Testing Drosophila life-history in the field: local adaptation in body size, development time and starvation resistance

Linde, Kim van der

Introduction

Normally, the first paragraphs of an article introduce the context for a study, and the relevant literature that is available. However, this type of introduction would merely be a condensed version of **chapter 1**, in which I explained why I carried out this research. I, therefore, refer to that chapter, instead of giving a new condensed version. The literature review can be found in **chapter 3**. The length of the review warranted a chapter on its own, and gives a broad overview of all the relevant literature for this (and the next chapter). Here I will start with the aim and outline of this chapter, followed by my expectations, before continuing with the 'Material & Methods'. In the 'Results' and 'Discussion' sections, each life-history trait is first examined or discussed independently, after which I focus on the interdependencies between the traits. Finally, I will discuss some more general aspects on which these experiments shed some light.

AIM AND OUTLINE

The aim of this study is to investigate the ecological and genetic covariances among three life-history traits in species of *Drosophila*: development time, starvation resistance, and adult body size using a combination of field and laboratory work. Practically, this has resulted in three experiments, two in the field, and one in the laboratory. Flies were collected at six sites in Panama located on two transects, each with a forest, an intermediate and a grassland site. Twelve species were present in at least three collection sites and the stocks were maintained in an open-air laboratory (See Material & Methods).

The aim of the first field experiment (table 1) was to measure the expression of the three life-history traits in the original field environment. I used all twelve species and this experiment will show whether differences between the habitats exist, and if so,

Table 1: Brief summary of the design principles for the three experiments.

First field experiment: original habitat only	Rationale:	Measuring traits for each population in its own habitat. Twelve species, and all populations of each species.
	Aim:	Gain insight into the realised phenotypic values under the original natural conditions the populations have evolved in. Provide insight into the differences among the populations, within and across species.
Second field experiment: transplantation	Rationale:	Measuring traits for each population of four species in their original, and in the two other habitats within the same field transect.
	Aim:	Gain insight into the relative importance of the genetic, environmental and GxE interaction factors.
Common environment experiment	Rationale:	Measuring traits of each population in a common environment in the laboratory. Twelve species, and all populations of each species.
	Aim:	Gain insight into the genetic differences between the different populations.

whether that variation is consistent over all species and is also habitat-related.

The aim of the second field experiment (table 1) was to measure the expression of different populations in all three habitats within a transect, using transplantation of (sub-)populations. The habitats are so close together that the differences between them are within the natural range of differences the flies encounter when they move between habitats. Four species were used for this experiment, which are representative for all the species. With this experiment, I measured the phenotypic plasticity within the different species as expressed in these field experiments and the consistency of this plasticity between the species. Furthermore, the results will also indicate whether genotype-by-environment interactions at the level of populations are present.

The common environment experiment (table 1) was carried out in the laboratory in the Netherlands, with all populations of the twelve species that were still available. This final experiment will give insight into the genetic differences between the different populations, and whether these differences are consistent over all the species.

EXPECTATIONS

Based on the literature review in **chapter 3**, I have drafted some expectations for the individual traits:

Body size: The published data on temperature selection, phenotypic plasticity, and geographical variation taken together predict that the open habitat will result in smaller individuals, both at the genotypic as well as the phenotypic level.

Development time: The latitudinal cline data and the temperature selection data predict that populations from locations with a lower temperature have genetically shorter development times (when measured in a common environment). However, when measured in the field, I expect the grassland populations to develop faster than the forest populations due to the higher environmental temperatures.

Starvation resistance: In the field, I expect that grassland populations have shorter realised starvation times than forest populations. Furthermore, based on latitudinal clines I expect that opening the canopy will result in genetically adapted populations with higher starvation resistances.

Furthermore, based on the life-history model of Sevenster & van Alphen (1993a, 1993b), I expect adaptation towards shorter development times and lower starvation resistances in the more disturbed habitats.

The transplantation (second field) experiment contains in total four main factors: original (or founding) habitat, experimental habitat, transect and sex. These four factors give rise to eleven interaction factors. To ease the interpretation, most of these main and interaction factors can be grouped into three categories: genetic, environmental and Genotype-by-Environment (GxE) interactions. The relative importance of these three categories sheds light to the evolutionary processes underlining the local adaptation. The genetic category (e.g. original or founding habitat related (interaction) factors) sheds light on the underlying genetic variation in the realised trait. The environmental category e.g. experimental habitat related (interaction) factors) underlines the importance of phenotypic plasticity in the realised trait values. Finally, the GxE interaction category incorporates all interaction factors between the original habitat with the experimental habitat. This last category signals whether asymmetry in the response of different populations to the different environments exists.

Material & Methods

FIELD SITE

The fieldwork was carried out in Panama at the Smithsonian Tropical Research Institute (STRI). Here, the closed canopy forest extends right up to the roads within the Canal Zone area. For this study, I established two transects with three habitats along each: closed canopy forest, open grassland with patches of scrub, and an intermediate zone. The distance between the sites within transects was just a few kilometres to ensure limited impact of large-scale factors, such as climate. Each site was of sufficient size so that it could accommodate a large resident population. This increased the likelihood of local adaptation being more important in shaping the life-history traits than immigration from the neighbouring habitats. However, I could not *a priori* exclude the possibility that mass migration between the habitats occurs, which could result in panmixia without local differentiation. The distance between the transects was larger than the length of the transects themselves so that I could test whether the differences within a transect are caused by habitat-related differences and not by local variation covering both transects simultaneously.

