J. Evol. Biol. 8: 539–557 (1995)

1010–061X/95/050539–19 $ 1.50 + 0.20/0

Research articles

The differential genetic and environmental canalization of fitness components in *Drosophila melanogaster*

Stephen C. Stearns,[1] Marcel Kaiser[1] and Tadeusz J. Kawecki[2]

[1]*Zoology Institute, University of Basle, Rheinsprung 9, CH-4051 Basle, Switzerland*
[2]*Department of Ecology and Systematics, Division of Population Biology, University of Helsinki, P.O. Box 17, FIN-00014 Helsinki, Finland*

Key words: Canalization; *Drosophila*; fitness components.

Abstract

Canalization describes the process by which phenotypic variation is reduced by developmental mechanisms. A trait can be canalized against environmental or genetic perturbations. Stabilizing selelction should favor improved canalization, and the degree of a trait's canalization should be positively correlated with its impact on fitness. Here we report, for *Drosophila melanogaster*, measurements of environmental canalization for five fitness components. We compare them with measurements of genetic canalization, and we discuss the impact of inbreeding on both. In three experiments we measured the variation of fitness components *within* lines nested within temperature, treatment, and experiment. Lines differed in the position of a *P* element insert or in genetic background. Within lines flies were genetically nearly identical. We designated trait variation within lines as environmental canalization. The canalization of the traits increased with their impact on fitness, and the pattern was similar to that found for the canalization of fitness components against genetic differences, measured as the variation *among* lines nested within temperature, treatment, and experiment. This suggests that developmental mechanisms buffer the phenotype against both genetic and environmental disturbance. The results also suggest, less strongly, that inbreeding weakens canalization.

Introduction

The phenotype has at least two key roles in evolution. One concerns natural selection: phenotypic variation among individuals in reproductive success *is* natural selection. Another concerns the expression of the genetic variation that enables a

response to selection: developmental mechanisms determine, for each environment, how much of what kind of genetic variation will be expressed, a process in which genotype by environment interactions are strongly involved. For whole organisms, the results of these processes can be described under two headings, phenotypic plasticity and canalization.

Much, but not all, of phenotypic plasticity can be described by reaction norms. A reaction norm, a property of a genotype, describes the set of phenotypes expressed by a genotype across a defined range of environmental conditions (Woltereck, 1909; Stearns, 1989). The analysis of phenotypic plasticity has recently generated considerable interest and debate (Via et al., 1995).

Canalization describes the process by which phenotypic variation due to genetic or environmental disturbances is reduced by developmental mechanisms. Waddington (1957) thought that each trait has one developmental pathway buffered by mechanisms against all types of disturbances, environmental or genetic. Here we consider the possibility that genetic and environmental canalization describe independent processes. Genetic canalization is a reduction in the average effect of an allele on a trait (Wagner, 1988, 1989). Environmental canalization is a reduction in the phenotypic variation caused by an environmental perturbation (Waddington, 1942). Both types of canalization should evolve by selection for genes that modify a trait's devlopmental expression.

For example, Prout (1962) showed that disruptive selection reduced the environmental canalization of developmental time in *Drosophila melanogaster*, and Scharloo et al. (1967) showed that a reduction in the phenotypic component of variation in *Drosophila* wing vein proportions did evolve under stabilizing selection. In a theoretical study of the standard one-locus model, Curnow (1964) showed that under stabilizing selection and with genes that have much stronger effects on the variance of a trait than on its mean, selection favors the least variable genotype unless the heterozygote is less variable than either homozygote.

Phenotypic plasticity and canalization are not opposites; they combine in canalized reaction norms (Stearns and Kawecki, 1994). The bundle of reaction norms that describes the genetic variation in response to an environmental gradient can be more tightly constricted in one environment than in another; canalization can be environment-specific (cf. Waddington, 1957, p. 136).

The concept of canalization has a long history. Schmalhausen (1938, 1949) referred to canalization as autonomous development. He saw its primary role as the progressive liberation of the development of the organism from the influences of the environment and foresaw (p. 195) the possibility of canalized reaction norms. Waddington (1940) implied canalization, the developmental buffering of the phenotype, when he defined the epigenetic landscape, and was the first to use the term (Waddington, 1942). Lerner (1954) suggested that heterozygosity would stabilize the phenotype and proposed a genetic model of canalization. Thoday (1958) showed that directional selection broke down genetic homeostasis as measured by the asymmetry of stenopleural bristle number in *Drosophila*, but this breakdown in homeostasis was not due to increasing homozygosity. Instead, Thoday implicated changes in modifier loci. Waddington (1960) demonstrated the effectiveness of

canalizing selection on eye facet number in *Drosophila*. Rendel (1967) and Scharloo (1991) reviewed experimental approaches to understanding the canalization of morphological traits, mostly selection experiments on bristle number, wing veins, and anal papillae in *Drosophila*.

Previous work in our laboratory showed that components of fitness in *D. melanogaster* – age and size at eclosion, early and late fecundity, and lifespan – are differentially canalized to genetic perturbations: the more important the trait is to fitness, the more strongly it is canalized (Stearns and Kawecki, 1994). In extending that work, we measured environmental canalization as the environmental component of the coefficients of variation of those traits and analyzed the relationship of canalization to fitness by calculating the sensitivity of fitness to detrimental changes in each of those traits.

In this paper we ask, first, are the components of fitness differentially canalized to environmental perturbations? If they are, the pattern of canalization of traits could be used as a clue to the selection pressures they had experienced.

Second, if they are so canalized, is the pattern similar to that found for genetic perturbation? If it is, then the same developmental mechanisms might be involved in generating both genetic and environmental canalization, the view held by Waddington (1957), Scharloo (pers. comm.), and Wagner (1989).

Third, does inbreeding reduce genetic and environmental canalization? If it does, then one component of inbreeding depression is an increase in variation about the optimal mean phenotype.

