particular trait, even in direction. In contrast, the laboratory evolution of traits closely related to fitness was less contingent on the circumstances of foundation. These findings suggest that the initial laboratory evolution of weakly selected characters may be unpredictable, even when the key adaptations under evolutionary domestication are predictable with respect to their trajectories.

KEY WORDS: **Adaptation, *Drosophila subobscura*, evolutionary contingency, founder effects, genetic background, life-history traits, repeatability.**

Evolutionary contingencies can be a source of differentiation among populations (Travisano et al. 1995; Joshi et al. 2003). In particular, differences in adaptive dynamics have been shown between populations with different ancestors that share a common environment in which they undergo subsequent adaptation (Cohan 1984a; Cohan and Hoffmann 1986, 1989). Several factors may be involved in such contingent differentiation. Small differences in the course of selection might result in substantially different evolutionary outcomes, particularly in populations studied without good environmental controls. Different genetic backgrounds may

explain disparate adaptive responses among populations exposed to the same selective pressures (Cohan and Hoffmann 1989; de Brito et al. 2005), particularly in carefully controlled laboratory studies.

Such differences among genetic backgrounds may specifically involve differences in additive genetic components of variance and covariance, nonadditive interactions among loci, or linkage disequilibrium (Lande 1980; Cheverud and Routman 1995; Falconer and Mackay 1996; Roff 2000; Wade et al. 2001; Steppan et al. 2002). These genetic differences may be due both to the past selective history and to random allele frequency changes associated with sampling processes, particularly with bottlenecked population sizes (Bryant et al. 1986; Cheverud et al. 1999; Naciri-Graven and Goudet 2003; Zhang et al. 2004). Furthermore, interactions between directional selection and genetic drift during the evolutionary process might increase the impact of different genetic backgrounds (Cohan 1984b; Cohan and Hoffmann 1989; de Brito et al. 2005).

Experimental evolution studies are particularly well suited to test the relevance of genetic background effects in adaptive evolution, because they allow us to reproducibly measure the adaptive response of replicated populations to defined environmental conditions (Rose et al. 1996; Lenski 2004; Chippindale 2006). The evolutionary responses to a laboratory selection regime can be measured relative to control populations already adapted to the laboratory. Several laboratory experiments have addressed the importance of genetic background effects in *Drosophila* laboratory evolution (Cohan and Hoffmann 1986; Teotónio and Rose 2000; Teotónio et al. 2002; Joshi et al. 2003; see reviews by Prasad and Joshi 2003; Rose et al. 2004). Overall, these studies have found some cases in which different genetic backgrounds clearly lead to different evolutionary patterns during adaptation to a common environment (e.g., Cohan and Hoffmann 1986; Teotónio and Rose 2000).

Our team has focused on the study of adaptation to the laboratory in *Drosophila subobscura*, studying the evolutionary trajectories of several life-history traits in populations collected from different natural locations (see Matos et al. 2000a, 2002). We have found evidence of variation in adaptive response during the first generations of laboratory adaptation between two sets of populations founded from the same natural location six years apart (Matos et al. 2002). In a subsequent study, we analyzed the evolutionary dynamics of two different sets of laboratory populations derived from synchronous foundations obtained in 2001 from two different natural sites: "AR" populations derived from a wild-caught sample from Arrábida, Portugal, and "TW" populations from Sintra, Portugal (see Simões et al. 2007). This study revealed a clear adaptive response in fecundity-related traits as well as significant differences in the evolutionary dynamics of these two sets of populations over their first 14 generations of laboratory culture, most likely as a result of different initial genetic composition of the populations. In spite of the clear results observed in that study, we could not specifically determine whether the differences in evolutionary dynamics observed were due to the genetic composition of the wild source populations or to sampling effects arising at the time of sample collection. Differences in genetic variation due to a limited number of field-collected individuals are expected to affect the subsequent evolutionary response of populations (James 1971; Powell and Richmond 1974; Reznick and Ghalambor 2005). Nevertheless, the importance of possible sampling effects for variation in the laboratory adaptation process is not easily predictable.

In this study we aim to test the repeatability of an adaptive process and the impact of contingent factors, such as chance events affecting the initial genetic background that may occur due to either geographical or temporal shifts in the genetic composition of populations. For this purpose, we present a broad analysis of the initial stages of laboratory evolution combining data from different sets of populations obtained from collections across different years (1998, 2001, and 2005) and different geographical locations (Arrábida and Sintra). We also compare the sensitivity of adaptive processes to differences arising from the geographical location of wild source populations (Arrábida vs. Sintra) with the impact of sampling effects among populations derived from the same wild source, in synchronous evolving populations (2005 populations).

The main questions addressed in this study are:

(1) Is there repeatability during laboratory evolution?
(2) Do contingent factors associated with the foundation process affect the evolutionary dynamics observed?
(3) If so, are these contingencies derived from temporal and/or spatial differences in the genetic composition of the populations? Do genetic sampling effects play a role in the evolutionary differences between populations?

Materials and Methods

FOUNDATION AND MAINTENANCE OF THE LABORATORY POPULATIONS

This study includes data from seven different sets of wild-caught samples of *D. subobscura*. These different sets of populations were founded in the calendar years 1998 (NW populations; see Matos et al. 2002), 2001 (AR and TW populations; see Simões et al. 2007), and 2005 (FWA, FWB, NARA, and NARB populations, the new data presented here)—see Table 1 for an overview of the experimental populations analyzed. Both NW and TW populations were collected from a pinewood near Sintra, Portugal whereas AR populations were collected from Arrábida, Portugal—see Simões et al. 2007.

Table 1. General description of the experimental populations used in this study.

Foundation	Location of sampling	Year of sampling	Sub-samples	No. of founder females	Average population sizes
NW	Sintra	1998	–	300	895.3
TW	Sintra	2001	–	110	870.5
AR	Arrábida	2001	–	59	795.6
NARA	Arrábida	2005	a	55	952.5
NARB	Arrábida	2005	b	68	928.3
FWA	Sintra	2005	a	60	926.7
FWB	Sintra	2005	b	75	856.7

The additional foundations reported here were performed in April 2005 and consisted of two independent collections from each of the two previously sampled natural sites: Sintra, Portugal—"FWA" and "FWB" populations; Arrábida, Portugal—"NARA" and NARB."

All populations were maintained under the same conditions: discrete generations of 28 days, reproduction close to peak fecundity, controlled temperature of 18°C, with a 12-h L: 12-h D photoperiod. Flies were kept in vials, with controlled adult densities of around 50 individuals per vial and larval densities of around 80 per vial. At each generation, emergences from the several vials within each replicate population were randomized using CO_2 anesthesia. Adult population sizes ranged, in general, between 600 and 1200 individuals.

Two generations after foundation, each population was split into three replicate populations (e.g., $FWA_{1–3}$ designating the three populations of the regime "FWA"), except the NW foundation, split into five replicates. A set of longer established laboratory populations "NB" was used as a control for all the experimental populations referred above. NB populations were at their 90^{th} laboratory generation when NW populations were founded (Matos et al. 2002). At the time the 2001 AR and TW populations were introduced into the laboratory, the NB populations were at their 136^{th} generation (Simões et al. 2007). As for the 2005 collections, the NB populations were at their 181^{st} generation at the time of their initial foundation. The early adaptation data analyzed in this study corresponds to the first 15 generations of NW culture and the first 20 generations of AR and TW laboratory populations. All new populations cultured from the 2005 foundations were cultured for 21 generations, during which their adaptation to laboratory conditions was also studied by means of phenotypic assays.

