

# Genetically idiosyncratic responses of *Drosophila melanogaster* populations to selection for improved learning ability

T. J. KAWECKI & F. MERY

Section of Ecology and Evolution, Department of Biology, University of Fribourg, Fribourg, Switzerland

## Keywords:

*Drosophila melanogaster*;  
experimental evolution;  
founder effect;  
learning;  
line-cross analysis;  
memory;  
outbreeding depression;  
parallel evolution;  
repeatability of evolution.

## Abstract

To what extent is adaptive evolution over short timescales repeatable? To address this question, we studied the performance of crosses between replicate *Drosophila melanogaster* lines previously subject to selection for improved learning response in the context of oviposition substrate choice. Of the 10 pairwise  $F_1$  crosses among the five selection lines, four performed in the original learning assay similarly to the parental lines, whereas the remaining six showed learning scores significantly below the average of the parental lines. In particular, four  $F_1$  crosses (three involving the same line) showed no detectable learning, on a par with unselected control lines. This indicates that the response to selection in some lines involved allelic substitutions at different loci. Additional assays of crosses between two selection lines indicated that the loss of performance in hybrids generalized to another type of learning assay, and held for both short- and long-term memory. Joint analysis of first- and second-generation crosses between these two lines supported the hypothesis that the response to selection in these different lines was based on the spread of recessive alleles at different loci. These results show that the evolutionary trajectories of populations of the same origin subject to uniform selection may sometimes diverge over very short evolutionary timescales.

## Introduction

There are often alternative ways in which a population may adapt to a given environment, involving different allele substitutions and different phenotypic changes. Therefore, given sufficient time, two initially identical isolated populations are likely to diverge even if they are subject to the same forces of natural selection (Wright, 1931; Lande, 1983; Cohan, 1984; Johnson *et al.*, 1995). This divergence will initially be due to drift (including founder effects) and mutation. However, if epistasis is pervasive, fitness effects of individual alleles will be contingent on changes already accumulated at other loci, so the divergence will be accelerated by selection – the populations will evolve towards different ‘adaptive peaks’ (Whitlock *et al.*, 1995). In sexual organisms such diver-

gence may lead to reduced performance of hybrids between populations (outbreeding depression; Lynch, 1991; Fenster *et al.*, 1997). This process is thought to be a major mechanisms of allopatric speciation (the Dobzhansky–Mueller model; Orr, 1995; Johnson, 2000; Welch, 2004). It is also relevant for the more general issues of the role of chance in adaptive evolution, of the repeatability and predictability of evolutionary change (e.g. Gould, 1989; Teotonio & Rose, 2001; Wood *et al.*, 2005).

Data directly addressing the effect of chance on adaptive evolution remain scarce; it remains unclear how readily and over what timescale populations subject to the same forces of selection will diverge. In a recent review of evidence from phylogenetic, experimental evolution and QTL studies, Wood *et al.* (2005) concluded that adaptive evolution is surprisingly repeatable, with parallel genetic changes occurring independently in isolated populations. Such parallelism, down to the level of nucleotide substitution, is even often observed in microbial experimental evolution studies, in which

Correspondence: Dr Tadeusz J. Kawecky, Section of Ecology and Evolution, Department of Biology, University of Fribourg, Chemin du Musée 10, CH-1700 Fribourg, Switzerland.  
Tel.: +41 26 300 88 71; fax +41 26 300 96 98;  
e-mail: tadeusz.kawecky@unifr.ch

selection lines are initiated with a single clone and adaptation is thus based entirely on new mutations arising independently in each line (reviewed in Wood *et al.*, 2005). Even more parallelism in experimental evolution is expected in multicellular sexual species, in which adaptation is based on standing genetic variation sampled from a genetically variable base population. Several studies (Graves *et al.*, 1992; Blows, 1993; Travisano *et al.*, 1995; Joshi *et al.*, 2003; Ungerer *et al.*, 2003; other examples reviewed in Teotonio & Rose, 2001) demonstrated that even initially different populations often tend to converge both at the phenotypic and genetic level when subject to the same selection regime. This led Wood *et al.* (2005) to suggest that the path of adaptive evolution may be highly constrained, at least on the micro-evolutionary timescale.

Nonetheless, there are also examples of divergence under uniform selection. For example, within <30 years of independent introductions of *Drosophila subobscura* into North and South America, the new populations evolved latitudinal clines in body and wing size, paralleling those in their ancestral range in Europe. However, the clines in wing size in Europe and South America are based on differences in cell number, whereas the cline in North America is due to cell size (Calboli *et al.*, 2003). Furthermore, the sections of wings responsible for the clines differ among the three continents (Gilchrist *et al.*, 2004), indicating genetically and developmentally idiosyncratic responses to presumably similar natural selection. In that case, subtle differences in the forces of selection cannot be excluded, but divergent responses to the same controlled selection regime have also been reported. Even though replicate lines of bacteria evolving in the same environment typically show parallel improvements in fitness (but see Korona *et al.*, 1994), they often vary in correlated responses, indicating different genetic bases of adaptation (e.g. Korona, 1996; Riley *et al.*, 2001; MacLean & Bell, 2003). Evolution of DDT resistance in two *D. melanogaster* populations of different origins involved changes in different metabolic pathways (Pedra *et al.*, 2005). Populations of a bean weevil of different geographic origins continued to differ in host preference (Kawecki & Mery, 2003) and in a number of life history traits (Bieri & Kawecki, 2003) in spite of 120 generations of adaptation to the same environment. Small initial variation in ethanol resistance among *D. melanogaster* populations of different origins was magnified by uniform selection on this trait (Cohan & Hoffman, 1989). Finally, replicate lines of mice selected on nest building behaviour showed heterosis in  $F_1$  (Bult & Lynch, 1996), suggesting that the response involved at least partially different genes, even though the lines originated from the same base population. Thus, divergence in response to uniform selection may occur over relatively short-time scales. More experimental results are needed in order to understand under what conditions and for what traits such divergence is more likely.