The first transect was located near Summit Gardens whilst the second was close to the town of Maria Eugenia (see figure 1). The two transects meet all the criteria mentioned above. The two forest locations are within the same stretch of forest, but the distance between them is around 10 kilometres. The intermediate and grassland locations are separated by this forest and are not connected by the same type of habitat. The forest sites are covered with closed canopy forest. Human activities such as logging, agriculture, settlements, and the Panama Canal have resulted in open areas with grasses as the dominant plants. Scattered in these grassland sites are patches with scrub and small trees, the remaining area is open grassland. The intermediate sites have a higher canopy cover than the grassland sites but lower than that of the forest sites. The intermediate sites differ somewhat

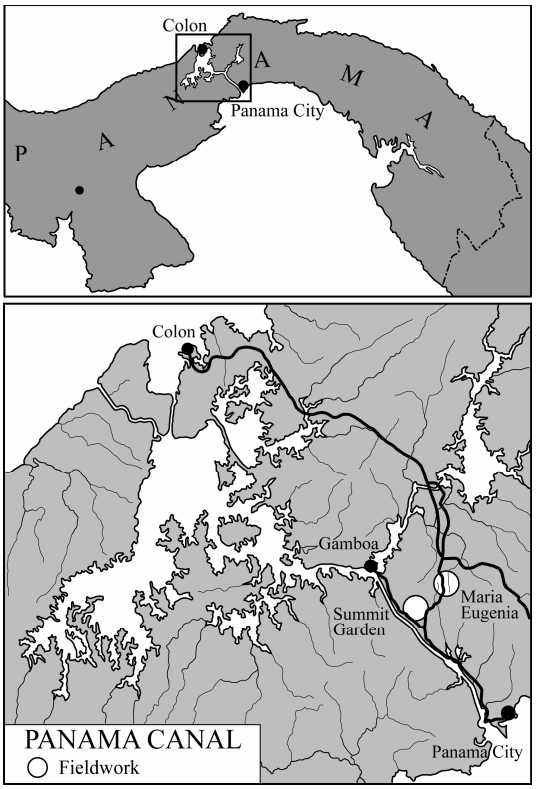


Figure 1: Map of research area. The two circles indicate the two different transects.

from each other in human land use; this might have an influence on the fly populations.

COLLECTIONS AND MAINTENANCE

Drosophila flies were collected using banana-bait traps with up to eight traps per collection site spaced at least 250 meters apart. The traps were constructed from 500-ml transparent containers (7 cm high, and 9-11 cm in diameter) each hanging on a nylon cord of about one-meter in length. A hole of Ø 2.5-cm, covered with 1.5-mm mesh, was positioned on one side of the trap. The hole faced slightly downwards to prevent rain from entering. The mesh allowed *Drosophila* access to the bait inside, but prevented larger animals from entering. The females were identified to the lowest taxonomic level possible (Bock 1980, Val *et al.* 1981) and single females were separately put in small vials with a small piece of banana dipped in yeast suspension. The male offspring were used for definitive identification to the species level. Offspring of iso-female lines of the same species and collection site were combined to give single species stocks. The stocks were maintained in 150-ml containers on pieces of banana dipped in yeast suspension as a breeding substrate and transferred every 10 days to a new vial with fresh breeding substrate.

The open-air laboratory was in Gamboa, under direct influence of the outdoor climate and providing natural variation in ambient temperatures and light regime. Humidity in the closed vials is higher than the ambient humidity.

The experimental space in the field comprised of field cages of approximately 70 x 70 x 120 cm, with a thick wooden floor. The sides were made of gauze (5 mm^2 mesh) and a slightly angled roof extending at the sides for 15 cm. The front was removable for easy access and each cage was on poles of approximately 1.25 meter on which I smeared mineral oil to prevent ants from entering. This set-up kept mammals, birds, and larger insects out. The roof prevented washing away of the experimental set-up (see next paragraph) during heavy rainfall, but also blocked

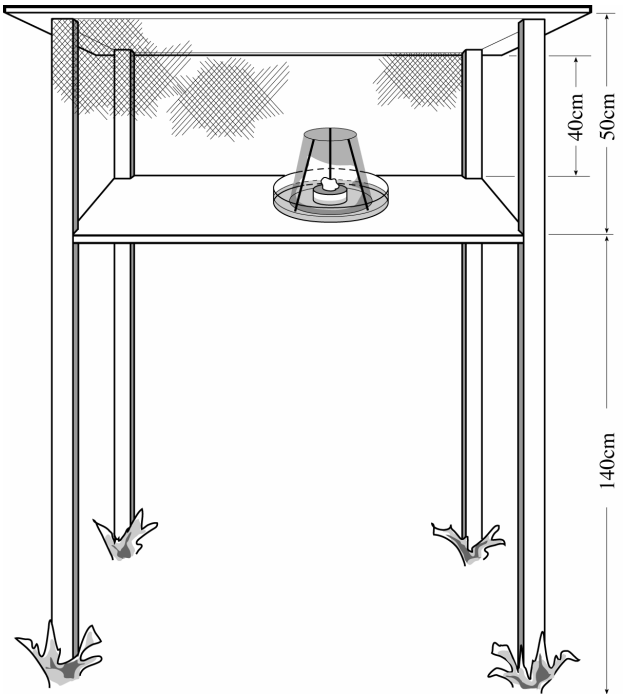


Figure 2: Semi-schematic drawing of the field set-up. The scales of the small cage and the large cage it is located in are not proportional, but modified in order to obtain a clear view of the field set-up.

direct sunshine as insects under natural circumstances can seek shaded places, but could not in these experimental cages (figure 2).