LIFE-HISTORY TRAIT ASSAYS

In each generation assayed an additional egg collection was made for the phenotypic assay, using the same basic procedure as described above for the standard maintenance of populations. Mated pairs of flies used individuals emerging in the same day, and were formed less than 6 h after eclosion started (stimulated by the light phase). These pairs were transferred daily to vials containing fresh medium, and the total number of eggs laid per female was counted daily for the first 12 days. After the fecundity assay was performed, each pair of flies was transferred to a vial containing plain agar medium where the number of hours of starvation resistance was measured. Four characters were analyzed: age of first reproduction (number of days between emergence and the day of first egg laying), early fecundity (total number of eggs laid during the first week), peak fecundity (total number of eggs laid between days 8 and 12) and female starvation resistance (number of hours until death, registered every 6 h).

Assays were performed at generations 4, 8, 13, and 15 of NW laboratory culture and at generations 3, 4, 6, 7, 12, 14, 18, and 20 of AR and TW populations. Phenotypic assays on the 2005 populations were carried out during generations 3, 6, 10, 12, 15, and 21. Sample sizes ranged between 14 and 21 pairs per replicate population. All assays involved synchronous analyses with NB populations.

STATISTICAL METHODS

The sampling methods used in this study resulted in a natural hierarchy of effects. At the very top of this hierarchy are samples that were collected in different years, 1998, 2001, and 2005 (Fig. 1). Within years are samples of flies from two different geographic locations, Sintra and Arrábida (except for 1998, with a foundation from Sintra only). In 2005 two collections were done in each location (designated as subsamples in each location: FWA and FWB from Sintra; and NARA and NARB from Arrábida). Every sample from a given location or a given subsample was replicated in the laboratory giving rise to three independent evolving laboratory populations, except in the case of the 1998 foundation that was fivefold replicated. The common replicates from a single subsample or a single sample from one location are called a foundation.

The data used involved phenotypic assays for each trait (using information at the individual level) concerning the initial adaptive process of the following laboratory foundations: NW (foundation in 1998; generations 4–15); AR and TW (foundations in 2001; generations 3–20); FWA, FWB, NARA and NARB (foundations in 2005; generations 3–21)—see also Figure 1. Differences relative to the average of the same-numbered NB replicate population (assayed synchronously with experimental populations) were used as input data for all the analyses.

Differences to a reference population (NB populations) were used to remove any inadvertent evolutionary change arising independently from the process of laboratory adaptation and to reduce the confounding effects of environmental heterogeneity between temporally spaced phenotypic assays that might obscure the actual

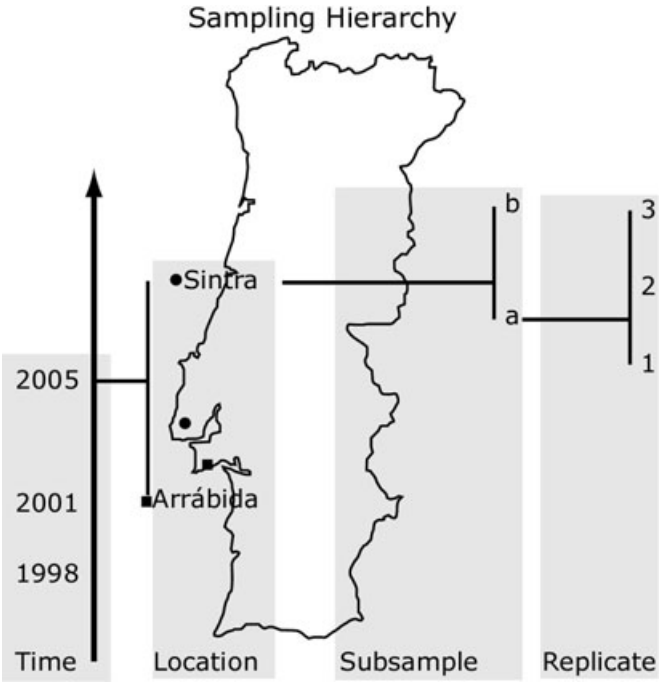

Figure 1. The sampling hierarchy used in this study. The circle and square symbols indicate the location of the Sintra and Arrábida locales in Portugal. Levels of hierarchy include year of foundation (1998, 2001, and 2005), location (Sintra and Arrábida) and subsamples (two collections from each location, in 2005). All foundations were threefold replicated, except the 1998 foundation (NW) which had five replicates. In both 2001 (AR from Sintra and TW from Arrábida) and 2005 (FWA, FWB from Sintra and NARA, NARB from Arrábida), foundations from the different geographical sources were carried out synchronously, using the same pinewood at each location.

evolutionary trends (vid. Matos et al. 2002; Simões et al. 2007). In fact, we have detected signs of environmental trends during the time period of each study, reinforcing the importance of using reference populations already stabilized in terms of laboratory adaptation. In any case, the average NB values were relatively stable among studies being, respectively for 1998, 2001, and 2005 (average ± SE for each study): (1) Age of first reproduction: 3.0477 ± 0.0464; 2.9881 ± 0.1454; 3.0357 ± 0.1166; (2) Early Fecundity: 122.6216 ± 4.0033; 121.0092 ± 5.0288; 128.5320 ± 9.4313; (3) Peak Fecundity: 172.5627 ± 4.2603; 142.0682 ± 2.0639; 165.0211 ± 6.8953; (4) Female starvation resistance: 43.8457 ± 1.3466; 42.5735 ± 1.1314; 39.3593 ± 1.3300.

Bootstrap Techniques: To study the effects of sampling hierarchy on the slope of the selection response we used bootstrap resampling techniques. Let the phenotype for year-i, location-j, subsample-k, replicate-l, and individual-v be y_{ijklv}. Because the goal is to predict these phenotypes from the generations of selection we let x_{ijklv} be the generation in which phenotype y_{ijklv} was measured. In a particular replicate population standard linear regression techniques can be used to estimate the regression coefficients ($\hat{\boldsymbol{\beta}}_{ijkl}$). From these regression coefficients residual error terms are determined from the relationship, $\hat{\varepsilon}_{ijklv} = y_{ijklv} - \hat{\beta}_{1ijkl} - x_{ijklv}\hat{\beta}_{2ijkl}$, $v = 1, 2, \cdots, n_{ijkl}$. These error terms were used to create the bootstrap samples. For each replicate population an empirical distribution of error terms was estimated, $\hat{\boldsymbol{\varepsilon}}_{ijkl} = \left(\hat{\varepsilon}_{ijkl1}, \cdots, \hat{\varepsilon}_{ijkln_{ijkl}}\right)$.

A bootstrap sample, $\boldsymbol{\varepsilon}^* = \left(\varepsilon_1^*, \cdots, \varepsilon_{n_{ijkl}}^*\right)$, was created by taking n_{ijkl} samples with replacement from $\hat{\boldsymbol{\varepsilon}}_{ijkl}$. From the bootstrap sample of error terms a vector of phenotypes is generated by, $y_v^* = \hat{\beta}_{1ijkl} + x_{ijklv}\hat{\beta}_{2ijkl} + \varepsilon_v^*$. From the vector of y^*'s a new bootstrap intercept ($\hat{\beta}_1^*$) and slope ($\hat{\beta}_2^*$) is estimated by standard linear regression techniques. This process was repeated to generate B bootstrap intercept ($\hat{\beta}_{1ijkl}^{*b}$) and slope estimates, $\hat{\beta}_{2ijkl}^{*b}$, $b = 1,\cdots,B$.

Bootstrapping residuals requires that the variance of the error terms be the same over all generations (Efron and Tibshirani 1993). For the composite phenotype this assumption appears to be well met. As an example the residual errors for the first replicate Sintra population in 1998 is shown in Figure 2.