In this paper, we address the repeatability of the response to selection for improved learning ability in five replicate *D. melanogaster* lines originating from the same large base population. The flies had been selected to avoid an oviposition substrate that was earlier associated with a bitter taste, and are characterized by faster learning and longer memory, compared with unselected control lines originating from the same base population (Mery & Kawecki, 2002). Here we study the learning ability and memory of crosses between replicate selected populations. If the response of two populations to selection has the same genetic basis, the phenotype of the  $F_1$  cross between them should not differ from the two selected populations, or should be intermediate if the mean phenotype differs between the selected populations. The latter pattern is expected if the frequencies of the favoured allele(s) differ between populations. If, on the contrary, the response of different populations involves allele substitutions at different loci, the  $F_1$  cross is likely to differ from the average of the two parents unless the effects of the favoured alleles are additive both within and between loci. In particular, loss of the evolved phenotype in an  $F_1$  cross between two selection lines would indicate that the response of the two lines involved substitution of recessive alleles at different loci (complementation), or that the alleles favoured in the two lines show strong antagonistic epistasis. We report such a partial or complete loss of the evolved phenotype in six of 10 pairwise crosses between replicate selection lines. Additional assays for two such 'incompatible' lines show that this loss of learning performance in  $F_1$  extends to another type of learning test (olfactory conditioning) and is manifested in both short- and long-term memory. Analysis of the second-generation crosses between these two lines supports the hypothesis that evolution of improved learning and memory in those lines was based on recessive alleles at different loci.

## Material and methods

### Selection lines and culture conditions

The origin of the *high-learning* selection lines and the selection regime have been described in detail elsewhere (Mery & Kawecki, 2002). Briefly, they were derived from a large base population established with about 2000 wild-caught females and maintained in the laboratory for 6 months before the beginning of the selection. In the course of selection, each selection line was maintained at the size of 150 adults. Every generation flies were given a choice between two oviposition media with different flavours (orange and pineapple). When naive flies were first presented with these media, one of them (pineapple in odd-, orange in even-numbered generations) was supplemented with quinine, which has an aversive taste, but apparently cannot be smelled by flies. After 3 h the flies were offered a new orange and new pineapple

medium, this time neither containing quinine. The next generation was bred from eggs laid 3–6 h later (from generation 48 onwards 0–3 h later) on the medium that had not contained quinine. This regime favoured flies which learned the association between a medium flavour and quinine, and continued to avoid this medium even though quinine was not present any more.

The original base population showed no detectable learning ability under the conditions of selection regime (i.e. no detectable change in oviposition medium preference as a result of experience with a quinine-containing medium), and it remained so for a set of unselected control (*low-learning*) populations derived from it. In contrast, within 15 generations of selection the selected *high-learning* lines evolved a substantially improved ability to avoid the medium that has previously contained quinine. Additional assays showed that the *high-learning* lines learn faster and remember longer than the *low-learning* controls (Mery & Kawecki, 2002) and more specifically that they have a better long-term memory (F. Mery and T.J. Kawecki, unpublished data). However, their evolution of improved learning was associated with a reduction in larval competitive ability, suggesting a cost of learning ability (Mery & Kawecki, 2003).

For some assays, we also used as a reference a population created by crossing several of the unselected *low-learning* lines. This population was established around generation 100 of experimental evolution and maintained at the size of >1000 individuals for over 30 generations before being used in the assays reported here.

Both in the course of selection and when bred for the assays, the flies were reared on a standard cornmeal medium at a density of 250 eggs per vial containing 25 g of medium, at 25 °C and complete darkness.

### Assay 1: performance of $F_1$ crosses in the oviposition learning test

For this assay, we used *high-learning* lines 1, 3, 5, 6 and 8 (line 4 had been lost, and lines 2 and 7 were left out to limit the size of the experiment); it took place after 34 generations of selection. For each of the five lines and for each of the 20 pairwise crosses between the lines (counting reciprocal crosses as two) we set up five matings; these matings are the main units of replication. For each mating about 100 freshly emerged males and 100 virgin females of appropriate lines were allowed to mass-mate and oviposit.

The resulting progeny were assayed for their learning ability in an oviposition test similar to that used in the course of selection. For each test a sample of 100 adults (males + females aged 3–5 days from emergence) were first conditioned for 3 h: they were presented with one Petri dish of each orange and pineapple medium, one of them supplemented with quinine. In the subsequent 3 h (test period), the flies could oviposit on fresh orange and

pineapple media, neither containing quinine. The eggs laid on the two media in the test period were subsequently counted. Two paired fly samples from each replicate mating were assayed in parallel; one was conditioned to avoid pineapple (i.e. the pineapple medium contained quinine during conditioning), the other to avoid orange. The difference between these two samples in the proportion of eggs laid on the orange medium in the test period estimates the effect of experience (conditioning) on oviposition medium choice. We refer to this difference as the *learning score* and use it as a dependent variable in subsequent analysis. A maximum possible learning score is one; zero indicates no effect of conditioning on preference (i.e. no learning). A third sample of 100 flies from each mating was assayed for their choice between the two media in the absence of conditioning ('innate' preference).

### Assay 2: outbreeding depression for short- and long-term memory

To get a better insight into the genetic architecture underlying the poor learning performance of  $F_1$  crosses between some pairs of *high-learning* lines (see Results), we assayed the memory of first- and second-generation crosses between one such pair of lines, *high-learning* lines 1 and 8. The two parental lines ( $P_1$  and  $P_8$ ), reciprocal  $F_1$  and  $F_2$  crosses, and the two backcrosses ( $B_1$  and  $B_8$ ) were set up by mass-mating at least 150 virgin females with at least 150 males of the appropriate line;  $F_1$  individuals were used as female parents for the backcrosses. This was done after 119 generations of selection, taking advantage of a new olfactory learning protocol which we had developed in the meantime.