FIELD EXPERIMENTS

Both field experiments had the same basic set-up, in order to reduce differences between them. Each experiment was carried out in three replicas, which were staggered in time, with gaps of three days. The first field experiment was carried out between 21st of August and the end of September 1998, while the second experiment was carried from mid October until mid November.

Field set-up

Instead of vials, I used small cages in the experiments so that the experimental temperature, humidity, and light intensity resembled the ambient temperature more closely. For the development times, I constructed slightly tapered cages of 12 cm high and a diameter of 10 cm made of steel thread and covered with very fine mosquito netting to keep the smallest local parasitic wasps away from the growing larvae. For the egg-laying phase, flies were put on fresh pieces of banana dipped in yeast suspension. This phase was carried out in vials in the open-air laboratory after which the pieces of banana were transported to the experimental cages in the field. The pieces of banana were placed on moist vermiculite in petridishes, over which the small cages were placed. The cages were placed in a thin layer of water, acting as a water-lock preventing entry of insects into the cages. For the starvation resistance, I used petridishes of five cm in diameter covered with the same fine mosquito netting. Five ml of agar was poured inside the petridish to serve as a water source for the flies.

The species for the second field experiment (table 1) were chosen based on their position within the phylogenetic tree and on the relative position within the range of the life-history traits. The four species were representatives for the four main

species groups within this study, *D. melanogaster* (*melanogaster* species group), *D. equinoxialis* (*willistoni* species groups) and *D. sturtevanti* (*saltans* species group), all three of the *Sophophora* subgenus, and *D. cardinoides* (*cardini* species group) from the *Drosophila* subgenus. These species covered the full range in all three traits.

Preparation and egg-laying

Three days before the experiment started, all flies were transferred to a new vial with fresh breeding substrate of banana slices dipped in yeast suspension. This allowed the flies to start producing eggs, and ripe eggs could be laid immediately preventing stowage of eggs. Stowage of eggs can result in shorter development times if the eggs start developing inside the female. On the first day of the actual experiment, groups of flies were transferred to fresh slices of banana dipped in yeast suspension for egg-laying. The parent flies were allowed to lay egg for 6 hours (06:00 - 12:00 hours) in the open-air laboratory after which the pieces of banana were transported to the different field sites.

Development time

Development time was measured as time from egg laying until eclosion as adult. The start of the daily check was as late as possible in the afternoon to have everything checked before sunset, but never before noon. Emerging flies show clear diurnal rhythms (Bakker & Nelissen 1963, Belcher & Brett 1973, Pavan *et al.* 1950): most individuals emerge in the early morning in the first hours after sunrise. Cutting such an emergence peak in two by a random morning visit would result in a random part of the peak counted for that day, while the rest is counted the next day. The daily checking was randomised over transects and sites within the transects to avoid systematic errors as travelling between the sites took considerable time.

Starvation resistance

Starvation resistance was measured as time from adult eclosion until death. In practice, the flies that had emerged between successive checks were transferred to petridishes. Deceased flies were counted once a day simultaneously with the daily check on newly emerged flies. No indications of diurnal rhythms in time of death are known, so randomisation of the counting times was the best way to avoid systematic errors due to variation between counting times and one day in-between periods.

Body size

The flies used in the starvation resistance experiment were stored in vials with 70 % alcohol. Body size was measured using thorax length as a proxy. Both thorax length and wing length are highly correlated with body size (Gu & Barker 1995, Karan *et al.* 1998c, Karan *et al.* 1998b, Karan *et al.* 2000, Parkash *et al.* 1998). However,

wing length shows more interspecific variation than thorax length and was therefore rejected (data not shown). Head width was also rejected, as it, unlike body size, is not affected by temperature (David & Clavel 1967, David *et al.* 1994, Noach *et al.* 1996). Reduction of food levels does not lead to change in the ratio between the characters (Robertson 1987, Thomas 1993).

I measured the length of the thorax between the end of the scutellum and the front rim of the thorax. All measurements were made using a stereomicroscope with drawing mirror and electronic drawing tablet (ACECAD Advanced Digitizer) connected to a computer. Each fly was measured three times and remeasured whenever the variation in the measurements exceeded the 3% tolerance limit. A test run with 25 randomly chosen individuals, measured double-blind two times in random order, showed that the average variation within a series was 0.38 %, while the average variance between the two series was 0.25 % (data not shown). Based on this level of repeatability, one run of three measurements was considered sufficient to obtain reliable data.

Temperature measurements

Variation in temperature is possibly a crucial variable in the explanation of the results. I measured temperatures continuously during the experiments using temperature data-loggers of Onset Computer Corporation (http://www.onsetcomp.com). Data on humidity could not be reliably obtained because condensation short-circuited the measuring element of the data logger.

COMMON ENVIRONMENT EXPERIMENT.

The common environment experiment (table 1) was carried out in the laboratory in Leiden where the flies were kept in a climate room at 25°C, 70-85% RH and 13:11 light:dark. The general set-up was similar to the field experiment, except that I used vials with moist vermiculite for the development time part of the experiment, and tubes with agar for the starvation part of the experiment.

STATISTICS

The "STATISTICA for Windows" software package (versions 5.5 and 6.0) of StatSoft, Inc. (1999, 2004) was used for all statistical calculations unless stated otherwise.