Variation across levels of the sampling hierarchy: To determine the importance of the sampling hierarchy to variation in the adaptive response we created different samples from the basic bootstrap intercepts and slopes. Because the techniques are the same with intercepts and slopes we illustrate the methods with slopes. A random sample of 1000 bootstrap slopes, $\hat{\beta}_{2ijkl}^{*b}$, were chosen from each of the 23 different replicate populations with each replicate ($l = 1,2, \cdots, 23$), and slope ($b = 1, \cdots, 1000$) having an equal chance of being chosen. With this particular sample we then computed the mean of the slopes for each collection from the same replicate population. We then subtracted the mean from the slopes within each replicate to obtain a centered slope value. This allows us to compare all slopes on the same scale.

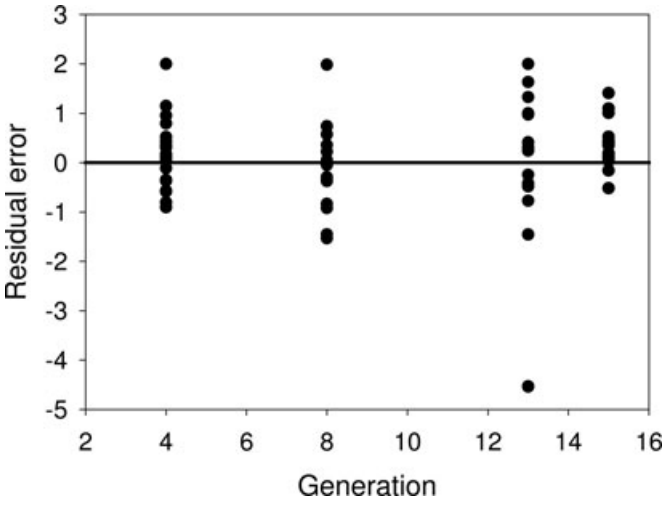

Figure 2. Residual errors from the bootstrap resampling analysis for the NW_1 replicate population (Sintra, 1998 foundation).

To compute the variation over levels of foundations a sample of 1000 slopes (or intercepts) was taken from all 23 replicates. The mean of those slopes that were from a common foundation were then computed and subtracted from the appropriate slopes.

In a similar fashion the subsample variation was computed by taking a sample of 1000 slopes from all the replicate populations in 2005. Then the mean of those slopes that originated from the same location was computed and then subtracted from each slope. Similar methods were followed to compute the centered slopes and intercepts for locations and time.

To compute the centered slopes and intercepts for different locations we did not use Sintra 1998 in the analysis. Consequently, there were $23 - 5 = 18$ different replicate populations representing Sintra and Arrábida in 2001 and 2005. For each of these 18 populations there were 1000 bootstrap slopes. To estimate the variances we sampled at random 1000 slopes from these 18 populations, thus each replicate population had an equal chance of being included in this sample. Not all of the bootstrap replicates were used because other comparisons had more (e.g., foundations and replicate variation) or less (subsamples) than 18 populations, so we wanted each level to be based on the same number of observations.

These 1000 randomly chosen slopes were divided into two groups based on the population they were sampled from: (1) Sintra 2001 and Arrábida 2001 and (2) Sintra 2005 and Arrábida 2005. The mean of the first group was subtracted from all the observations in the first group and likewise we subtracted the mean of the second group from the observations in the second group. These centered slopes then reflect the variation among locations with a common year of sampling. The variance in the figures is the variance of these 1000 centered values.

For variation among years the same 18 populations were used but the two groups were (1) Sintra 2001 and 2005, and (2) Arrábida 2001 and 2005. So now the variation around the mean reflected variation among different years.

Effects of hierarchy on rates of evolution: The bootstrap samples can also be used to test for significant differences in the slope of the selection response and in the *y*-intercept. Our procedure follows the methods used in the previous section. For instance to test for differences between subsamples we sampled one bootstrap slope or intercept from each replicate within each subsample from a common location. The mean of slopes from each subsample was then computed and their average was taken and saved. This process was repeated 1000 times. A statistical test consists of computing the fraction of these 1000 differences that are greater than zero. Two times this fraction or 1 minus this fraction (whichever is less) constitute the *P*-value for significant differences.

Composite phenotype: The effects of laboratory adaptation on the phenotypes measured in this study are almost certainly not independent. It is likely that there are genetic correlations between traits like early fecundity and peak fecundity. Alternatively each of these traits may contribute to fitness in these laboratory environments to different degrees and thus all would be affected by natural selection. It is worthwhile to consider a composite of all the life-history traits and follow the change in this composite trait. This composite trait would also reduce the number of variables tested and simplify the analysis of the basic evolutionary questions.

One way to deduce an appropriate composite trait is to empirically determine what trait is changed by selection to the greatest extent. We have done this by using all four female phenotypes (standardized as difference relative to the NB population) from each of the 23 replicate samples. In each replicate we used phenotype measurements at the earliest generation, called the start of selection, and phenotype measurements at the last generation, called the end of selection. Using linear discriminant analysis, the linear combination of all four traits that produced the largest separation between the start and the end of selection was found (Morrison 1976).

A composite phenotype is defined from the coefficients in Table 2 as

$$y = \sum_{i=1}^{4} p_i l_i, \tag{1}$$

where p_i is the ith phenotype and l_i is the ith linear discriminant coefficient. There are some general trends that fall out of the sign and magnitude of the discriminant functions. Age of first reproduction and female starvation resistance are roughly on the same scale and they have the same negative sign so the composite function will increase as these two phenotypes decrease. However, the magnitude of the coefficient for age of first reproduction is about 1000 times larger and thus the value of the composite phenotype depends to a much greater degree on this phenotype than on female starvation resistance. The two fecundity phenotypes are roughly on the same scale but of opposite sign. Thus, the composite phenotype increases with increasing early fecundity and decreasing peak fecundity. However, because the coefficient for early fecundity is roughly three times the peak fecundity coefficient it contributes much more to the final value of the composite

Table 2. The linear discriminant coefficients for the four standardized female life-history traits obtained from the discriminant analysis.

Phenotype	Linear discriminant coefficient (l_i)
Age of first reproduction	−0.409
Early fecundity	0.0214
Peak fecundity	−0.00883
Female starvation resistance	−0.000510

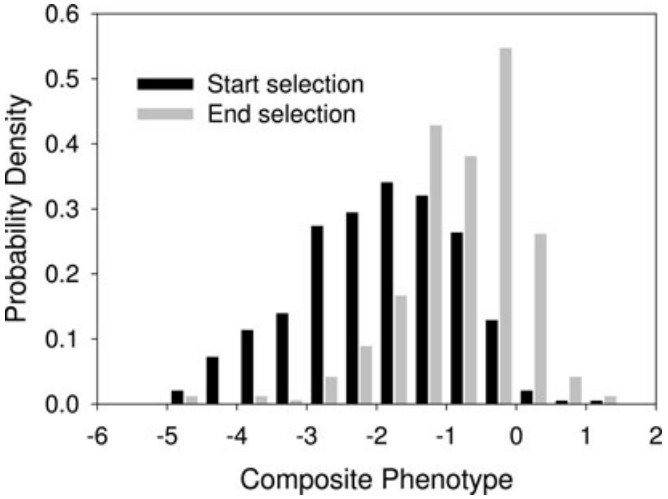

Figure 3. Distribution of the composite female phenotype. This distribution follows from applying equation (1) to data collected at the start and at the end of laboratory selection over all replicate populations.

phenotype. Overall, age of first reproduction and early fecundity contribute the most to the numerical value of the discriminant function. The distribution of the composite phenotype shows substantial differentiation between the start and end of laboratory selection (Fig. 3).