Methods

In three experiments of similar design we measured the canalization of five fitness components: age at eclosion, dry weight at eclosion, lifespan, early fecundity, and late fecundity. Their degree of environmental canalization was estimated from the variation among individuals within treatments, lines, and temperatures. The experiments were planned originally to measure the impact on fitness components of inserting an additional copy of the gene for elongation factor, *EF-1α*, into the *Drosophila* genome. The nature of the genetic transformation required us to make measurements both for C-inserts (controls without the additional gene for elongation factor) and for T-inserts (treatments with the additional gene for elongation factor) at 25° C and 29.5° C. After these experiments were planned and executed, it was discovered that the gene for elongation factor was not expressed in any of the lines at either temperature (Shikama et al., 1994; Ackermann and Gasser, in prep.). Thus the variation between treatments is due not to overexpression of the gene but to the different lengths and insert positions of the constructs.

Differences among treatments

T (treatment)-lines consisted of *D. melanogaster* transformed with a *P*-element containing the ry^+ marker plus the *F1* copy of the gene for elongation factor *EF-1α*

flanked by initiation and termination sequences for heat shock protein. C (control)-lines had been similarly manipulated but lacked the gene for elongation factor; their insert was shorter and at a different position in the *Drosophila* genome (treatment constructs: 11.8 kb, control constructs: 9.9 kb, 16% smaller).

Differences among experiments

Experiment I was done with inbred flies. Three T- and three C-lines (provided by W. Gehring) were first maintained on our standard medium at 25° C for 4 generations. In Expt. I all flies were held at 25° C until eclosion; adults were tested either at 25° C or at 29.5° C. Longevity and fecundity were measured with 40 flies per line, age and weight at eclosion with 24 flies per line.

Experiment II was done on outbred flies with six different genetic backgrounds. Because they resulted from the cross of inbred lines derived from different populations, they were probably more heterozygous than wild flies. The T- and C-lines were obtained from a jumpstart cross in which one arranges for P elements already inserted in the genome to jump to new insert positions. The use of balancer chromosomes ensured that the only surviving flies carried the plasmid at some position on the third chromosome (Bellen et al., 1989). After one fly with the desired marker combinations was crossed again with the balancer stock, a brother-sister cross between flies with the correct markers yielded flies homozygous for the plasmid construct.

The test lines in Expt. II were obtained by outcrossing the one inbred T- and one inbred C-line with six different inbred lab stocks to yield six pairs of heterozygous T- and C-lines with the genetic backgrounds to be tested. Genetic backgrounds were provided by isofemale lines held in the laboratory since 1987 and maintained by full-sib mating for about 60 generations.

In Expts. II and III larvae and pupae were reared at the same temperature as the adults.

Experiment III was done on outbred flies with one genetic background, that of Line 5 in Expt. II. The test flies were obtained by outcrossing six inbred T- and six inbred C-lines from the jumpstart cross with one inbred lab stock to yield heterozygous treatment and six heterozygous control lines with the insert position to be tested. At least five of the six smaller C-inserts were on the third chromosome – C1: 75C, C2: 64C, C3: 89B, C4: 85D, C5: 61C. The position of the sixth C-insert cannot be determined because the line was lost after the experiment was done and before the inserts were localized. All the larger T inserts were on the third chromosome – at T1: centromeric region, T2: 79F, T3: 96D, T4: 96D, T5: 99B, T6: 99B. T3 and T4 had the same band position, as did T5 and T6.

The macroposition from banding patterns does not resolve insert positions in the DNA sequence, and analysis of effects of inserts on traits indicated that the pairs of treatment lines with the same macropositions often had traits that differed significantly. Treatment lines 3 and 4 (with 3 significant differences out of 24 possibilities, 12 traits × 2 temperatures, Scheffé test for post hoc comparison) and

treatment lines 5 and 6 (with 6 significant differences), with the same macroposition within each pair, differed significantly nearly as often as an average pair of lines (6.6 significant differences) with different macropositions. Furthermore, treatment lines 1 and 2, with different macropositions, differed significantly only once in 24 trait comparisions, and lines 1 and 3 and 2 and 3, with different macropositions, differed significantly 3 times, just as often as did lines 3 and 4, which had ostensibly the same insert position. Therefore we treated every insert as though it were a different position regardless of its macroposition.

The differences among Expts. I, II, and III relevant to the questions posed were these: (a) Expts. II and III were better replicated; there were three C-lines and three T-lines in Expt. I, six C-lines and six T-lines in Expts. II and III. (b) The flies were inbred in Expt. I and outbred in Expts. II and III. (c) The larvae and pupae were reared at 25° in Expt. I, 25° and 29.5° in Expts. II and III. (d) Sample sizes were larger in Expts. II and III than in Expt. I. (e) There were P-element inserts in 2 positions (one T and one C) in Expt. II, in 6 positions (three T and three C) in Expt. I, and in 12 positions (six T and 6 C) in Expt. III.

Note that every fly carried only one P-element insert, and that this insert was in the same position in all individuals within a line. Thus our measures of environmental variation were calculated within sets of flies homogeneous for P-element insert position and type.

Fly culture

All measurements were made at 25° C and at 29.5° C (on different flies at each temperature). Flies were reared from first instar larvae gathered within 8 hours of hatching and reared at 12 larvae per vial. Ages were calculated from the mid-point of the period during which the mothers were allowed to lay eggs; this lasted 4 hours. Here we discuss only the traits of mated females. To measure the lifespan of mated females, we established 4 vials per line in Expt. I and 10 vials per line in Expts. II and III with 10 two-day old virgin females and 15 two-day old males. Three times a week the flies were transferred to new vials and the number of dead flies were recorded until the last fly died. Males were replaced by younger males (aged within one week of having eclosed) from the same line if there were fewer males than females in the vial. Females had access to appropriate substrate for oviposition and did lay eggs.

To measure developmental time and dry weight at eclosion, we placed 12 larvae in each of 10 vials per line. When all larvae had pupated, the vials were placed in an eclosion fractionator (Stearns, et al. 1987) that collected freshly eclosed flies at six hour intervals. After sex was determined, the flies were dried at 50° C for 3 hours and weighted to 0.01 mg.