The female progeny (aged 3–6 days) were assayed for short- and long-term memory following a classical (Pavlovian) olfactory conditioning protocol, in which flies could associate an odorant (3-octanol or 4-methylcyclohexanol) with aversive mechanical shock (Mery & Kawecki, 2005). In each conditioning cycle a group of 50 female flies were first presented for 30 s with odour A and simultaneously subject to vibrations delivered by a test tube shaker (2000 rpm vibration pulses of 1 s at 5 s intervals). This was followed by a 60 s rest period during which the flies received humid air (no odours and no shock). Then odour B was delivered for 30 s without shock, followed by another 60 s rest period.

To assay short-term memory, the flies were subject to two consecutive conditioning cycles and tested 20 min later. For the long-term memory the flies were subject to five conditioning cycles at 20 min intervals and tested 24 h after the end of conditioning (for the logic underlying this design see Tully *et al.*, 1994; Mery & Kawecki, 2005). The test involved a choice between the two odours. The flies were transferred to a central point of a T-maze where two air currents carrying the two odorants converged, and given 60 s to choose an arm of the maze.

The proportion of flies choosing each odour was then calculated (the flies that remained in the central chamber of the maze were excluded). A unit of replication consisted of two fly groups from the same line assayed simultaneously, one conditioned to avoid methylcyclohexanol, the other octanol. A memory score was calculated as the difference between these two groups in the proportion of flies choosing octanol. Six replicate learning scores were obtained for each line and cross.

### Assay 3: olfaction

Differences in the learning score could potentially be caused by differences in olfactory ability or motivation to move away from a repulsive odour in the T-maze. In order to exclude this confounding factor, we assayed the behaviour of unconditioned (naïve) female flies in the T-maze when given a choice between one of the odours used in the T-maze learning assay (octanol or methylcyclohexanol) and humid air. Both of these odours are moderately repulsive at the concentrations used in the experiment. Differences in olfaction or locomotor response to a repulsive odour should then be reflected in the proportion of flies moving towards air (away from octanol or methylcyclohexanol). This assay was performed on the *high-learning* lines 1 and 8, the  $F_1$  cross between them and the mixed *low-learning* population. This assay was carried out at generation 121.

### Analysis of deviations from additivity

We used the estimates of mean and standard error of the learning score for the crosses and original selection lines to test for departures from additive gene action on this trait. In assay 1, we also analysed the total number of eggs laid in the test period on both media (realized fecundity), and the proportion of eggs laid on the orange medium in the absence of conditioning (innate preference). The analysis was based on fitting composite genetic parameters with weighted linear regression and testing their significance with likelihood ratio tests (Lynch & Walsh, 1998, pp. 213–221), and was carried out with Mathematica 4.1 (Wolfram, 1999). We did not detect any effect of the direction of the cross for any pair of reciprocal  $F_1$  and  $F_2$  crosses ( $t$ -test, all  $P > 0.5$ ), so we pooled the reciprocal crosses for the analysis. We only report analysis done on learning scores based on untransformed proportions; angular transformation had negligible influence on the results (linear correlation between untransformed and angular learning scores was  $>0.99$ ).

#### Assay 1.

Fitting and testing a separate additive-dominance model for each of the 10 crosses would involve multiple tests using partially overlapping data. With up to three parameters for each cross (intercept, composite additive

and composite dominance effects), the total number of estimated parameters (30) would exceed the number of lines (five parental + 10 crosses), leading to a high degree of redundancy. To avoid this problem, we analysed data from all parental lines and  $F_1$  crosses jointly. We first fitted a saturated additive-dominance model with 15 parameters. The first five parameters ( $m_1$ ,  $m_3$ ,  $m_5$ ,  $m_6$  and  $m_8$ ) estimate the means of the five parental lines; the remaining 10 parameters ( $[d_{1 \times 3}]$  to  $[d_{6 \times 8}]$ ) estimate departures of the 10  $F_1$  crosses from the values expected under additivity. This parameterization differs from the usual parameterization of analysis of crosses between two lines, which includes an overall intercept parameter  $m$ , and a composite additive parameter  $[a]$  estimating half of the difference between the parental lines (Mather & Jinks, 1982; Lynch & Walsh, 1998). Because we simultaneously analyse more than two parental lines, this type of parameterization would be impractical in our case. Nonetheless, the interpretation of the composite dominance parameters  $[d_{1 \times 3}]$  to  $[d_{6 \times 8}]$  remains the same. To test the significance of the parameters, for each parameter we fitted a reduced model, created from the saturated model by dropping the focal parameter (i.e. setting it to zero). Under the null hypothesis that the focal parameter indeed equals zero, the weighted residual sum of squares ( $RSS_w$ ) of the corresponding reduced model is distributed as chi-square with one degree of freedom (Lynch & Walsh, 1998).

The tests of significance of the 10 dominance parameters are not independent; they test a set of overlapping biological hypotheses. For example, if the performance of one set of lines is based on different genes than that of another set of lines, one expects significant dominance components between lines belonging to different sets, but no dominance components in crosses between lines within each set. It is not clear how one might correct the significance values corresponding to individual parameters to account for a given experimentwise type I error in this case. To circumvent this problem and strengthen our conclusions, we thus also used the Akaike Information Criterion (AIC) to find the most parsimonious model within the above parameterization. The most parsimonious model is thought to represent the optimal compromise between the number of parameters in the model and the amount of variation explained by it (Burnham & Anderson, 1998; for an application of AIC to line cross analysis see Bieri & Kawecki, 2003).

#### Assay 2.

The first- and second-generation crosses between *high-learning* lines 1 and 8 were analysed by fitting composite additive, dominance and epistasis parameters. We followed the classic parameterization which takes the mean of the parental lines as the point of reference (Mather & Jinks, 1982, see also Bieri & Kawecki, 2003); note that Lynch & Walsh (1998, Table 9.1) follow a different

parameterization, under which the interpretation of the epistatic parameters is somewhat different.