Results

The response to laboratory adaptation of the individual life-history traits is variable between traits and levels of sampling. Early fecundity and peak fecundity generally increase over time whereas age of first reproduction typically declines over time (see Fig. 4). These changes are in accord with our general expectations for improvements in fitness. On the other hand female starvation resistance may increase, decrease, or not change at all, depending on the source population and year (see Fig. 4D).

The progression of the composite phenotype shows consistent increases over all locations and years (Fig. 5) suggesting that this measure is responding consistently to laboratory adaptation (see also Fig. 3). We next examine if the response to selection varies across subsamples (within each location), locations, or years.

By looking at variation within each sampling level we can get an impression of the magnitude of variation that originated with each hierarchical level of sampling. One of the most variably evolving single phenotypes is female starvation resistance (Fig. 6). The slope for this character shows a large change from the individual replicates to the differences among replicates within foundations and then smaller increases at the higher levels of subsample and time (Fig. 6A). The changes are still more dramatic for the intercept of female starvation resistance (Fig. 6B). There is a considerable increase in the variation of this parameter proceeding from the individual replicate level to differences among replicates and then a further, although smaller, increase from that level to the higher levels of the analysis, with the exception of effects of location.

The composite phenotype shows much less variation due to the different sampling levels (Fig. 7). Variation in slopes increases roughly threefold between the individual replicate and among replicates within the same foundation (Fig. 7A). However, the other sampling levels (subsamples, location and year) have no more variation than is seen at the level of the foundation. This would suggest that most of the variability among these inferred evolutionary trajectories arises between independent replicate populations founded from a common wild-caught sample.

Regression intercepts for the composite phenotype are more affected by levels of sampling hierarchy (Fig. 7B). Variation at the foundation level presents also a roughly threefold increase relative to the variation of individual replicates. However, variation is further increased by both subsamples and spatially distinct samples (Fig. 7B). The trend in Figures 6 and 7 and for the other traits not shown is that higher sampling levels add greater variation to the intercept than they do to the slope of the selection response.

A summary of the tests for significant differences in slopes at the different sampling levels is given in Table 3. The individual phenotypes show significant effects at 1 to 5 of the 7 categories. Among the individual phenotypes there is at least one example of significant effects from each level of hierarchy. These significant effects are more frequently associated with female starvation resistance, which showed significant effects at all levels of hierarchy except location. In contrast, fecundity-related traits as well as the composite phenotype only presented significant differences for 1 or 2 comparisons (see Table 3). We found three significant effects of location out of five tests, all involving fecundity-related traits as well as the composite phenotype. On the other hand, for these traits there was no significant effect of the subsample level, in contrast with the results obtained for female starvation resistance.

The intercept represents the phenotype value at the start of the selection experiments. The evolutionary intercepts for individual phenotypes show significant effects at 3 to 4 of the 7 categories, with numerous significant effects of sampling hierarchy (Table 4). This indicates that temporal and spatial sampling produces populations whose initial phenotypes are significantly different.

Discussion

Our results clearly support the importance of contingency for the starting point of adaptive evolution. Contingent effects can apparently arise from both temporal and/or spatial effects during the process of sampling from the wild, changing among populations

even within the same geographical region. This variation occurs predominantly in the initial performance of populations and, to a lesser extent, in their subsequent evolutionary trajectory. Female starvation resistance shows clearly a more contingent pattern of evolution both in the differences of initial state and early adaptive rates between foundations, even changing sign. The more consistent general response to laboratory adaptation of fecundity characters is expected because of their relevance to fitness. In the same vein, the composite phenotype shows generally parallel evolution across temporal foundations.

As a whole, these results clearly highlight the prominent role of fecundity characters during laboratory adaptation that has also been observed in other recent studies (see also Hercus and Hoffmann 1999; Sgrò and Partridge 2000; Matos et al. 2002; Simões et al. 2007). On the other hand, the importance of stress resistance characters is apparently equivocal. In fact, it is not

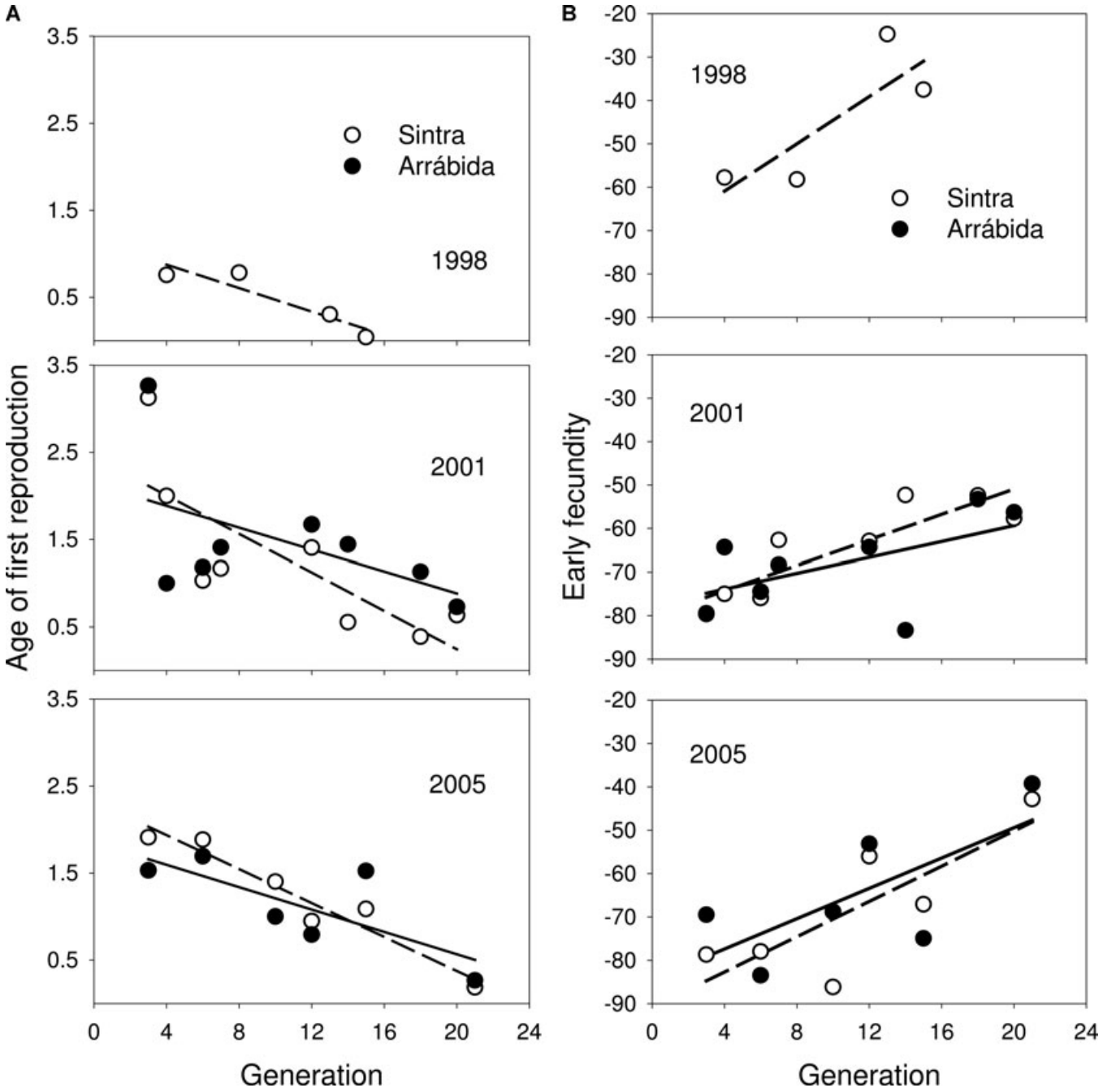

Figure 4. Evolutionary response of individual life-history traits to laboratory selection. (A) Age of first reproduction; (B) Early fecundity; (C) Peak fecundity; (D) Female starvation resistance. The symbols show the average values in each location and year. For 2005 the averages for each location included both subsamples. The lines show the linear fit to the phenotypic change in populations derived from each location.