To measure lifetime egg production, we established 30 vials per line with one female and two males per vial. The laying surface was replaced daily by a new one with a drop of fresh yeast. Dead males were removed and replaced by a young virgin male from the same line. The number of eggs laid by each female in 24 hours

were counted daily until the last female died. For this analysis, fecundity early in life was defined as the number of eggs laid per day from the 4th to the 14th day after eclosion. This was 13–23 days after being laid as an egg at 25° C and 11–21 days at 29° C. Fecundity late in life was defined as the number of eggs laid from the 15th to the 5th day before the death of the last female. This was 32–42 days after being laid at 25° C and 23–33 days at 29.5° C. If a female died, her fecundity was treated as a missing value, not as zero.

Sample sizes

Expt. I. – For lifespan, 40 individuals from each line were tested at each temperature: 3 backgrounds × 2 treatments × 2 temperatures = 480 mated females. For fecundity, 40 mated females from each line, temperature, and treatment were tested (480 total). For age and weight at eclosion, 24 larvae from each line were tested at each temperature (288 total). In this experiment we did not get reliable data for early and late fecundity and lifespan at 29.5° because many of these inbred flies died at that temperature. Therefore we have not reported coefficients of variation for those traits under those conditions.

Expts. II and III. – For lifespan, 100 individuals for each line were tested at each temperature: 6 backgrounds × 2 treatments × 2 temperatures = 2400 mated females. For fecundity, 30 mated females from each line, temperature, and treatment were tested (720 total). For age and weight at eclosion, 120 larvae from each line were tested at each temperature (2880 total, of which about 2400 survived to eclosion).

Statistics

We measured genetic and environmental canalization as the inverse of the appropriate coefficients of variation. Therefore, canalization increases as the coefficients of variation decreases. For each combination of experiment, treatment and temperature, we used the restricted maximum likelihood approach to partition total phenotypic variance between the within-line and among-line components (procedure VARCOMP, option REML, SAS Institute, 1989). By taking square roots of these variance components and dividing them by character mean we calculated among-line CVs, which measure genetic variation (CV_{gen}), and within-line CVs, which measure environmental variation (CV_{env}). The environmental variation was induced by microenvironmental variation of unknown nature and by developmental noise. In a previous paper (Stearns and Kawecki, 1994), we analyzed the CVs calculated directly from the raw data, rather than the variance components. We compared the two methods and found that this difference in definitions makes no difference in qualitative results in this data set.

For each trait we analyzed seven plausible mortality regimes to estimate the percent reduction in fitness caused by a 10% detrimental change in the trait. We chose the seven regimes to span the range of conditions that *Drosophila* encounters

in both field and laboratory. The analysis was taken directly from Stearns and Kawecki (1994), which presents details.

Because the means and the variances of the coefficients of variation were correlated, we did further analysis on log-transformed CV's. The scaling effects to which Bryant (1986) called attention were not problems in this data set. Lande (1979) pointed out that if traits differ in the probability of a zero measurement, then comparison of their CV's will be biased, but there were no zero measurements in our data set. To analyze the relationship between canalization and fitness sensitivity, experiment, treatment and temperature, we performed a three-way analysis of covariance for each of the seven fitness sensitivity scenarios. We also calculated the regressions of environmental CV's on fitness sensitivity for each scenario and each experiment × treatment × temperature combination. All analyses were done with SAS statistics software (SAS Institute, 1989).

To check the relationship between genetic perturbation and canalization, we also analysed the magnitudes of effects on traits of the difference between the small and the large inserts in the experiment on position effects where we could compare the effects on eight female traits of six small inserts and six large inserts at different positions on the third chromosome. We estimated the means of the traits in the groups with small and with large inserts in an ANOVA with lines nested within insert sizes, divided the larger mean by the smaller, multiplied by 100, then substracted 100 to get the absolute value of the perturbation of the trait expressed as a percentage. This was done for measurements made at both 25° and 29.5°. In this case we included three more traits whose relationship to fitness could not be estimated from the demographic model.

Results

First, we report the mean environmental CVs of a series of traits whose impact on fitness is known or can be inferred (Tab. 1): age at eclosion, weight at eclosion, early fecundity, late fecundity, lifespan. In all three experiments, for both treatments, and at both temperatures, there was a general increase in the coefficients of variation from age at eclosion to weight at eclosion to early fecundity to late fecundity. Lifespan was less variable than late fecundity.

Second, we took from Stearns and Kawecki (1994) the sensitivity of fitness to those 5 traits for seven different scenarios spanning a range of plausible mortality regimes (Tab. 2). In general, fitness was most sensitive to changes in age and weight at eclosion, rather sensitive to changes in early fecundity, and not very sensitive to changes in late fecundity or lifespan.

Third, we calculated the regressions of trait variation on fitness sensitivity for all seven scenarios for both temperatures (25° C and 29.5° C). The regressions showed that trait variation was negatively associated with impact on fitness for all scenarios. The coefficient of determination ranged from 0.25 to 0.90. Figure 1 depicts the best and worst fits for genetic and environmental CVs.

Table 1. Environmental coefficients of variation obtained from variance components (CV_{env}), expressed as percentages. EF = elongation factor insert, C = control insert.

Exper.	Temp. (°C)	Treat	Age at eclosion	Dry weight	Early fecundity	Late fecundity	Lifespan
I	25	EF	2.85	12.92	34.03	81.18	42.61
		C	3.36	13.20	57.50	107.11	42.34
	29.5	EF	5.85	10.03	51.81	131.03	40.07
		C	4.39	11.18	43.72	91.00	35.12
II	25	EF	2.86	10.19	29.97	44.69	36.47
		C	3.33	10.44	27.31	49.81	31.49
	29.5	EF	2.81	9.70	30.80	68.98	25.21
		C	3.22	10.19	33.33	73.13	25.63
III	25	EF	4.54	14.45	21.07	98.55	33.24
		C	5.15	15.14	27.05	108.11	37.65
	29.5	EF	4.03	14.11	25.52	155.99	32.47
		C	3.95	12.42	34.41	140.55	28.72

Table 2. Sensitivity of fitness to different traits measured as the percent reduction in fitness resulting from a 10% detrimental change in the trait, for different mortality and fecundity scenarios. The scenarios span a plausible range of conditions from laboratory (low adult mortality) to field (high adult mortality). Fitness was measured as the expected lifetime reproductive success, R_0, or the finite rate of increase of the population over 15 days, e^{15r}, where r is the malthusian parameter calculated from the Euler–Lotka equation. Oviposition was assumed to begin on the third day after adult eclosion; daily fecundity changed with age following the curve proposed by Roff (1981), or it was made independent of age (indicated by 'Const'). In both cases daily fecundity was proportional to dry weight at eclosion. The mortality values are instantaneous daily rates. For details see Stearns and Kawecki (1994).