Removing the effect of variation due to sample size

The first step in the analysis was to remove the impact of species and sample size effects on the data. The number of larvae in each piece of banana was uncontrolled and therefore a possible source of errors in the statistics due to crowding (See **Chapter 3**) or Allee effects (Courchamp *et al.* 1999, Hoffmeister & Rohlfs 2001, Rohlfs & Hoffmeister 2003, Stephens & Sutherland 1999, see also Etienne *et al.*

2002, Wertheim *et al.* 2002). I therefore estimated, for each species, a second-degree relationship between the number of flies in the sample and the realised trait values. The residuals of this analysis were used in further analysis. The use of residuals requires a correction in the degrees of freedom for the denominator of the F-distribution. However, when the number of individuals in tests is sufficient large (N > 50), the corresponding impact is small, and therefore the correction can safely be omitted. In case of doubt (results close to the 5% criterion), I tested for the effect, and only cases with significant differences are reported.

Levels of analysis

Collections of flies were made at six locations, three in each transect. In the **basic analysis**, all collection sites were within one single categorical variable: *'collection site'*, while in the **extended analysis**, the six sites were categorised by two variables: *'transect'* and *'habitat'*. The interaction factor is the same as collection site, but without the variation attributable to the main effects. Whenever the two analyses resulted in the same conclusions, the basic analysis was omitted in the results section of that experiment and trait.

There was more than one species in each of the three experiments, which enabled analysis not only at a species level but also at the community level. In general, the analysis commenced at the intraspecific level. For each specific trait, every species was analysed independently for responses to the different factors (**species-specific level**) using the 'Visual General Linear Model' (VGLM) module of STATISTICA with the trait values as the dependent variable, and the different factors and interactions between the factors as the independent variables. The second level of analysis was to combine the species-specific effects, and to test whether all species combined showed a significant response to the factors (**combined-effect level**). This combined effect does not account for the (lack of) similarity between the different species, so opposing species-specific effects could still result in a significant combined effect. The combined effect was estimated using a Fisher-Omnibus test, to examine whether different p-values, as estimated in different tests, show an overall significant effect. These estimates were calculated in a standard spreadsheet program. This test does not consider the direction of the effects, for which the arguments are mentioned above. The rationale behind not using the VGLM module with species as an independent variable is discussed under the next header. The final level of analysis was the overall analysis; to test whether different factors had a similar effect on all species combined (**overall level**). For this analysis, the VGLM module was used in a similar manner to the species-specific analysis, but now with all data from all species.

The variation within an experiment is partitioned into a portion that is explained by the variables in the model (the **explained** variation, which I for clarity will call **'non-error'** variation) and a portion that yields the unexplained variation (**error** component of the model). This error component of the analysis consists of the pure error component, but also all the variation that could have been explained if more variables were added. However, these additional variables are not of interest for the

research questions under investigation, but merely complicate the interpretation of the analysis. One such variable could be the emergence of flies, which takes place over a period of several days and consequently causes a spread in the development times that is both logical and explainable. In addition, body sizes vary in concordance with this spread in development times. Working with averages would have eliminated this aspect in the error component, but also would have decreased the number of data points dramatically and with that, the power of the analysis. I therefore will not indicate the percentage of the total variation that is explained by the different factors, but instead give the proportion of the **'non-error'** variation.

The number of estimated main and interaction factors in a General Linear Model increases exponentially with the inclusion of more independent variables. As valuable as they are, they easily can distort the overall underlying picture. I therefore grouped, when appropriate, the main and interaction factors into broad categories, which are of importance for interpreting the experiment (table 2). For example, in the second field experiment, four main factors are included: sex, transect, original habitat and experimental habitat. This effectively results in 16 estimated main and interaction factors (table 2). For the issue of what proportion of the **'non-error'** variation can be explained by the underlying genetics (e.g. the

Table 2: Grouping of the different main and interaction factors as estimated in the second field experiment analyses, into the broad categories indicative respectively for *'genetic', 'environmental'* and *'GxE interaction*' related factors or *'habitat and collection site'* related factors. See text for a more extensive explanation.

	Grouping into genetic, environmental and GxE interaction related factors	Grouping along habitat and collection site related factors
transect	-	-
original habitat	Genetic	Habitat
experimental habitat	Environment	Habitat
sex	-	-
transect*original habitat	Genetic	Collection site
transect*experimental habitat	Environment	Collection site
original*experimental habitat	GxE	Habitat
transect*sex	-	-
original habitat*sex	Genetic	Habitat
experimental habitat*sex	Environment	Habitat
transect*original*experimental habitat	GxE	Collection site
transect*original habitat*sex	Genetic	Collection site
transect*experimental habitat*sex	Environment	Collection site
original*experimental habitat*sex	GxE	Habitat
transect*original*experimental habitat*sex	GxE	Collection site

adaptive effects of the original habitat), the variation in the '*original habitat'* factor, the '*original habitat*sex'* factor, the '*transect*original habitat'* factor, and the in the *'transect*original habitat*sex'* factor are all of interest. In a similar way, the remaining main and interaction factors can be grouped into an environmental category and a GxE interaction category. Alternatively, grouping can also take place along any other subdivision into categories, such as *'habitat'* (patterns within transect similar) versus *'collection site'* (patterns within transect different).