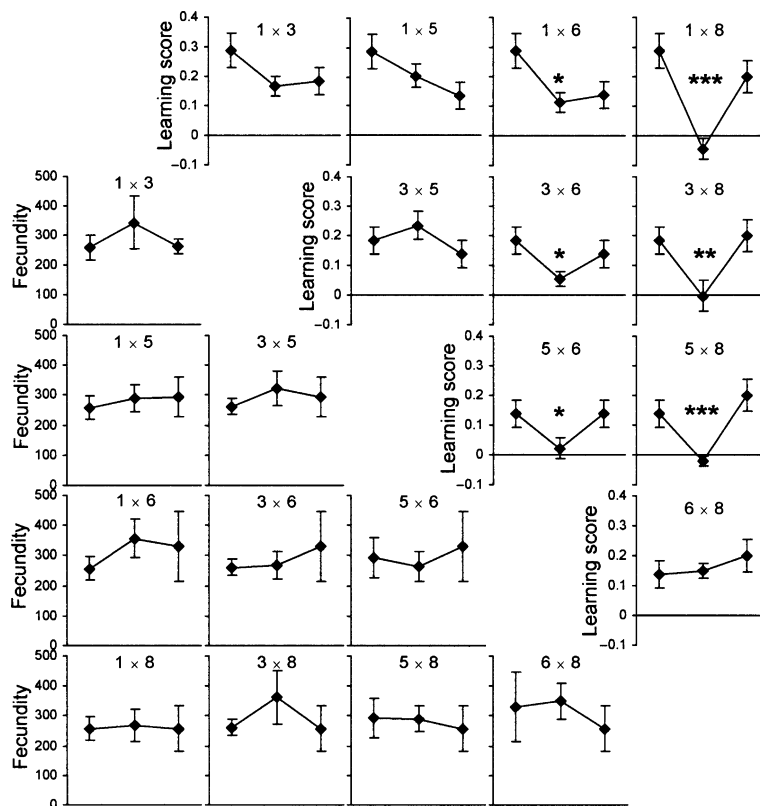
## Results

Based on the analysis of the performance of the  $F_1$  crosses in the oviposition learning test one can divide the five *high-learning* lines in two groups, the first consisting of lines 1, 3 and 5, and the other of lines 6 and 8 (Fig. 1, panels above diagonal). The four  $F_1$  crosses between lines belonging to the same group did not show deviations from additivity (nonsignificant dominance parameters, smallest  $P = 0.17$ ). In contrast, all six  $F_1$  crosses between lines belonging to different groups showed poorer performance than expected under the additive model, as indicated by significantly negative dominance parameter estimates in the full model. In particular, three  $F_1$  crosses involving line 8 showed no hint of responding to conditioning. This pattern is supported by the analysis based on the AIC. The most parsimonious model (Table 1) included all six dominance coefficients which were significant in the full model, although  $[d_{1 \times 6}]$  was only marginally significant when the most parsimonious model was used as the base for likelihood ratio tests. Additionally, the most parsimonious model included a negative (but nonsignificant) dominance coefficient between lines 1 and 3. The pattern of significant

reduction of performance of  $F_1$  crosses was only observed for the learning score. Neither realized fecundity (Fig. 1, below diagonal) nor unconditioned (innate) preference for orange vs. pineapple (results not shown) showed any deviation from additivity (all  $P > 0.2$ ).

The assays of response to olfactory conditioning in the T-maze test demonstrated very poor performance of  $F_1$  crosses between *high-learning* lines 1 and 8 for both short- and long-term memory; in the latter the  $F_1$  did not show any response (Fig. 2). This contrasts with very high memory scores of line 1 and reasonably high scores of line 8 (Fig. 2). The difference between the two parental lines (estimated by twice the additive composite parameter  $[a]$  in Table 2) was not significant for either score, but the trend for line 1 to perform somewhat better than line 8 parallels results from other assays of short- and long-term memory (F. Mery, J. Pont, T. Preat, T. J. Kawecki, unpublished data). Overall, the pattern of performance of all crosses for both memory scores fits very well the additive-dominance model (Fig. 2, Table 2).

The response of the  $F_1$  between *high-learning* lines 1 and 8 to olfactory conditioning seemed no better than typical responses of the unselected control lines (short-term memory score usually between 0.1 and 0.2, long-term memory score around 0.05; F. Mery and T. J. Kawecki, unpublished data). In an additional assay



**Fig. 1** Learning score in the oviposition test (above diagonal) and realized fecundity (below diagonal) of  $F_1$  crosses between pairs of *high-learning* (mean  $\pm$  SE). In each panel the left and right symbol correspond to the respective *high-learning* lines and the middle symbol to the  $F_1$ . Asterisks indicate significant deviations from an additive model (\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.0001$ ).

**Table 1** The most parsimonious model fitted to the performance of the five *high-learning* lines and  $F_1$  crosses in the oviposition learning test.

Parameter	Estimate	SE	P-value
$m_1$	0.28	0.05	0.0000
$m_3$	0.21	0.04	0.0000
$m_5$	0.15	0.04	0.0001
$m_6$	0.13	0.04	0.0012
$m_8$	0.19	0.04	0.0000
$[d_{1 \times 3}]$	-0.07	0.05	0.11
$[d_{1 \times 5}]$	-	-	-
$[d_{1 \times 6}]$	-0.09	0.05	0.051
$[d_{1 \times 8}]$	-0.28	0.05	0.0000
$[d_{3 \times 5}]$	-	-	-
$[d_{3 \times 6}]$	-0.11	0.04	0.0031
$[d_{3 \times 8}]$	-0.20	0.06	0.0008
$[d_{5 \times 6}]$	-0.12	0.04	0.0081
$[d_{5 \times 8}]$	-0.19	0.03	0.0000
$[d_{6 \times 8}]$	-	-	-

Missing values indicate that the corresponding parameter was excluded from the model. The significance of parameters was tested with likelihood ratio tests by dropping the focal parameter from the model. AIC = 20.0, goodness to fit:  $\chi^2 = 2.0$ , d.f. = 3,  $P = 0.57$ .