A.

	Scenario						
	1	2	3	4	5	6	7
Mortality values							
Larval	0.100	0.200	0.050	0.100	0.100	0.100	0.100
Adult	0.400	0.400	0.000	0.400	0.001	0.400	0.000
Fecundity values	Roff	Roff	Roff	Const	Const	Roff	Roff
Fitness measure	R_0	R_0	R_0	R_0	R_0	e^{15r}	e^{15r}

B. Percentage reduction in fitness caused by a 10% detrimental change in each of the following traits for each of the scenarios:

	Scenario						
	1	2	3	4	5	6	7
Age at eclosion	9.52	18.13	4.87	9.52	9.52	17.9	25.8
Dry weight	10.00	10.00	10.00	10.00	10.00	10.30	9.38
Early fecundity	5.92	5.92	6.20	4.41	2.66	5.40	6.50
Late fecundity	0.00	0.00	0.31	0.00	2.59	0.00	0.00
Lifespan	0.00	0.00	0.83	0.00	10.35	0.00	0.00

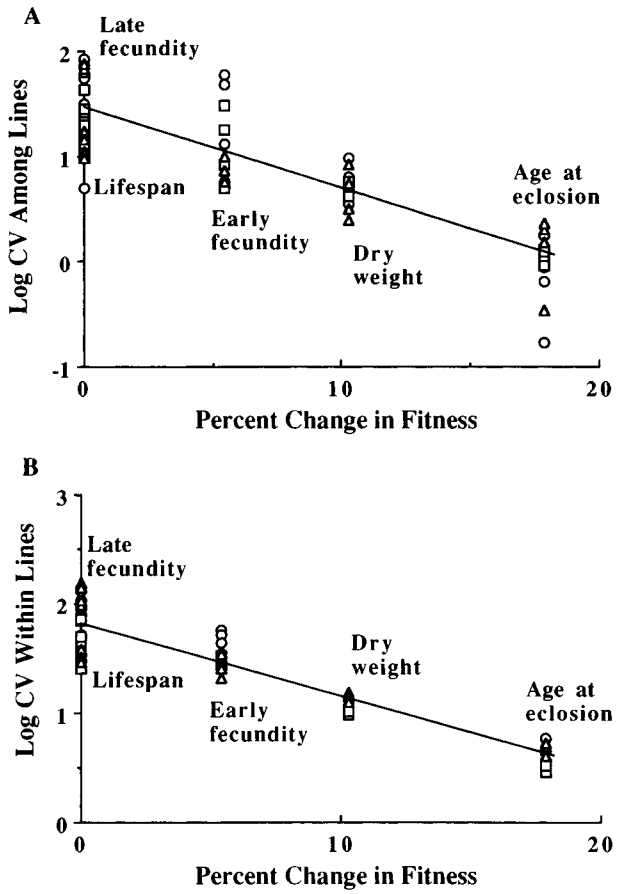

Fig. 1(A), (B).

Fourth, we did a 3-way ANCOVA on the logs of the CVs for each of the seven scenarios and for both genetic and environmental CVs. Table 3 gives an example. The three factors were experiment, temperature, and treatment, and the covariate was fitness sensitivity of the traits. The results were consistent among scenarios. In all seven ANCOVAs on logs of genetic CVs, the only significant factor was the covariate, fitness sensitivity, with $p < 0.0001$ for six of the scenarios and $p = 0.0023$ for Scenario 3, which always gave the worst fit. In no case was experiment, temperature, treatment, or any of the interaction effects significant. For logs of environmental CVs the ANCOVA results were similar, with fitness sensitivity always significant ($p < 0.0001$), and in two cases, Scenarios 2 and 6, experiment effects were also significant ($p = 0.0496$ and $p = 0.0397$).

We therefore reduced the model to main effects – experiment, temperature, treatment, and fitness sensitivity as covariate – allowing the interaction effects to be