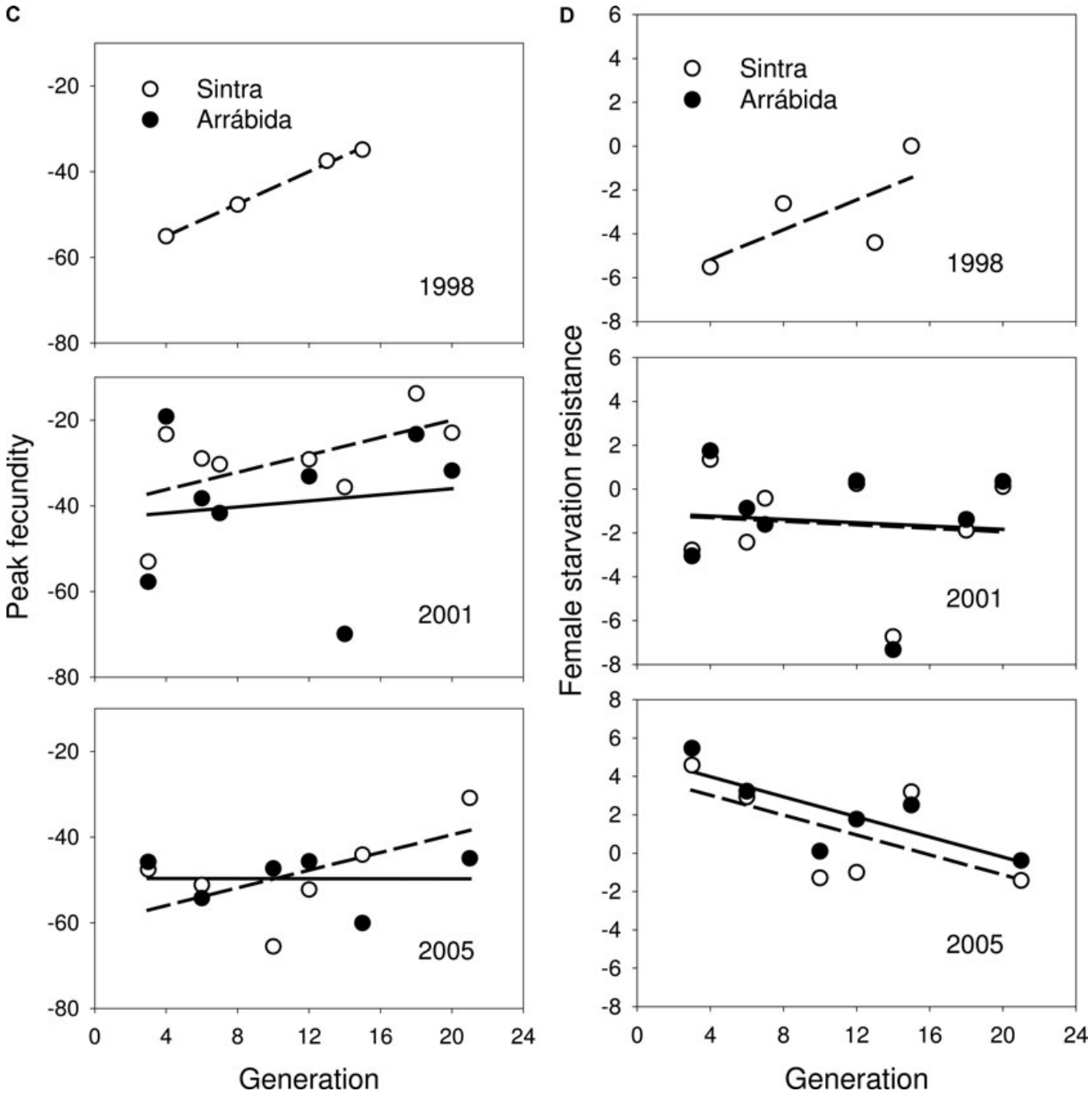

Figure 4. Continued.

possible to generalize or predict the evolutionary patterns during laboratory adaptation for this particular trait, given its varying evolutionary responses to the year of sampling from the wild (see Tables 3 and 4 and Fig. 4).

The differences in the relative impact of these evolutionary contingencies among traits appear thus to depend on the relevance of the specific trait for the overall performance of a population in this new environment. In particular, the relationship between starvation resistance and fitness is probably complex or weak relative to that of fecundity traits (Hoffmann et al. 2001; Matos et al. 2002; Simões et al. 2007). This would explain our finding that the evolutionary response of resistance traits is more contingent on the initial composition of the laboratory populations, leading to significant variation in evolutionary patterns across foundations. This might also in turn explain disparities between laboratories with respect to changes in starvation resistance during captivity (e.g., Hoffmann et al. 2001, Griffiths et al. 2005).

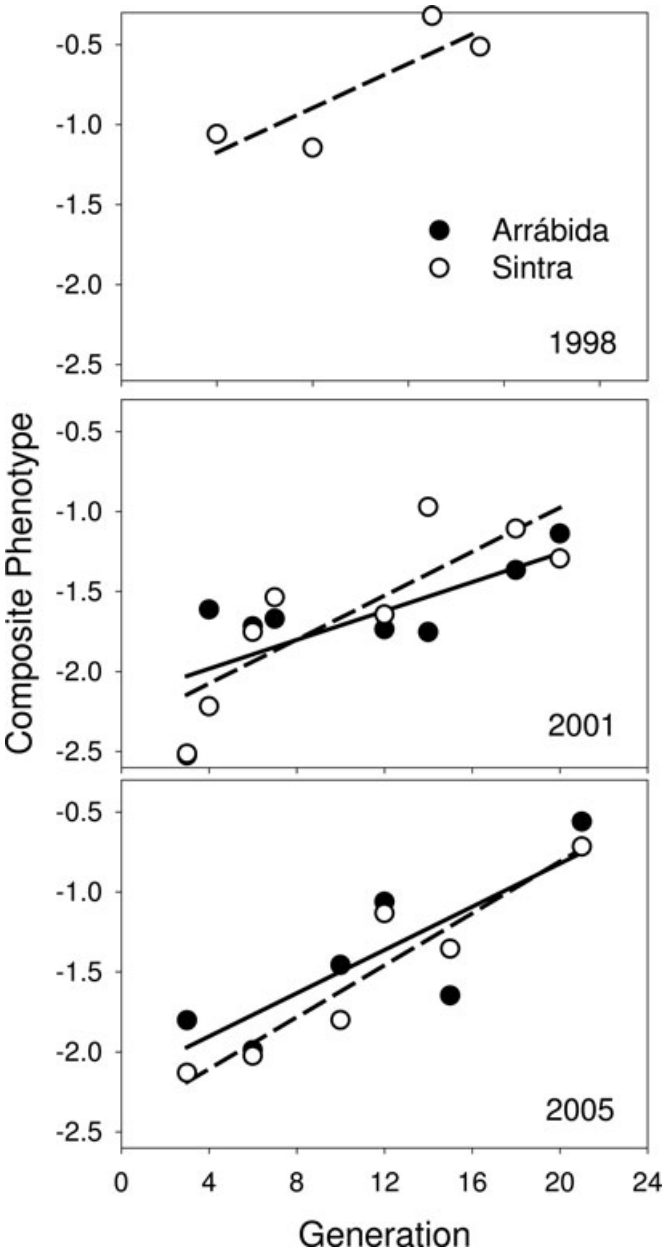

Figure 5. The response of the composite phenotype to laboratory selection. The symbols show the mean composite phenotype calculated from equation (1) in each location and year. For 2005, the averages for each location included both subsamples. The lines show the linear fit to the phenotypic change in populations derived from each location.