Troubleshooting

The impacts of different factors on the data were analysed using General Linear Models (GLM), which allows custom-made designs with multiple categorical and/or continuous variables. The analyses are primarily carried out using the Type VI sums-of-squares or 'Effective hypothesis decomposition' (Hocking 1996, StatSoft 1999), and not the often used Type III sum-of-squares. The 'Effective hypothesis decomposition' is based on the philosophy that the estimate should be based only on the variation uniquely attributable to the effect. In an ANOVA design with missing cells, this results in fewer degrees of freedom than in designs without missing cells and for some missing cell designs, the degrees of freedom can drop to zero. Elimination of the higher interaction factors often eliminates at least some of the empty cells in the design and makes estimation of the other variables possible. In those cases, higher interaction factors, which are excluded in the design to obtain useful estimates, are indicated in tables with the word 'Zeroed'. In the case of a nested design, the type VI sum-of-squares cannot be used, and a type III sum-of-squares will be used instead.

The overall analysis is sensitive to disproportional impacts of single species and a jack-knife procedure, excluding one species at the time from the analysis, was used to detect such species. The outcome of the analysis is not robust if the outcome of the analysis was altered by elimination of a single species, e.g. when significant effects became non-significant or vice versa.

Between trait variation

To test whether different traits covaried, I estimated the correlations of all possible two-trait combinations. Homogeneity-of-slopes models (factorial analysis with an interaction factor between collection site and independent continuous variable) were used to test for consistency of the interspecific correlations over the six different habitats.

Inter-experiment comparison

The realised phenotypic life-history trait values are a result of the underlying genotype, the environment, and genotype-by-environment (GxE) interactions. In order to estimate the relative importance of the genetic background on the field values, the population averages of the first-field experiment were plotted against the population averages of the common environment experiment, both for averages

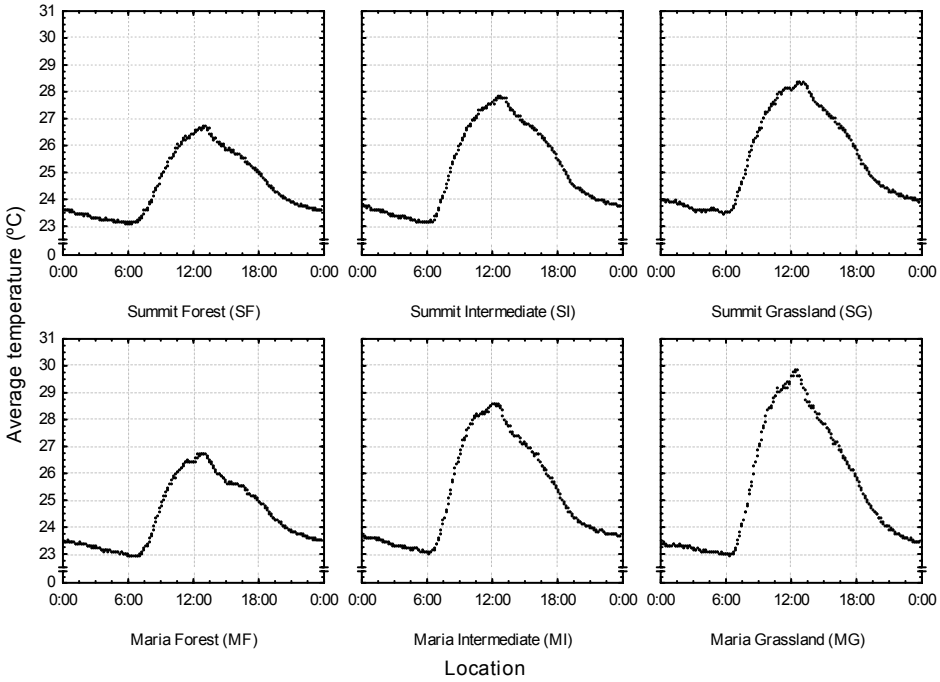


Figure 3: Daily temperature curves for the six habitats during the first field experiment period. The measurements were made between August 21st and September 23rd 1998, with time interval of 6 minutes, and temperatures are averaged per measurement per interval over 33 days. The left two graphs are for the forest locations, the middle two graphs for the intermediate locations and the right two graphs for the grassland populations.

based on the raw data and on the residuals as described before. A change in slope and/or intercept within both types of plots is indicative of the influence of the environmental change on the realised life-history values. When the genetics dominates the realised phenotype, and GxE interactions are absent, the intercept will be zero and the slope of the line will be one, regardless of which averages are used. Positive or negative environmental effects, but also GxE interactions will change the slope and intercept of the fitted line. Both variables are estimated with a degree of error, therefore, Reduced Major Axis regression was used (Bohonak 2002, Bohonak & van der Linde 2004, Kermack & Haldane 1950, Ricker 1973, see also: Sokal & Rohlf 1981).

Results

TEMPERATURES

Temperature measurements were made between August 21st and September 23rd 1998, with time intervals of 6 minutes. Figure 3 shows that the temperatures at the

different collection sites fluctuate widely. Temperature measurements varied between 20 ºC and 35 ºC, but the average midday (between 11:00 and 15:00 hours) and midnight (between 23:00 and 03:00 hours) temperatures are less extreme (table 3). Average day temperature (24 hour period), average midday temperature, day temperature standard deviation, and habitat are all highly correlated with each other. Forest habitats have the lowest day and midday averages, and standard deviations, while grassland has the highest day and midday averages, and standard deviations (habitat - day average: $r = 0.86$, $p = 0.029$; habitat - day SD: $r = 0.83$, $p = 0.042$; habitat - midday average: $r = 0.84$, $p = 0.039$; day average - day SD: $r = 0.89$, $p = 0.017$; day average - midday average: $r = 0.95$, $p = 0.004$; day SD - midday average: $r = 0.99$, $p < 0.001$).