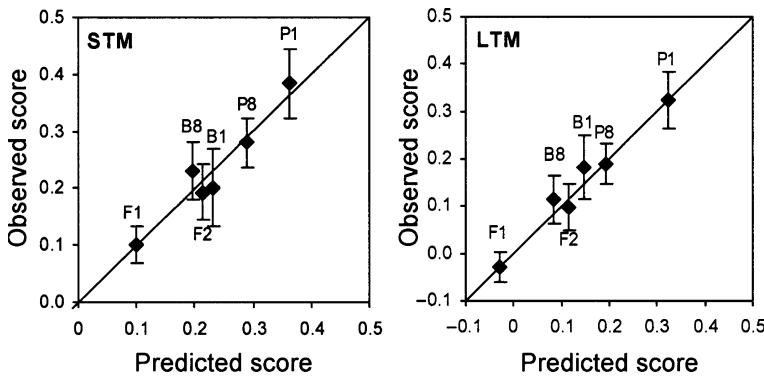
directly comparing short-term memory of the  $F_1$  with the mixed *low-learning* population, the former even tended to perform less well than the latter (mean memory scores  $\pm$ SE,  $F_1$ :  $0.19 \pm 0.03$ , mixed *low-learning* population:  $0.23 \pm 0.5$ ;  $t$ -test,  $t = 0.68$ ,  $n = 12$ ,  $P = 0.5$ ). Thus, the response to selection as measured by the learning score is lost entirely in the  $F_1$  cross between these two *high-learning* lines.

In the absence of conditioning flies from *high-learning* lines 1 and 8, the  $F_1$  cross between them, and the mixed *low-learning* population showed the same avoidance of odours (Fig. 3; two-way ANOVA on arcsine-transformed proportions, line  $F_{3,24} = 0.2$ ,  $P = 0.92$ , odorant  $F_{3,24} = 1.0$ ,  $P = 0.32$ ). This excludes that the large differences in learning scores observed above are due to differences in olfaction or locomotor response to a repulsive odour.

## Discussion

Crossing the *high-learning* line 1, 3 or 5 with *high-learning* line 8 resulted in complete reversion of the learning performance of the  $F_1$  hybrids to the level of the unselected *low-learning* lines. Additional analysis of crosses between lines 1 and 8 shows that this also held for short- and long-term memory assays in an olfactory learning task, different from the task used to impose selection. The two assays were carried out 85 generations apart (at generation 34 and 119, respectively, from the beginning of selection), so the phenomenon is stable. It indicates that the improved learning performance of line 8 had a different genetic basis than that of lines 1, 3 and 5.

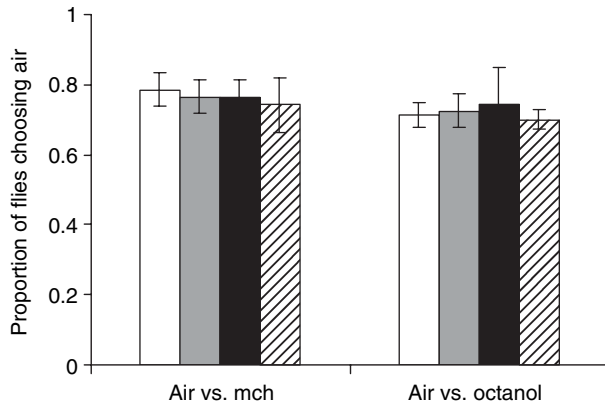
$F_1$  crosses between the *high-learning* lines 1, 3 or 5 and line 6 also showed significant reduction of learning performance compared with the mid-parent expectation, but for the cross between line 1 and 6 this reduction was relatively small. At the same time, the  $F_1$  cross between lines 6 and 8 fitted the additive expectation well, suggesting that the improved learning of lines 6 and 8 had a similar genetic basis. One possibility is that the response to selection in line 6 was mostly based on the



**Fig. 2** Short-term memory (STM) and long-term memory (LTM) scores in the olfactory conditioning paradigm of *high-learning* lines 1 and 8 (labelled P1 and P8) and crosses between them. Observed means  $\pm$  SE are plotted against values predicted by the additive-dominance model.

Parameter	Short-term memory			Long-term memory		
	Estimate $\pm$ SE	$\chi^2$	P-value	Estimate $\pm$ SE	$\chi^2$	P-value
Intercept, $m$	$0.33 \pm 0.03$	98.6	<0.0001	$0.26 \pm 0.04$	34.9	<0.0001
Additive, $[a]$	$0.04 \pm 0.03$	1.2	0.27	$0.06 \pm 0.05$	1.8	0.18
Dominance, $[d]$	$-0.23 \pm 0.05$	21.8	<0.0001	$-0.29 \pm 0.05$	31.6	<0.0001
Goodness of fit, d.f. = 3		1.0	0.80		0.5	0.92

**Table 2** Parameters of the additive-dominance models for short- and long-term memory scores fitted to first- and second-generation crosses between *high-learning* lines 1 and 8.



**Fig. 3** Olfaction assay: the proportion of flies (mean  $\pm$  SE) choosing humid air over either methylcyclohexanol (mch) or octanol. Lines tested: *high-learning* line 1 (white bars) and 8 (black),  $F_1$  cross between them (grey) and mixed *low-learning* population (dashed). Four samples of about 50 females were tested for each line and odour, except for  $F_1$ , for which three samples were tested with methylcyclohexanol and two with octanol.

same allele(s) as that of line 8, but the allele(s) responsible for the response of line 1 were also segregating in line 6 at intermediate frequencies. In the absence of direct genetic data this remains speculation.