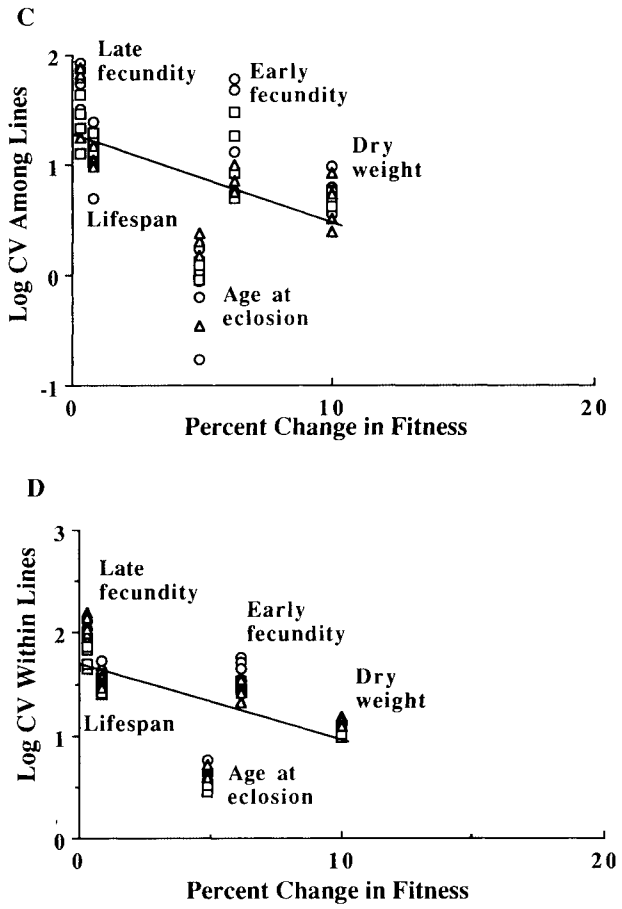

Fig. 1. The best and worst cases for the relationship between logs of CVs and fitness sensitivities of those traits. Best cases (Scenario 6): (A) Genetic, slope = -0.077, $p < 0.001$; (B) Environmental, slope = -0.066, $p < 0.0001$. Worst cases (Scenario 3): (C) Genetic, slope = -0.075, $p < 0.0007$; (D) Environmental, slope = -0.074, $p < 0.0001$.

absorbed in residual and main effects. The patterns of significance did not change. Both genetic and environmental CVs had a highly significant and consistently negative relation to fitness sensitivity (Tab. 4). Again, in Scenarios 2 and 6 experiment effects were also significant ($p = 0.026$ and $p = 0.021$).

Fifth, the genetic and environmental coefficients of variation were similar for all traits (Fig. 2).

Sixth, mortality scenarios 2, 6, and 7, with moderate to high juvenile mortality, consistently resulted in the highest F-values for the relation of log(CV) to the covariate fitness sensitivity (Tab. 4). Scenarios 3 and 5, with moderate to low

Table 3. The full ANCOVA for Scenario 1. Dependent variable: $\log(CV_{gen,rml})$ from variance components. In none of the scenarios were any of the temperature, treatment, or interaction effects significant for either genetic or environmental CVs. In two of the scenarios, 2 and 6, experiment effects were significant in the full ANCOVA for environmental, but not for genetic, CVs.

Source	df	SS	MS	F	p
Experiment	2	0.275	0.137	0.601	0.5532
Temperature	1	0.146	0.146	0.641	0.4280
Treatment	1	0.082	0.082	0.357	0.5534
Fitness sensitivity	1	11.416	11.416	49.965	0.0001
Experiment * Temp	2	0.006	0.003	0.014	0.9859
Experiment * Treatment	2	0.182	0.091	0.398	0.6742
Experiment * FitSens	2	0.160	0.080	0.351	0.7060
Temp * Treatment	1	0.082	0.082	0.360	0.5516
Temp * FitSens	1	0.003	0.003	0.014	0.9073
Treatment * FitSens	1	0.142	0.142	0.620	0.4355
Experiment * Temp * Treatment	2	0.065	0.032	0.142	0.8682
Residual	41	9.367	0.228		

Table 4. All effects with $p < 0.05$ in the ANCOVAs without interactions. (a) Genetic CVs. (b) Environmental CVs. The slope is given for the relation between log(CV) and fitness sensitivity.

Scenario	Effect	F	p	Slope
(a) Genetic				
1	Fitness sensitivity	59.24	0.0001	−0.101
2	Fitness sensitivity	148.81	0.0001	−0.076
3	Fitness sensitivity	12.99	0.0007	−0.075
4	Fitness sensitivity	68.94	0.0001	−0.104
5	Fitness sensitivity	27.79	0.0001	−0.100
6	Fitness sensitivity	150.64	0.0001	−0.077
7	Fitness sensitivity	147.97	0.0001	−0.055
(b) Environmental				
1	Fitness sensitivity	115.38	0.0001	−0.090
2	Fitness sensitivity	357.22	0.0001	−0.065
	Experiment	3.91	0.0260	
3	Fitness sensitivity	23.53	0.0007	−0.074
4	Fitness sensitivity	147.80	0.0001	−0.094
5	Fitness sensitivity	38.61	0.0001	−0.086
6	Fitness sensitivity	383.17	0.0001	−0.066
	Experiment	4.16	0.0210	
7	Fitness sensitivity	276.54	0.0001	−0.046

juvenile mortality and very low adult mortality, consistently yielded the lowest F-values for impact of fitness sensitivity as a covariate.

Seventh, the ANCOVAs for Scenarios 2 and 6 suggested significant effects of experiments on environmental CV's. Analysis of pariwise contrasts suggested that the CVs in Expt. I were larger but marginally insignificant ($p = 0.0804$ for Scenario

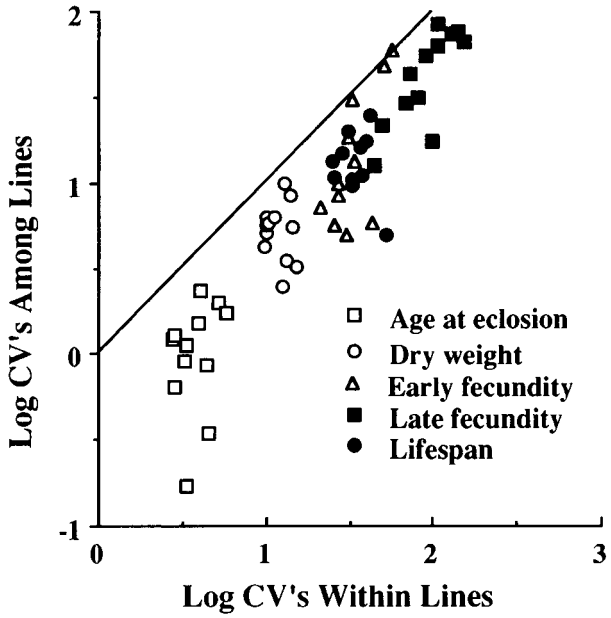

Fig. 2. The genetic (variation among lines) and environmental (variation within lines) coefficients of variation were very similar for each trait but not precisely the same. The line shown is where $CV_{\text{within lines}} = CV_{\text{among lines}}$.