CONSISTENCY OF ADAPTATION DURING LABORATORY EVOLUTION

This study reveals substantial variation in the adaptive response to a new environment, both at the start of adaptation as well as, for weakly selected characters, in initial evolutionary rates among foundations. Disparities in rates of adaptation have been inferred in all our previous studies, which have separately treated temporal (e.g., Matos et al. 2002) and spatial (e.g., Simões et al. 2007) effects. The present study involves both further replication of that earlier work and an overall analysis of the entire series of experiments, including seven independent foundations stretching over almost 10 years. As such, our analysis provides the most complete analysis of the adaptive evolution of a species in response to the same selective regime, when multiple foundations from wild populations at different locations and different times are employed.

The results of the present study also provide a useful contrast with those of other authors who started their laboratory evolution experiments from different source populations. For example, Cohan and Hoffmann (1986) found that *Drosophila melanogaster* populations derived from collections obtained from wild populations at very different latitudes along the west side North American coast showed different correlated responses to selection for knockdown resistance to ethanol. Teotónio and his collaborators (Teotónio and Rose 2000; Teotónio et al. 2002) performed a reverse evolution experiment, in which genetically differentiated populations were returned to their common ancestral environment and then allowed to evolve in parallel for 50 generations. They found significant heterogeneity among the evolutionary trajectories of these populations in the same environment. Teotónio et al. concluded that past selective history along with a variable relationship between life-history characters and fitness were responsible for the heterogeneity that they observed. Both of their studies indicated effects of genetic background on evolutionary response.

It is possible that the differences that we observe in the initial state and, to some extent, in the subsequent process of adaptation between our laboratory populations are reduced in the long term. In fact, our previous analyses of the 2001 foundations from Arrábida and Sintra showed significantly different evolutionary rates of adaptation in the short term no longer obtained with more generations (see Simões et al. 2007). In this sense, our results are more in keeping with other experimental evolution studies in *Drosophila*, which have shown that the effects of evolutionary history are transient (e.g., Joshi et al. 2003).

The long-standing experimental evolution work of Lenski and colleagues using *E. coli* has also addressed the effects of historical contingencies in evolution (Travisano et al. 1995; Elena and Lenski 2003). In one particular experiment, *E. coli* lines that had previously evolved in glucose for 2000 generations were placed in a maltose environment for anot her 1000 generations. The evolutionary response in this new environment was measured in 36 *E. coli* populations—as the design employed three replicate populations from each of the 12 populations that previously evolved in glucose (see Travisano et al. 1995). The results showed that the replicate derivatives of these populations evolving in the new maltose environment achieved similar fitness levels despite prior history and/or subsequent chance events. On the other hand,

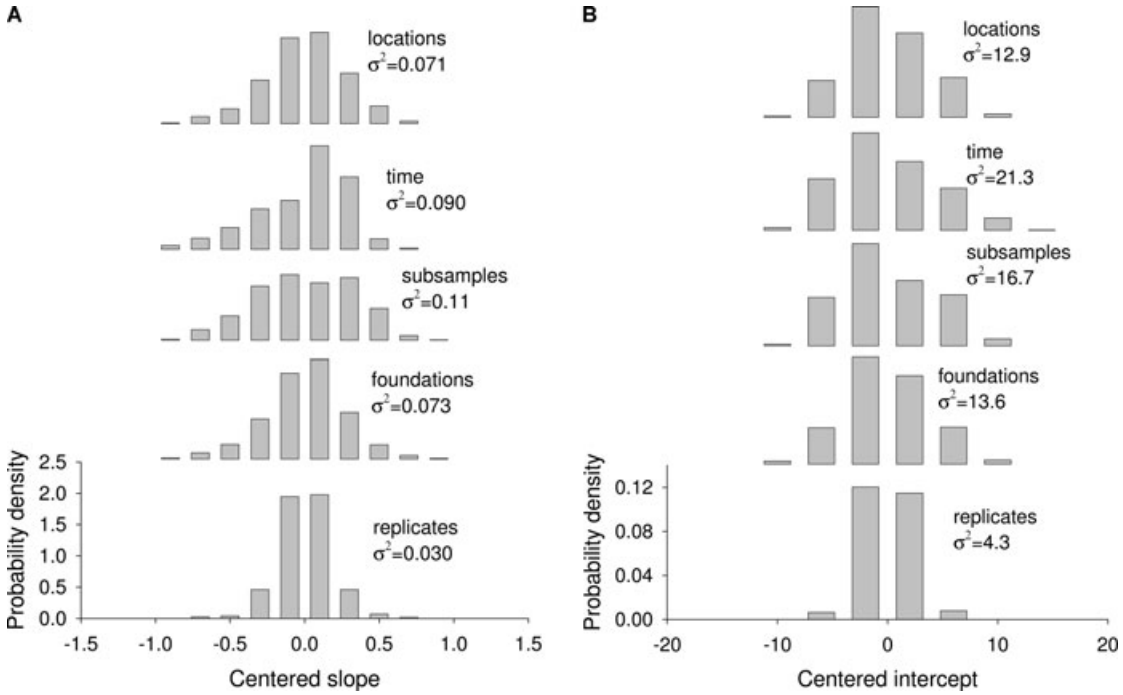

Figure 6. The distribution of the parameters slope and intercept over levels of hierarchy for female starvation resistance: (A) distribution of slopes; (B) distribution of intercepts. The distributions are normalized so their area sums to 1. Slopes within a given level of hierarchy are centered about their mean. The variance of each level is given as σ^2. The "replicates" level corresponds to the variation obtained for each individual replicate whereas the "foundations" level refers to the variation between replicates within the same foundation, "subsample" level refers to the variation between replicate populations from different subsamples (within each location), and the same rationale applies to the two higher hierarchical levels (location and time).