No sign of genetic differentiation or any deviation of  $F_1$  from the paternal lines was observed for the unconditioned (innate) resource preference. Likewise, no sign of reduced performance of  $F_1$  crosses was observed for a short-term measure of fecundity; if anything, some crosses tended to show weak (statistically nonsignificant) heterosis. These two traits did not show any correlated response to selection (i.e. did not differ between the selected *high-learning* and the unselected *low-learning* lines, F. Mery and T.J. Kawecki, unpublished data). This indicates that the pattern observed for learning and memory scores is due to a specific response to selection on learning performance rather than to generally poor vigour of crosses between the lines.

For both short- and long-term memory, the performance of second-generation crosses between lines 1 and 8 is consistent with an additive-dominance model. The simplest genetic model consistent with these results assumes that the improved learning of line 8 was due to a recessive allele at one locus, while that of lines 1, 3 and 5 was due to a recessive allele at another locus. In the  $F_1$  both loci would be heterozygous and the ancestral dominant alleles would complement, causing reversion to the ancestral phenotype with poor learning ability. This model is also consistent with preliminary data from  $F_1$  crosses between the *high-* and *low-learning* lines (F. Mery and T. J. Kawecki, unpublished data), which indicate that the improved learning phenotype is mostly or fully recessive. The very good fit of the additive-dominance model implies no epistasis, so it predicts that

flies homozygous for both selected alleles would show better performance than both parental lines (transgression). However, such double homozygotes would only occur in  $F_2$  at frequency 1/16 (if the two loci were on different chromosomes) or not at all (if the loci were on the same chromosome; there is no crossing-over in male *Drosophila*). So the failure to detect epistasis does not preclude an antagonistic interaction between alleles responsible for better learning in lines 1 and 8. In reality more than one locus may be involved in each line. Unfortunately, we cannot assign learning scores to individual flies, so we cannot use the comparison of variance between the different types of crosses to estimate the minimum number of loci, or the recombination rate assuming two loci (Lynch & Walsh, 1998, Chapter 9).

Irrespective of the genetic details, the loss of evolved phenotype in the  $F_1$  crosses between some pairs of *high-learning* lines indicates that the response to selection for improved learning had a different genetic basis in some lines than in others. This is rather remarkable because the lines were derived from the same base population. Given the population size and the rapidity of the response to selection, the favoured alleles are unlikely to have been new mutants. The alleles responsible for the response to selection must thus have segregated in the original base population. The simplest explanation for the genetically idiosyncratic responses would be founder effects – only a subset of alleles that could lead to improved learning might have been sampled in each selection line. This possibility cannot be excluded. However, the mean learning score of the selected populations under the conditions of the selection regime increased from nearly 0 to about 0.15 within about 20 generations, and then reached a plateau at about 0.2. Assuming (as the data suggest) that it was due to fully or mostly recessive alleles, such a rapid response indicates that the initial frequencies of favoured alleles were unlikely to be smaller than about 0.1 (a detailed argument and a supporting model are described in the Appendix). Assuming an allele frequency of 0.1 in the base population, its sampling variance due to founding a selection line with 150 diploid individuals is  $(0.1)(0.9)/300 = 0.0003$ . Even if the real sampling variance was somewhat larger (some of those individuals might not have reproduced), the likelihood of losing the allele in one generation due to the founder effect is negligible. Subsequent loss of the allele due to drift is also unlikely, given the initial frequency of the order of 0.1, relatively strong selection, and a population size of 150 adults. For those reasons, the fact that the response of different selection lines was based on different alleles would be difficult to explain by founder effects and drift alone. One may thus speculate that the alleles which were behind the response of different lines to selection might show antagonistic epistasis, under which either one or the other allele would increase under selection, but not both simultaneously.

An alternative mechanism preventing both alleles from simultaneously increasing under selection would be tight linkage with strong negative linkage disequilibrium in the base population. In the absence of direct evidence this remains a speculation.

This study has demonstrated genetic divergence of replicated populations under uniform selection, which occurred within a very short evolutionary time in populations derived from the same base population. Furthermore, the genetic architecture of this divergence effectively led to outbreeding depression, i.e. the loss of performance of hybrids relative to the original selected lines. Although divergence in direct or correlated responses to uniform selection has been observed in several experiments (see the Introduction), the few reports of crosses between replicate selection lines that we found in the literature (Cohan *et al.*, 1989; Blows, 1993; Bult & Lynch, 1996; Boake *et al.*, 2003) include no cases of outbreeding depression. One study (Blows, 1993) even reports that weak outbreeding depression for developmental time initially observed between *D. serrata* populations of different origin was eliminated by 14 generations of laboratory selection for desiccation resistance. One might argue that a complex behavioural character like the one studied here is particularly likely to be regulated by complex interactions among a large number of loci. Selection on such a character would not only be more likely to lead to idiosyncratic genetic changes, but would also be more sensitive to changes in the genetic background (van Swinderen & Greenspan, 2005). However, replicate lines of mice selected for another complex behaviour (nest building) showed heterosis rather than outbreeding depression when crossed with each other (Bult & Lynch, 1996). More studies are needed before we can make any generalizations concerning the likelihood that parallel populations under the same selection regime will follow idiosyncratic evolutionary trajectories.

## Acknowledgments

We thank L. Sygnarski and N. Vouilloz for help with the experiments, and two referees for comments on a previous version of the manuscript. This work was supported by a Swiss National Science Foundation grant to TJK.

## References

Bieri, J. & Kawecki, T.J. 2003. Genetic architecture of differences between populations of cowpea weevil (*Callosobruchus maculatus*) evolved in the same environment. *Evolution* **57**: 274–287.

Blows, M.W. 1993. The genetics of central and marginal populations of *Drosophila serrata*. II. Hybrid breakdown in fitness components as a correlated response to selection for desiccation resistance. *Evolution* **47**: 1271–1285.

Boake, C.R.B., McDonald, K., Maitra, S. & Ganguly, R. 2003. Forty years of solitude: life-history divergence and behavioural isolation between laboratory lines of *Drosophila melanogaster*. *J. Evol. Biol.* **16**: 83–90.