2, $p = 0.0717$ for Scenario 6). In all other scenarios, environmental CVs were larger in Expt. I than in Expts. II and III, but the effects were not significant. Genetic CV's were also larger in Expt. I, but the effect was never significant (the flies in Expt. I were more inbred than the flies in Expts. II and III).

Eight, we calculated the effect of size of insert on the mean of each of eight female traits at both temperatures. The order of the traits ranked by magnitude of effect of size of insert (Fig. 3) was similar to the order of the traits ranked by genetic or environmental CV (Fig. 1), and the pattern was similar at both temperatures.

Discussion

In the introduction we posed three questions: (1) Are the components of fitness differentially canalized to environmental perturbations? (2) Is the pattern similar to that found for genetic perturbations? (3) Does inbreeding have consistent effects on genetic and environmental canalization? We will also discuss (4) whether the selection that produced that canalization is more likely to have occurred in the field or in the laboratory. We managed to answer the first two questions fairly reliably, and we got interesting indications of possible answers to the last two.

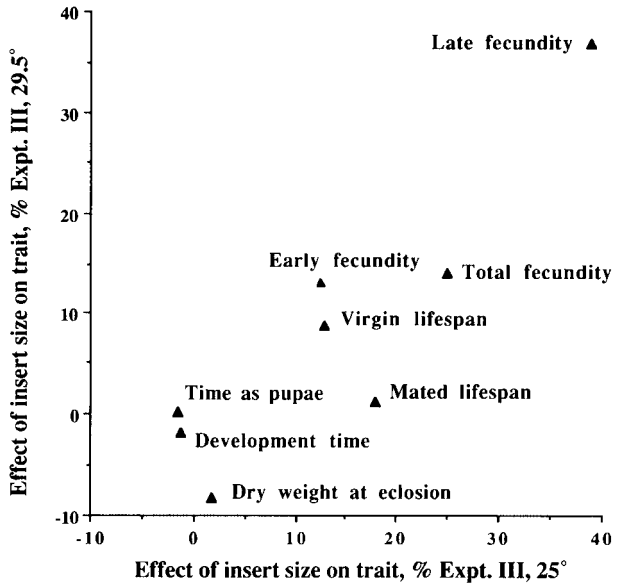

Fig. 3. In Expt. III, in which six small and six large inserts were placed on Chromosome III, the difference in the average effect of large and small inserts was calculated for 8 female traits and expressed as a percentage of trait means. The order of the traits ranked by impact of size of insert was similar to the order of the traits ranked by CVs (Fig. 1), and the effects of insert size on traits was similar at both temperatures ($r = 0.89$, $p = 0.0032$).

Canalization of fitness components

Components of fitness in *Drosophila melanogaster* were differentially canalized to environmental perturbations: the more important a trait was to fitness, the more strongly canalized it was to environmental variation. Our analysis assumes that the traits are either under stabilizing selection or under directional selection that is pushing the trait up against some constraint.

Waddington (1957, pp. 136–138) stated that there should be a direct relation between the canalization of a trait and its impact on fitness, but that this relation would be complicated by two factors: the genetic architecture of the trait (the proportion of additive to interaction variance), and the problem of finding a common scale in which to express the canalization of different traits. By dealing directly with coefficients of variation as an inverse measure of canalization, we think we have found a practical solution to the problem of finding a common scale. Coefficients of variation may not solve all the problems, but they are a reasonable measure with which to start.

As for the problem of genetic architecture, it is certainly a serious problem for comparing traits in artificial selection experiments, where differences among traits in proportion of additive variance would complicate the canalizing response to

stabilizing selection – the context in which Waddington mentioned it. It should be less of a problem for traits whose canalization is close to equilibrium after a long period of selection – the condition that we think we have examined here.

Working with *D. melanogaster*, Prout (1962) applied stabilizing selection on development time to one line (S) and disruptive selection on development time to another line (D) for 40 generations. He found that the coefficients of variation decreased in the S line and increased in the D line. The response in the S line was almost entirely due to a reduction in additive genetic variation, whereas the response in the D line was due mostly to a loss in "buffering capacity." That one can select for decreased environmental canalization for development time in *D. melanogaster* suggests that the trait was in a reasonably well canalized state before the disruptive selection began.

Similarity of genetic and environmental canalization

The patterns of genetic and environmental canalization were very similar (Fig. 2). This is consistent with Waddington's (1957) and Wagner's (1989) idea there is a single developmental pathway for any given trait and that the mechanisms that buffer it do so against disturbances of all types.

Effect of inbreeding on canalization

If heterozygosity improves canalization, then inbreeding should weaken it (Lerner, 1954). It has long been known that inbreeding increases phenotypic variation, which, according to our definition, is equivalent to weakening canalization: Falconer (1989, p. 269) provides a table of six studies in which this effect was found. This idea also lies behind much of the work on fluctuating asymmetry, reviewed by Palmer and Strobeck (1986), where the impact of inbreeding is less consistent than it is on straightforward phenotypic variation.

The data reported here support the idea that inbreeding weakens canalization. The inbred flies of Expt. I were always less well canalized than the outbred flies of Expts. II and III, and while the effect was never significant, it was consistent over all scenarios and both kinds of canalization, and it was almost significant for environmental canalization in Scenarios 2 and 6. These results suggest that an experiment designed explicitly to examine the impact of inbreeding on the canalization of fitness traits would probably give positive results.

Selection regime under which the patterns evolved

If we assume that the degree of canalization of a trait is a good indication of the strength of stabilizing selection on that trait, and that traits with a stronger correlation to fitness should be better canalized than those with a weaker correla-

tion to fitness, then the results presented in Table 4 suggest that the canalization patterns measured here evolved under conditions where juvenile mortality rates were high. We suspect that juvenile mortality rates are lower in the laboratory than in the field. This suggests that for traits like these fitness components, patterns of canalization do not evolve very rapidly in laboratory culture and that the rate of evolution of canalization mechanisms might be slow enough that they could function as an evolutionary constraint. This conclusion is hinted at, but not strongly supported, by our data and deserves further investigation.