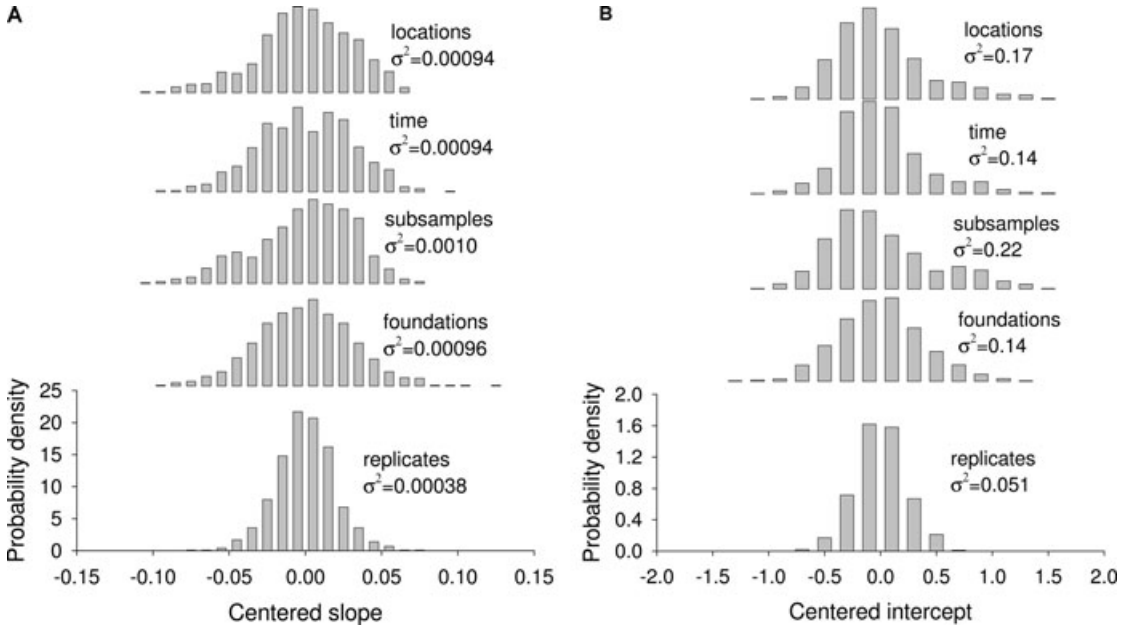

Figure 7. The distribution of the slope and intercept parameters over levels of hierarchy for the composite phenotype: (A) distribution of slopes; (B) distribution of intercepts. The distributions are normalized so that their area sums to 1. Slopes within a given level of hierarchy are centered about their mean. The variance of each level is given as σ^2. The "replicates" level corresponds to the variation obtained for each individual replicate whereas the "foundations" level refers to the variation between replicates within the same foundation, "subsample" level refers to the variation between replicate populations from different subsamples (within each location), and the same rationale applies to the two higher hierarchical levels (location and time).

Table 3. The *P*-values for tests of significant differences between slopes for different levels of sampling using bootstrap analysis. Tests with *P*-values less than 5% are shown in bold face type. The comparisons between locations involved the 2001 and 2005 data.

Level	Test	Trait *P*-values				
		Age	Early fecundity	Peak fecundity	Female starvation resistance	Composite
Subsample	Arrábida	0.31	0.68	0.12	**0.022**	0.30
	Sintra	0.33	0.33	0.082	0.95	0.25
Location	Arrábida-Sintra	**0.004**	0.18	**0.01**	0.74	**0.024**
Year	Arrábida, 2001–2005	0.85	**0.022**	0.16	**0.008**	0.092
	Sintra, 1998–2001	**0.026**	**0.010**	0.28	**0.018**	0.98
	Sintra, 2001–2005	0.27	0.098	0.75	**0.006**	0.31
	Sintra, 1998–2005	0.16	0.11	0.18	**<0.001**	0.47

chance and historical events were found to have a clear impact on traits less correlated with overall fitness (e.g., cell size), leading to differentiation among populations in these particular traits (Travisano et al. 1995). Our data similarly suggests that evolutionary contingencies have more impact on traits less relevant to fitness. Specifically, starvation resistance, a trait expected to have less direct impact on fitness in the laboratory environment relative to fecundity traits, shows an evolutionary pattern that is more contingent on the effects of foundation. Nevertheless, comparisons of studies in sexual populations with those involving asexual populations should be made with caution, given that the former do not allow us to disentangle clearly the effects of prior evolutionary history from other sources of initial genetic differences between founder populations (e.g., sampling effects upon foundation). Studies involving more systematic multiple samplings from the same natural populations—using replicate foundations within each given year of sampling (as done for the 2005 data)—may improve our capacity to distinguish between these several sources of variation.

WHAT CAN WE SAY ABOUT THE WILD POPULATIONS THAT WE SAMPLE AND SUBJECT TO LABORATORY EVOLUTION?

The present study is different from those of Teotónio and colleagues (e.g., Teotónio and Rose 2000; Teotónio et al. 2002) or Lenski and colleagues (Travisano et al. 1995) in that the genetic differentiation that our experimental populations start with comes from nature. This naturally raises the issue of the relationship between laboratory studies of experimental evolution and the properties of populations in the wild.

Some authors have argued that the use of long-established laboratory populations limits inferences about evolutionary processes (Harshman and Hoffmann 2000; Linnen et al. 2001). This view assumes that experimental evolution studies aim at extrapolating specific evolutionary patterns shown by laboratory populations to populations in the wild. However, this is not necessarily the goal of such studies. Experimental evolution often focuses primarily on tests of predictions regarding evolutionary processes in general. In our view, any particular laboratory environment is

Table 4. The *P*-values for tests of significant differences between intercepts for different levels of sampling using bootstrap analysis. Tests with *P*-values less than 5% are shown in bold face type. The comparisons between locations involved the 2001 and 2005 data.

Level	Test	Trait *P*-values				
		Age	Early fecundity	Peak fecundity	Female starvation resistance	Composite
Subsample	Arrábida	0.34	0.53	0.76	0.35	0.29
	Sintra	0.14	**0.026**	**0.002**	0.60	0.066
Location	Arrábida-Sintra	**0.010**	0.11	0.11	0.47	**0.016**
Year	Arrábida, 2001–2005	0.088	0.10	0.44	**<0.001**	0.66
	Sintra, 1998–2001	**<0.001**	0.12	**0.004**	**<0.001**	**<0.001**
	Sintra, 2001–2005	0.20	**0.012**	**0.002**	**<0.001**	**<0.001**
	Sintra, 1998–2005	**<0.001**	**<0.001**	0.69	**<0.001**	**<0.001**

just another environment featuring a particular selection regime to which populations may adapt (Matos et al. 2000b). Moreover, we propose that the laboratory environment can in fact be an ideal setting to address the potential for adaptive responses and to test general predictions concerning evolutionary patterns, such as convergence. The analysis of the laboratory adaptation of recently wild-collected samples allows experimenters to study in detail the evolutionary response of populations with high starting genetic variability as well as to address issues such as the impact of different genetic backgrounds on evolution, or the repeatability of evolutionary patterns across temporally and spatially sampled populations, as we have here.

Our findings suggest that *D. subobscura* populations can vary significantly in their initial performance in a novel laboratory environment even when sampled from a relatively small geographical area over a relatively short period of time. Our long-term research program provides a powerful and novel window into the potential for adaptive evolution of populations in the wild, a window very different from that provided by the collection of data pertaining solely to standing genetic variation in the wild. The latter has been a traditional research topic within population genetics, from the pioneering studies of Dobzhansky and Lewontin (Dobzhansky 1937; Lewontin and Hubby 1966) to the recent attempts to detect selection in natural populations (see Ford 2002, for a review). It is also a different angle on this question from that afforded by studies of the phenotypics of selection in nature (e.g., Lande 1979; Lande and Arnold 1983; Arnold and Wade 1984a,b; Grant and Grant 1995; Reznick et al. 1997).

What we are suggesting is that, although the laboratory evolution of a sample collected from a wild population is necessarily limited to the particular selection regime(s) chosen by an experimenter, it provides one of the most readily interpretable assays of the potential for adaptation of a population. As such, this particular type of assay should perhaps be more common among the experimental designs used in evolutionary research.

ACKNOWLEDGMENTS

This study was partially financed by "Fundação para a Ciência e a Tecnologia" (FCT) project no. POCTI/BSE/33673/2000 and by FCT and POCI 2010 project no. POCI-PPCDT/BIA-BDE/55853/2004 (both with co-participation of FEDER). JS had a BTI grant and P S had a Ph.D. grant (SFRH/BD/10604/2002) from FCT.