The general picture shows that the daily temperature fluctuations are the largest in the grassland locations, and the smallest in the forest locations, with the intermediate locations nicely in between. The temperature fluctuations between days show the same picture, with the highest fluctuations between days in the grassland locations and the smallest in the forest locations. In more detail, most fluctuation between days occurs around noon (SD's ranging between 1.5 and 3 °C) while the lowest fluctuations are found during the night (SD's between 0.5 and 0.7 °C). The average midnight temperature is much higher in the Summit Grassland series. This location is located directly next to the Panama Canal, which could dampen the daily temperature fluctuations. Average midnight temperature showed no correlation with any of the other variables, and showed a limited variation of just over 0.5 ºC. The main source of variation between the habitats is the midday temperatures, which are much higher in the grassland, while the midnight temperatures are similar in all habitats (figure 3).

Table 3: Average 24 hour, midday and midnight temperatures, and standard deviation for the average 24 hour temperatures per site (in °C).

Temperature regime (°C)

	Summit Forest	Summit Intermediary	Summit Grassland	Maria Forest	Maria Intermediary	Maria Grassland
Average day (0.00 - 24:00 hours)	24.47	25.01	25.38	24.40	25.15	25.39
SD day	1.47	1.94	2.07	1.54	2.27	2.69
Average midday (11:00 - 15:00 hours)	26.35	27.42	27.97	26.34	27.98	28.97
Average midnight (23:00 - 03:00 hours)	23.48	23.62	23.86	23.39	23.54	23.32

SAMPLE SIZES

Body size, development time and starvation resistance are all influenced by high crowding levels resulting in smaller adults (Bakker 1961, Borash & Ho 2001, Chiang & Hodson 1950, Santos 1996, Zwaan *et al.* 1991). Such adults need a longer time to develop (Borash & Ho 2001, Chiang & Hodson 1950, Zwaan *et al.* 1991), and at eclosion, they have a shorter starvation time (Baldal *et al.* in press, Borash & Ho

2001). Significant interaction effects between sample size and the realised trait value were observed in all experiments for all three traits. Not all species showed a significant impact, but the total number of significant results was higher than expected based on the type 1 errors. I therefore used residuals for the rest of the analysis, except when noted otherwise.

BODY SIZE

First-field experiment

As the basic and extended analyses resulted in the same conclusions, only the extended analysis results are presented. In this extended analysis, with the two main factors *'habitat'* and *'transect'*, *'sex'* was highly significant with all individual species showing significant differentiation between the sexes (table 4). Females were larger than males. The *'habitat*transect'* interaction factor, which is equivalent to *'collection site'*, was significant (table 4), as were the underlying *'habitat'* and *'transect'* factors (table 4). The sensitivity analysis showed that transect had a non-significant effect after removal of *D. melanogaster* (after removal: χ^2= 25.43; df=22; p=0.27682), but that both other factors were robust. The overall estimate of the *'intercept'* was significant (table 4) but not robust to removal of *D. malerkotliana*, which resulted in a non-significant estimate (after removal: χ^2 = 30.42, df = 22, p = 0.11). The remaining interaction factors were non-significant (table 4).

In the overall analysis, the error component explained 37.7% of all variation in both the basic as well as the extended analysis. The sexual dimorphism in body size with larger females was confirmed and explained over 99% of all the non-error variation in each analyses ($F_{1;2704}$ = 4406.55, p < 0.0001). It is unsurprising that this sexual size dimorphism overshadows the other factors in the analysis, as these differences are of a completely different nature from those between sites or habitats. Consequently, the non-error variation that can be explained by the remaining factors is very small, but that does not make them less meaningful.

Both the *'site'* and *'site*sex'* factors were significant in the basic analysis (*site*: $F_{5;2704}$ = 4.82, p = 0.0002, non-error variation explained = 0.54 %; *site*sex*: $F_{5;2704}$ = 2.77, p = 0.017, non-error variation explained = 0.31 %). However, the robustness analysis using the jack-knife method revealed that the *'site'* effect was solely attributable to *D. melanogaster* (remaining species: $F_{5;2243}$ = 2.04, p = 0.07). The *'site*sex'* factor was sensitive to three removals of species in the jack-knife procedure, and the jack-knife procedure showed that excluding *D. melanogaster* resulted in a non-significant estimate for the *'site*sex'* factor. The contrast analysis on *'collection site'* varied with the inclusion or exclusion of *D. melanogaster* and was, therefore, not considered further.

In the extended analysis, only *'transect'* and *'habitat*transect*sex'* showed significant effects (*transect*: $F_{1;2704}$ = 15.75, p = 0.0001, non-error variation explained = 0.35 %; *habitat*transect*sex*: $F_{2;2704}$ = 4.64, p = 0.0097, non-error

Table 4: P-values for the different factors as estimated in the extended analysis, for each species separately and based on body size residuals. The overall p-values were calculated with the Fisher-omnibus test. See Material & Methods for 'zeroed' values. The last column contains the sample sizes.