Alternative interpretations

Accumulation of errors with age

Most traits used in this study were expressed at different ages, and the increasing variation of traits with age may have resulted from the accumulation of errors with age. We see two ways to test this alternative. First, one could compare the canalization of several traits expressed at the same age but with different impact on fitness. Second, one could locate a representative sample of loci segregating for each trait using QTL (quantitative trait locus) mapping, then measure the average effect of a substitution at loci coding for different traits. This would be a direct test of our interpretation of genetic canalization.

Are the genetic CVs the result of differential heterozygosity?

Could the pattern in the genetic CVs be explained by different numbers of loci segregating for each of the traits measured? The strongest evidence against this interpretation comes from Expt. III, where the genetic differences among lines were due to the different sizes and positions of the P-element inserts but not to differences in genetic backgrounds. To check the results inferred from the pattern of the CVs, we also calculated the effect of size of insert on the mean of each trait. The order of the traits ranked by magnitude of effect of size of insert (Fig. 3) was similar to the order of the traits ranked by genetic or environmental CV (Fig. 1), and the pattern was similar at both temperatures. This result is consistent with the following interpretation: large inserts disturb the genome more than small inserts, and the effect of such disturbance is greater on traits that are less canalized.

While this supports our interpretation of the pattern of genetic CVs as a reflection of differential canalization rather than of different numbers of loci segregating for different traits, it does not demonstrate it conclusively. More direct tests are needed of the magnitude of allelic effects on traits that appear to be differently canalized.

Should the CVs increase rather than decrease with impact on fitness?

Price and Schluter (1991) argued that the reason why life history traits have lower heritabilities than morphological traits is that they are the product of causal interactions over longer time periods and through more complex pathways. These should inflate the environmental variance, thus reducing the heritability. The

implication is that the environmental coefficients of variation for life history traits, which are closely related to fitness, should be larger, not smaller, than those for morphological traits, which are less closely related to fitness. Furthermore, Houle (1992) analyzed the consequences of standardising genetic variances with means, as CVs, rather than with phenotypic variances, as is done in calculating heritabilities. He also found that traits more closely related to fitness were more variable than morphological traits, as measured by genetic CVs. Both Price and Schluter's theory and Houle's results thus appear to contradict the pattern we claim to have found. This raises several interesting issues.

First, neither Price and Schulter (1991) nor Houle (1992) calculated the sensitivity of fitness to changes in a series of traits, as we did; they simply assume that life history traits have greater impact on fitness than morphological traits. This is a plausible assumption, but it is not necessarily true. If one could perturb wing length or stenopleural bristle number by 10% through mutations or changes to the larval environment, one might well discover that such changes were associated with high fitness costs in the form of viability, development time, or fecundity.

Second, different evolutionary biologists make different implicit assumptions about the relation of traits to fitness. Specialists on life history evolution, like us, assume that all traits are involved in tradeoffs, and that these tradeoffs must be taken into account when calculating the sensitivity of fitness to changes in a trait. For example, in *Drosophila* an increase in fecundity usually involves an increase in weight at eclosion that can only be achieved by prolonging development. This means that neither fecundity nor development time can be under directional selection, for to change one in a direction that increases fitness implies a change in the other that decreases fitness.

Robertson (1955) argued the opposite, that precisely these traits would be under continuous directional selection, fecundity upward and development time downward, a view echoed by Lewontin (1965) and still widespread among evolutionary geneticists. Strong selection, in this view, should reduce the proportion of genetic variance due to additive effects and increase the proportion due to dominance, epistasis, and other interactions.

However, Bürger (1993) has recently shown that this traditional view concentrated too much on the elimination of disadvantageous mutations and did not take sufficient account of the behaviour of advantageous mutants under directional selection in large populations ($N_e > 500$). If there is a slow, steady flow of moderately advantageous mutations into the population, then directional selection will continually be pulling them through the fixation, and the number present at intermediate frequencies at any one time will be large enough to increase additive genetic variance. This is not the case for stabilizing selection (Bürger, 1993), where the traditional view still holds, and one expects a reduction in additive genetic variance with increasing selection pressure.

Thus whether fitness components are under stabilizing or directional selection becomes interesting, for directional selection favors extreme types with two results: it can select for decanalization rather than canalization if the directional selection is not pushing the trait up against some immovable constraint (Thoday, 1958), and it can increase rather than decrease additive genetic variance.

Our results for development time and size at eclosion suggest the opposite, indicating that these traits are under stabilizing rather than directional selection, an interpretation supported by Houle's (1992) analysis (his Tab. 1), which also confirms our ranking of the genetic and environmental coefficients of variation of development time, longevity, and fecundity. The view that development time is under stabilizing selection is also supported by Zwann et al. (1994). They selected development time in *D. melanogaster* in both directions, with significant responses of about +50 h in the high lines and −15 h in the low lines. Their results suggest that directional selection in the original population had not reduced development time to a minimum determined by some constraint.

Some basic assumptions

The discussion in the previous section indicates that the field has not achieved consensus on the following issues: How should the relationship between a trait and fitness be modelled, with or without tradeoffs? Should one expect an increase or a decrease in additive genetic variance for traits under directional selection? Should one expect an increase or a decrease in the canalization of traits under directional and stabilizing selection? Should one assume that fitness components (life history traits) are under directional or stabilizing selection? Is selection on morphological traits necessarily weaker than selection on fitness components? We have assumed that all fitness components are involved in tradeoffs, that they are under stabilizing selection, that stabilizing selection favors canalization, and that stronger stabilizing selection produces stronger canalization. We do not know, a priori, whether life history traits are under stronger selection than morphological traits. Others disagree with our assumptions (Scharloo, pers. comm.; anonymous referee pers. comm.), and that disagreement makes the patterns we present here interesting, for it shows that they elicit reactions that expose implicit assumptions about central issues in evolutionary biology.

Conclusion

We can state two points with some confidence: (1) The canalization of fitness components against environmental perturbations improved with the correlation of those traits with fitness as we calculated it. (2) This pattern is extremely similar to that found for the canalization of fitness components against genetic perturbations. Two other points suggested by these data deserve further investigation: (3) Inbreeding weakens canalization. (4) The patterns of canalization measured in the laboratory may have evolved in the field.