Bult, A. & Lynch, C.B. 1996. Multiple selection responses in house mice bidirectionally selected for thermoregulatory nest-building behavior: crosses of replicate lines. *Behav. Genet.* **26**: 439–446.

Burnham, K.P. & Anderson, D.A. 1998. *Model Selection and Inference*. Springer Mathematics, New York.

Calboli, F.C.F., Gilchrist, G.W. & Partridge, L. 2003. Different cell size and cell number contribution in two newly established and one ancient body size cline of *Drosophila subobscura*. *Evolution* **57**: 566–573.

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

Cohan, F.M. & Hoffman, A.A. 1989. Uniform selection as a diversifying force in evolution: evidence from *Drosophila*. *Am. Nat.* **134**: 613–637.

Cohan, F.M., Hoffmann, A.A. & Gayley, T.W. 1989. A test of the role of epistasis in divergence under uniform selection. *Evolution* **43**: 766–774.

Fenster, C.B., Galloway, L.F. & Chao, L. 1997. Epistasis and its consequences for the evolution of natural populations. *Trends Ecol. Evol.* **12**: 282–286.

Gilchrist, G.W., Huey, R.B., Balanya, J., Pascual, M. & Serra, L. 2004. A time series of evolution in action: a latitudinal cline in wing size in South American *Drosophila subobscura*. *Evolution* **58**: 768–780.

Gillespie, J.H. 1973. Natural selection with varying selection coefficients – a haploid model. *Genet. Res. Camb.* **21**: 115–120.

Gould, S.J. 1989. *Wonderful Life: The Burgess Shale and the Nature of History*. Norton, New York.

Graves, J.L., Toolston, E.C., Jeong, C., Vu, L.N. & Rose, M.R. 1992. Desiccation, flight, glycogen, and postponed senescence in *Drosophila melanogaster*. *Physiol. Zool.* **65**: 268–286.

Hartl, D.L. & Clark, A.G. 1997. *Principles of Population Genetics*. Sinauer Associates, Sunderland, MA.

Johnson, N.A. 2000. Gene interactions and the origin of species. In: *Epistasis and the Evolutionary Process* (J.B. Wolf, E.D. Brodie & M.J. Wade, eds), Oxford University Press, New York.

Johnson, P.A., Lenski, R.E. & Hoppensteadt, F.C. 1995. Theoretical analysis of divergence in mean fitness between initially identical populations. *Proc. R. Soc. Lond. B* **259**: 125–130.

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

Kawecki, T.J. & Mery, F. 2003. Evolutionary conservatism of host preference in *Callosobruchus maculatus*. *Ecol. Entomol.* **28**: 449–456.

Korona, R. 1996. Genetic divergence and fitness convergence under uniform selection in experimental populations of bacteria. *Genetics* **143**: 637–644.

Korona, R., Nakatsu, C.H., Forney, L.J. & Lenski, R.E. 1994. Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. *Proc. Natl. Acad. Sci. U.S.A.* **91**: 9037–9041.

Lande, R. 1983. The response to selection on major and minor mutations affecting a metrical trait. *Heredity* **50**: 47–65.



- Lynch, M. 1991. The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution* **45**: 622–629.
- Lynch, M. & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.
- MacLean, R.C. & Bell, G. 2003. Divergent evolution during an experimental adaptive radiation. *Proc. R. Soc. Lond. B* **270**: 1645–1650.
- Mather, S.K. & Jinks, J.L. 1982. *Biometrical Genetics*. Chapman & Hall, London.
- Mery, F. & Kawecki, T.J. 2002. Experimental evolution of learning ability in fruit flies. *Proc. Natl. Acad. Sci. U.S.A.* **99**: 14274–14279.
- Mery, F. & Kawecki, T.J. 2003. A fitness cost of learning ability in *Drosophila melanogaster*. *Proc. R. Soc. Lond. B* **270**: 2465–2469.
- Mery, F. & Kawecki, T.J. 2005. A cost of long-term memory in *Drosophila*. *Science* **308**: 1148–1148.
- Orr, H.A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* **139**: 1805–1813.
- Pedra, J.H.F., Festucci-Buselli, R.A., Sun, W.L., Muir, W.M., Scharf, M.E. & Pittendrigh, B.R. 2005. Profiling of abundant proteins associated with dichlorodiphenyltrichloroethane resistance in *Drosophila melanogaster*. *Proteomics* **5**: 258–269.
- Riley, M.S., Cooper, V.S., Lenski, R.E., Forney, L.J. & Marsh, T.L. 2001. Rapid phenotypic change and diversification of a soil bacterium during 1000 generations of experimental evolution. *Microbiology-UK* **147**: 995–1006.
- van Swinderen, B. & Greenspan, R.J. 2005. Flexibility in a gene network affecting a simple behavior in *Drosophila melanogaster*. *Genetics* **169**: 2151–2163.
- Teotonio, H. & Rose, M.R. 2001. Perspective: reverse evolution. *Evolution* **55**: 653–660.
- Travisano, M., Mongold, J.A., Bennett, A.F. & Lenski, R.E. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**: 87–90.
- Tully, T., Preat, T., Boynton, S.C. & Del Vecchio, M. 1994. Genetic dissection of consolidated memory in *Drosophila*. *Cell* **79**: 35–47.
- Ungerer, M.C., Linder, C.R. & Rieseberg, L.H. 2003. Effects of genetic background on response to selection in experimental populations of *Arabidopsis thaliana*. *Genetics* **163**: 277–286.
- Welch, J.J. 2004. Accumulating Dobzhansky–Muller incompatibilities: reconciling theory and data. *Evolution* **58**: 1145–1156.
- Whitlock, M.C., Phillips, P.C., Moore, F.B.G. & Tonsor, S.J. 1995. Multiple fitness peaks and epistasis. *Annu. Rev. Ecol. Syst.* **26**: 601–629.
- Wolfram, S. 1999. *The Mathematica Book*, 4th edn. Wolfram Media/Cambridge University Press, Champaign, IL, USA.
- Wood, T.E., Burke, J.M. & Rieseberg, L.H. 2005. Parallel genotypic adaptation: when evolution repeats itself. *Genetica* **123**: 157–170.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* **16**: 97–159.