	intercept	habitat	transect	sex	hab*tran	hab*sex	tran*sex	hab*tran*sex	N
D. cardinoides	p=0.8512	p=0.4165	p=0.9917	**p<0.0001**	p=0.4564	p=0.5783	p=0.5410	p=0.9634	54
D. equinoxialis	**p=0.0065**	**p=0.0162**	**p=0.0306**	**p<0.0001**	p=0.5538	p=0.6520	p=0.0611	p=0.2794	283
D. melanogaster	p=0.0607	p=0.1858	**p<0.0001**	**p<0.0001**	**p<0.0001**	**p=0.0414**	p=0.5330	p=0.3698	468
D. malerkotliana	**p<0.0001**	**p=0.0001**	**p<0.0001**	**p<0.0001**	Zeroed	p=0.3765	p=0.2918	Zeroed	133
D. nebulosa	p=0.7413	p=0.2620	p=0.7669	**p<0.0001**	**p<0.0001**	p=0.5234	p=0.1333	p=0.3414	255
D. neomorpha	p=0.8153	**p<0.0001**	p=0.8598	**p<0.0001**	Zeroed	p=0.2162	p=0.9137	Zeroed	62
D. saltans	p=0.7244	**p=0.0035**	p=0.1096	**p<0.0001**	p=0.1356	p=0.1341	p=0.0617	p=0.9107	259
D. simulans	**p<0.0001**	**p=0.0098**	p=0.1153	**p<0.0001**	**p<0.0001**	p=0.2928	p=0.2584	p=0.2058	601
D. septentriosaltans	**p<0.0001**	p=0.1366	p=0.4501	**p<0.0001**	Zeroed	p=0.4993	p=0.4676	Zeroed	87
D. sturtevanti	p=0.0620	p=0.3417	p=0.5876	**p<0.0001**	p=0.8566	p=0.1836	p=0.4646	p=0.1868	127
D. tropicalis	p=0.4310	**p=0.0022**	p=0.1498	**p<0.0001**	Zeroed	p=0.2120	p=0.5956	Zeroed	227
D. willistoni	p=0.4439	p=0.0503	p=0.9628	**p<0.0001**	Zeroed	p=0.1103	**p=0.0244**	Zeroed	184
Chi-square value	46.29	45.2	51.6	>344.60	42.38	14.5	15	5.85	
Df	24	24	24	24	14	24	24	14	
p-value	**p=0.0041**	**p=0.0055**	**p=0.0009**	**p<0.0001**	**p=0.0001**	p=0.9344	p=0.9206	p=0.9702	

variation explained = 0.21 %). In the robustness analysis, *'transect'* seems to be solely accounted for by *D. melanogaster* (remaining species: $F_{5;2243} = 1.64$, $p = 0.20$). When *D. nebulosa* was removed from the species pool, the *'habitat'* and *'habitat*transect'* factors became significant (remaining species: *transect*: $F_{1;2410} = 4.59$, $p = 0.0102$; *habitat*transect*sex*: $F_{2;2410} = 15.75$, $p < 0.0001$). The robustness analysis on the remaining species showed that *'habitat*transect'* was robust, but *'habitat'* was not. All the other interaction factors did not show robust significant results. The '*habitat*transect'* factor showed the same pattern in the contrast analysis as in the basic analysis. The contrast analysis for *'habitat'* showed as expected no differentiation between the habitats, unless *D. nebulosa* was removed from the data set. After removal, the contrast analysis (with superscripts indicating similarity groups) for *'habitat'* showed the following order: Intermediate [a] < Grassland [a, b] < Forest [b].

Most species showed clear intraspecific variation between the samples collected at different sites and part of that variation was related to the differences in habitat. However, in the combined analyses and the overall analysis, the results were not robust. This leads to the conclusion that neither *'habitat'*, nor *'collection site'* had a consistent impact on the realised body sizes in the field despite clear variation within the different species.

Second-field experiment

The four species in the second-field experiment were present in unequal numbers; 171 individuals for *D. sturtevanti*, 210 for *D. cardinoides*, 1208 for *D. melanogaster* and 1239 for *D. equinoxialis*. Thus, the last two species have a relatively large influence on the overall analysis, and one species could easily dominate the outcome of the whole analysis. Therefore, the different species were first analysed independently.

The analysis of *D. cardinoides* suffered heavily from many empty cells, especially in combination with the 'effective hypothesis decomposition' (Hocking 1996) that I used for the sums-of-squares calculations. Many cells in several interaction factors could not be estimated, and none of those factors, which were estimated, were significant. Removal of the highest interaction factor resulted in four significant factors: *'original habitat'* ($F_{1,207} = 8.14$, $p = 0.0048$), '*original*experimental habitat'* ($F_{2,207} = 6.71$, $p = 0.0015$), '*transect*original*experimental habitat'* ($F_{2,207} = 4.52$, $p = 0.012$) and '*transect*experimental habitat*sex'* ($F_{2,207} = 8.08$, $p = 0.0004$). The use of type III sums-of-squares resulted in the same significant results, but also in some additional results which were significant (*intercept*: $F_{1,206} = 26.13$, $p < 0.0001$; *transect:* $F_{1,206} = 12.78$, $p = 0.0004$; *experimental habitat*: $F_{2,206} = 5.52$, $p = 0.0046$; *sex*: $F_{1,206} = 272.55$, $p < 0.0001$; *transect*experimental habitat*: $F_{2,206} = 3.96$, $p = 0.021$).

The interpretation for *D. equinoxialis* was more straightforward. This species showed a clear sexual dimorphism (table 5). Of the other factors, seven were significant. It is noteworthy that the two transects differed significantly (table 5). Of

Table 5:: F-statistics and p-values for the different main and interaction factors as estimated in the second field experiment, each based on body size residuals and for all four species separately. The use of the 'Effective Hypothesis Decomposition' model causes the reduction to zero of the degrees of freedom (see Material & Methods).