Acknowledgements

We thank Anni Mislin, Barbara Sykes, and Hanni Zingerli for their reliable, cheerful, and precise work. Comments by Michi Doebeli, David Berrigan, Ray Huey, Russ Lande, Isabelle Olivieri, Reinhard

Bürger, Wim Scharloo, Bob Vrijenhoek, and two reviewers improved the presentation. S. C. S and M. K. were supported by the Swiss Nationalfonds (Nr. 31-28511.90). T. J. K. was supported by the Finnish Academy of Sciences.

References

Bellen, H. J., C. J. O'Kane, C. Wilson, U. Grossniklaus, R. K. Pearson and W. J. Gehring. 1989. P-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes and Development 3: 1288–1300.

Bryant, E. H. 1986. On use of logarithms to accomodate scale. Syst. Zool. 35: 552–559.

Bürger, R. 1993. Predictions of the dyanamics of a polygenic character under directional selection. J. Theor. Biol. 162: 487–513.

Curnow, R. N. 1964. The effect of continued selection of phenotypic intermediates on gene frequency. Genet. Res., Camb. 5: 341–353.

Falconer, D. S. 1989. Introduction to Quantitative Genetics. 3rd Edition. Longman Scientific and Technical, Harlow.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. Genetics 130: 195–204.

Lande, R. 1977. On comparing coefficients of variation. Syst. Zool. 26: 214–217.

Lerner, I. M. 1954. Genetics Homeostasis. Oliver & Boyd, Edinburgh.

Lewontin, R. C. 1965. Selection for colonizing ability, p. 79–94. *In* H. G. Baker and G. L. Stebbins (eds.), The Genetics of Colonizing Species. Academic Press, New York.

Palmer, A. R. and C. Strobeck. 1986. Fluctuating asymmetry: measurement, analysis, patterns. Ann. Rev. Ecol. Syst. 127: 391–421.

Price, T. and D. Schluter. 1991. On the low heritability of life-history traits. Evolution 45: 853–861.

Prout, T. 1962. The effects of stabilizing selection on the time of development in *Drosophila melanogaster*. Genet. Res., Camb. 3: 364–382.

Rendel, J. M. 1967. Canalisation and Gene Control. Logos Press, London.

Robertson, A. 1965. Selection in animals: Synthesis. Cold Spr. Harb. Symp. Quant. Biol. 20: 225–229.

Roff, D. A. 1981. On being the right size. Am. Nat. 118: 405–422.

SAS Institute, Inc. 1989. SAS User's Guide: Statistics, Version 5 Edition. SAS Institute, Inc. Cary, NC USA.

Scharloo, W., M. S. Hoogmoed and A. Ter Kuile. 1967. Stabilizing and disruptive selection on a mutant character in Drosophila. I. The phenotypic variance and its components. Genetics 56: 709–726.

Scharloo, W. 1991. Canalization: Genetic and developmental aspects. Ann. Rev. Ecol. Syst. 22: 65–93.

Schmalhausen, I. I. 1938. The Organism as a Whole in Development and Evolution. Moscow. (In Russian, not read).

Schmalhausen, I. I. 1949. Factors of Evolution. The Theory of Stabilizing Selection. Blakiston, Philadelphia. Reprinted 1986 by University of Chicago Press, Chicago.

Shikama, N., R. Ackermann and C. Brack. 1994. Protein synthesis, elongation factor EF-1α expression and longevity in *Drosophila melanogaster*. Proc. Nat. Acad. Sci. USA 91: 4199–4203.

Stearns, S. C. 1989. The evolutionary significance of reaction norms. Bioscience 39: 436–446.

Stearns, S. C. 1992. The Evolution of Life Histories. Oxford University Press, Oxford, England.

Stearns, S. C., T. Diggelmann, M. Gebhardt, H. Bachmann and R. Wechsler. 1987. A device for collecting flies of precisely determined post-hatching age. Drosophila Information Service 66: 167–169.

Stearns, S. C. and M. Kaiser. 1993. The effects of enhanced expression of elongation factor *EF-1α* on lifespan in *Drosophila melanogaster*. IV. A summary of three experiments. Genetica 91: 167–182.

Stearns, S. C. and T. J. Kawecki. 1994. Fitness sensitivity and the canalization of life history traits. Evolution 48: 1438–1450.

Thoday, J. M. 1958. Homeostasis in a selection experiment. Heredity 12: 401–415.

Via, S., R. Gomulkiewicz, G. de Jong, S. M. Scheiner, C. D. Schlichting and P. H. van Tienderen. 1995. Adaptive phenotypic plasticity: Consensus and controversy. TREE (in press, May 1995).

Waddington, C. H. 1940. Organisers and Genes. Cambridge University Press, Cambridge.

Waddington, C. H. 1942. Canalization of development and the inheritance of acquired characters. Nature 150: 563–565.

Waddington, C. H. 1957. The Strategy of the Genes. George Allen & Unwin Ltd, London.

Waddington, C. H. 1960. Experiments on canalising selection. Genet. Res., Camb. 1: 140–150.

Wagner, G. P. 1988. The influence of variation and of developmental constraints on the rate of multivariate phenotypic evolution. J. Evol. Biol. 1: 45–66.

Wagner, G. P. 1989. Multivariate mutation-selection balance with constrained pleiotropic effects. Genetics 122: 223–234.

Woltereck, R. 1909. Weitere experimentelle Untersuchungen über Artveränderung, spezielle über das Wesen quantitativer Artunterschiede bei Daphniden. Verh. Deutsch. Zool. Gesell. 1909: 110–172.

Zwaan, B., R. Bijlsma and R. F. Hoekstra. 1994. Artificial selection for developmental time in *Drosophila melanogaster* in relation to the evolution of ageing: direct and correlated responses. *Evolution* (in press).

Received 24 November 1994;
accepted 27 April 1995.
Corresponding Editor: M. Akam