## Appendix

### Estimating initial allele frequency

In this appendix, we make educated guesses about the frequencies of the alleles favoured by selection in the base population, based on the observed dynamics of

mean learning score in the course of experimental evolution and population genetic theory. This requires that we specify the relationship between the learning performance and fitness. Under our selection regime fitness was proportional to the number of eggs laid on the ‘correct’ medium within a 3 h time window. Assuming a constant total number of eggs, fitness would thus be proportional to the proportion of eggs laid in that period on the correct medium. In the absence of learning the flies laid about 57% of eggs on the orange and 43% on the pineapple medium. Because orange was the correct medium in odd-numbered and pineapple in even-numbered generations, the proportion of eggs laid on the ‘correct’ medium by a genotype that does not learn would be on average 0.5. (Strictly speaking, when direction of selection alternates between generations a geometric mean should be used (Gillespie, 1973), which in this case would be 0.495. This makes a negligible difference, compared with other approximations in this model.) From the definition of the learning score (see Material and methods), the proportion of eggs laid on the correct medium by a genotype with a learning score  $z$  would be on average  $0.5(1 + z)$ .

We assume first that the response to selection in any given line is due to spread of a ‘high-learning’ allele at a single locus. Under the conditions of the selection regime, the learning score of the initial base population was indistinguishable from zero, so it is reasonable to assume that the homozygote for the ‘low-learning’ allele shows no learning. Thus, taking the fitness of the ‘low-learning’ homozygote as 1, the relative fitnesses of the heterozygote and the homozygote for the ‘high-learning’ allele would be  $1 + hz$  and  $1 + z$ , respectively, where  $hz$  and  $z$  are the learning scores of these two genotypes. From the standard model of selection in a single-locus system (Hartl & Clark, 1997), the change in the frequency  $p$  of the ‘high-learning’ allele would then be described by the recurrence:

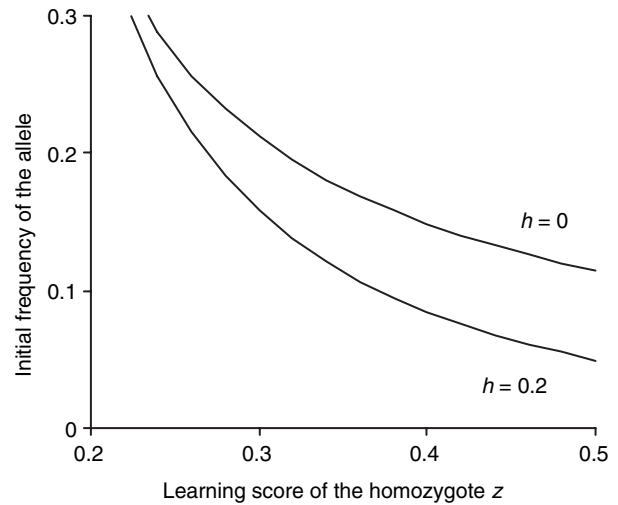
$$p' = p \frac{1 + (1 - p)hz + pz}{1 + 2p(1 - p)hz + p^2z} \quad (\text{A1})$$

(Hartl & Clark, 1997), where the prime denotes the frequency in the next generation. The mean learning score of the population would be given by

$$\bar{z} = p^2z + 2p(1 - p)hz. \quad (\text{A2})$$

In the course of experimental evolution, the mean learning score relevant for selection reached about 0.15 within about 20 generations and then appeared to plateau at around 0.2 (Mery & Kawecki, 2002; also F. Mery and T.J. Kawecki, unpublished data). (Learning scores reported for some lines in this paper are somewhat higher because here we scored the response within 3 h of the end of conditioning, whereas during selection the flies were selected for a response between 3 and 6 h after conditioning; see Material and Methods.)

We used equations (A1–A2) to back-calculate, for a range of  $z$  and  $h$  values, the initial frequency of the favoured allele that would lead to the evolution of the mean learning score of 0.15 within 20 generations (Fig. A1). For a fully recessive allele and assuming that the learning score of the homozygote  $z$  is about 0.2 or 0.25, the model predicts the initial frequency of the favoured allele to be  $>0.2$ . The predicted initial frequency is somewhat smaller if the allele is not fully recessive ( $h = 0.2$ ), but even then the initial frequency is below 0.1 only if  $z > 0.35$  is assumed. Yet, with such high value of  $z$  it would be difficult to explain why the mean learning score reached a plateau at about 0.2; with  $z = 0.35$  and  $h = 0.2$  this would require that the frequency of the ‘high-learning’ allele stabilize at 0.7. Such an intermediate equilibrium might be due to negative effects of the ‘high-learning’ allele on some other aspects of fitness. This is possible, e.g. the *high-learning* lines show a reduced larval competitive ability under highly limiting food (although apparently not under the food conditions used in the selection; Mery & Kawecki, 2003). But if this was the case the effective selection coefficient in favour of the ‘high-learning’ allele would be smaller than the one predicted by the learning score alone, so the predictions in Fig. A1 would underestimate the initial allele frequency. Simple calculations (details not shown) indicate that the initial allele frequencies would have to be even higher if the response to selection were based on two loci with equal and additive effects. These arguments strongly suggest that the initial frequencies of



**Fig. A1** The initial frequency of an allele for improved learning, predicted by a single-locus model of selection to result in the mean learning score of 0.15 after 20 generations of selection. The results are plotted as a function of the learning score of the homozygote for the favoured allele  $z$ , for two values of its dominance coefficient  $h$ ; the homozygote for the other allele is assumed to show no learning.

the ‘high-learning’ alleles are unlikely to have been substantially smaller than 0.1.