LITERATURE CITED

Arnold, S. J., and M. J. Wade. 1984a. On the measurement of natural and sexual selection: theory. Evolution 38:709–719.

———. 1984b. On the measurement of natural and sexual selection: applications. Evolution 38:720–734.

Bryant, E. H., S. A. McCommas, and L. M. Combs. 1986. The effect of an experimental bottleneck upon quantitative genetic variation in the housefly. Genetics 114:1191–1211.

Cheverud, J. M., and E. J. Routman. 1995. Epistasis and its contribution to genetic variance components. Genetics 139:1455–1461.

Cheverud, J. M., T. T. Vaughn, L. S. Pletscher, K. King-Ellison, J. Bailiff, E. Adams, C. Erickson, and A. Bonislawski. 1999. Epistasis and the evolution of additive genetic variance in populations that pass through a bottleneck. Evolution 53:1009–1018.

Chippindale, A. K. 2006. Experimental Evolution. Pp. 482–501 *in* C. Fox and J. Wolf, eds. Evolutionary genetics: concepts and case studies. Oxford Univ. Press, London.

Cohan, F. M. 1984a. Can uniform selection retard random genetic divergence between isolated conspecific populations? Evolution 38:495–504.

———. 1984b. Genetic divergence under uniform selection. I. Similarity among populations of *Drosophila melanogaster* in their responses to artificial selection for modifiers of ci^D. Evolution 38:55–71.

Cohan, F. M., and A. A. Hoffmann. 1986. Genetic divergence under uniform selection. II. Different responses to selection for knockdown resistance to ethanol among *Drosophila melanogaster* populations and their replicate lines. Genetics 114:145–163.

———. 1989. Uniform selection as a diversifying force in evolution: evidence from *Drosophila*. Am. Nat. 134:613–637.

de Brito, R. A., L. S. Pletscher, and J. M. Cheverud. 2005. The evolution of genetic architecture. I. Diversification of genetic backgrounds by genetic drift. Evolution 59:2333–2342.

Dobzhansky, T. 1937. Genetics and the origin of species. Columbia Univ. Press, New York.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman & Hall, New York.

Elena, S. F., and R. E. Lenski. 2003. Evolution experiments with microorganisms: the dynamics and genetic basis of adaptation. Nat. Rev. Genet. 4:457–469.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, Harlow.

Ford, M. J. 2002. Applications of selective neutrality tests to molecular ecology. Mol. Ecol. 11:1245–1262.

Grant, P. R., and B. R. Grant. 1995. Predicting microevolutionary responses to directional selection on heritable variation. Evolution 49:241–251.

Griffiths, J. A., M. Schiffer, and A. A. Hoffmann. 2005. Clinal variation and laboratory adaptation in the rainforest species *Drosophila birchii* for stress resistance, wing size, wing shape and development time. J. Evol. Biol. 18:213–222.

Harshman, L. G., and A. A. Hoffmann. 2000. Laboratory selection experiments using *Drosophila*: what do they really tell us? Trends Ecol. Evol. 15:32–36.

Hercus, M. J., and A. A. Hoffmann. 1999. Desiccation resistance in interspecific *Drosophila* crosses: genetic interactions and trait correlations. Genetics 151:1493–1502.

Hoffmann, A. A., R. Hallas, C. Sinclair, and L. Partridge. 2001. Rapid loss of stress resistance in *Drosophila melanogaster* under adaptation to laboratory culture. Evolution 55:436–438.

James, J. W. 1971. The founder effect and response to artificial selection. Genet. Res. Camb. 16:241–250.

Joshi, A., R. B. Castillo, and L. D. Mueller. 2003. The contribution of ancestry, chance, and past and ongoing selection to adaptive evolution. J. Genet. 82:147–162.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution 33:402–416.

———. 1980. The genetic covariance between characters maintained by pleiotropic mutations. Genetics 94:203–215.

Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution 37:1210–1226.

Lenski, R. E. 2004. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. Plant Breeding Rev. 24:225–265.

Lewontin, R. C., and J. L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. The amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54:595–609.

Linnen, C., M. Tatar, and D. Promislow. 2001. Cultural artifacts: a comparison of senescence in natural, laboratory-adapted and artificially selected lines of *Drosophila melanogaster*. Evol. Ecol. Res. 3:877–888.

Matos, M., M. R. Rose, M. T. Rocha Pité, C. Rego, and T. Avelar. 2000a. Adaptation to the laboratory environment in *Drosophila subobscura*. J. Evol. Biol. 13:9–19.

Matos, M., C. Rego, A. Levy, H. Teotónio, and M. R. Rose. 2000b. An evolutionary no man's land. Trends Ecol. Evol. 15:206.

Matos, M., T. Avelar, and M. R. Rose. 2002. Variation in the rate of convergent evolution: adaptation to a laboratory environment in *Drosophila subobscura*. J. Evol. Biol. 15:673–682.

Morrison, D. F. 1976. Multivariate statistical methods. 2nd ed. McGraw-Hill, New York.

Naciri-Graven, Y., and J. Goudet. 2003. The additive genetic variance after bottlenecks is affected by the number of loci involved in epistatic interactions. Evolution 57:706–716.

Powell, J. R., and R. C. Richmond. 1974. Founder effects and linkage disequilibria in experimental populations of *Drosophila*. Proc. Natl. Acad. Sci. USA 71:1663–1665.

Prasad, N. G., and A. Joshi. 2003. What have two decades of laboratory life-history evolution studies on *Drosophila melanogaster* taught us? J. Genet. 82:45–76.

Reznick, D. N., and C. K. Ghalambor. 2005. Selection in nature: experimental manipulations of natural populations. Integr. Comp. Biol. 45:456–462.

Reznick, D. N., F. H. Shaw, F. H. Rodd, and R. G. Shaw. 1997. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). Science 275:1934–1937.

Roff, D. 2000. The evolution of the G matrix: selection or drift? Heredity 84:135–142.

Rose, M. R., T. J. Nusbaum, and A. K. Chippindale. 1996. Laboratory evolution: the experimental Wonderland and the Cheshire Cat syndrome. Pp. 221–241 *in* M. R. Rose and G. V. Lauder, eds. Adaptation. Academic Press, San Diego, CA.

Rose, M. R., H. B. Passananti, and M. Matos, eds. 2004. Methuselah Flies: a case study in the evolution of aging. World Scientific Publishing, Singapore.

Sgrò , C. M., and L. Partridge. 2000. Evolutionary responses of the life history of wild-caught *Drosophila melanogaster* to two standard methods of laboratory culture. Am. Nat. 156:341–353.

Simões, P., M. R. Rose, A. Duarte, R. Gonçalves, and M. Matos. 2007. Evolutionary domestication in *Drosophila subobscura*. J. Evol. Biol. 20:758–766.

Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: evolution of the G matrix. Trends Ecol. Evol. 17:320–327.

Teotónio, H., and M. R. Rose. 2000. Variation in the reversibility of evolution. Nature 408:463–466.

Teotónio, H., M. Matos, and M. R. Rose. 2002. Reverse evolution of fitness in *Drosophila melanogaster*. J. Evol. Biol. 15:608–617.

Travisano, M., J. A. Mongold, A. F. Bennett, and R. E. Lenski. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science 267:87–90.

Wade, M. J., R. G. Winther, A. F. Agrawal, and C. J. Goodnight. 2001. Alternative definitions of epistasis: dependence and interaction. Trends Ecol. Evol. 16:498–504.

Zhang, X.-S., J. Wang, and W. G. Hill. 2004. Redistribution of gene frequency and changes of genetic variation following a bottleneck in population size. Genetics 167:1475–1492.

Associate Editor: M. Travisano