Species	*D. cardinoides*			*D. equinoxialis*			*D. melanogaster*			*D. sturtevanti*		
	DF	F	p	DF	F	p	DF	F	p	DF	F	p
Intercept	0		-	1	2.92	0.0875	1	3.63	0.057	0		-
transect (tran)	0		-	**1**	**7.16**	**0.0076**	**1**	**187.85**	**<0.0001**	0		-
original habitat (or)	0		-	**2**	**17.44**	**<0.0001**	**1**	**26.40**	**<0.0001**	1	1.45	0.2309
Experimental habitat (ex)	0		-	**2**	**29.55**	**<0.0001**	**2**	**10.35**	**<0.0001**	1	0.75	0.3875
Sex	0		-	**1**	**4872.01**	**<0.0001**	**1**	**1060.20**	**<0.0001**	0		-
tran*or	0		-	**2**	**4.37**	**0.0129**	**1**	**22.56**	**<0.0001**	1	1.40	0.238
tran*ex	0		-	**2**	**5.35**	**0.0049**	**2**	**6.04**	**0.0025**	1	0.08	0.7733
or*ex	1	2.98	0.0856	4	1.27	0.2795	**2**	**7.81**	**0.0004**	3	0.40	0.7533
tran*sex	0		-	1	0.13	0.723	**1**	**4.76**	**0.0294**	0		-
or*sex	0		-	2	0.81	0.4431	**1**	**4.61**	**0.0321**	1	0.07	0.7905
ex*sex	0		-	2	0.56	0.5697	2	1.34	0.2633	1	0.03	0.8648
tran*or*ex	1	2.41	0.122	**4**	**5.46**	**0.0002**	**2**	**3.37**	**0.0348**	3	1.68	0.1735
tran*or*sex	0		-	2	0.66	0.5159	1	1.01	0.3145	1	3.73	0.0552
tran*ex*sex	0		-	**2**	**4.24**	**0.0146**	2	2.42	0.0891	1	0.28	0.5948
or*ex*sex	1	0.46	0.4984	4	1.35	0.2511	**2**	**3.78**	**0.0231**	**3**	**3.13**	**0.0274**
tran*or*ex*sex	1	1.35	0.2472	4	0.44	0.7765	2	2.02	0.1334	3	1.30	0.2763
Error	206			1201			1180			151		

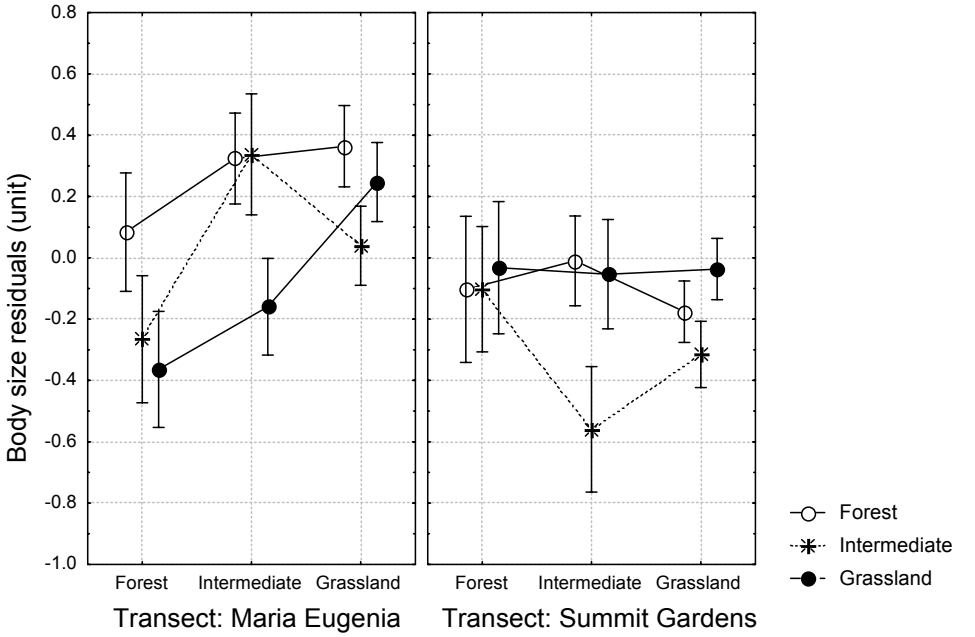


Figure 4: Transect-specific plot for body size residuals (with standard errors) with effects of experimental habitat (lines), original habitat (forest-grassland), and transect (Maria-Summit).

the remaining significant factors, '*experimental habitat', 'transect*experimental habitat'* and *'transect*experimental habitat*sex'* were related to the experimental habitat (table 5), two to the original habitat (table 5: *original habitat; transect*original habitat*) and the last one was related to the interaction between original and experimental habitat (table 5: *transect*original*experimental habitat*).

The analysis of *D. melanogaster* showed the most significant factors. This species also showed a clear sexual dimorphism (table 5), and transect effect (table 5: *transect; transect*sex*). Of the remaining significant factors, two were related to the experimental habitat (table 5: *experimental habitat; transect*experimental habitat*), three to the original habitat (table 5: *original habitat*; *transect*original habitat*; *original habitat*sex*) and the last three to the interaction between original and experimental habitat (table 5: *original*experimental habitat*; *transect*original*experimental habitat*; *original*experimental habitat*sex*).

Finally, *D. sturtevanti* also had some empty cells that made the analysis less straightforward. Removal of the highest interaction factor resulted in a change of the estimated factors: the '*original*experimental habitat*sex'* factor dropped out, while '*transect'* and *'sex'* became significant (table 5). When I analysed the data using the type III sums-of-squares, all three mentioned significant factors remained significant.

The overall analysis for body size suffered somewhat from the low number of species, and the large differences in sample size between the species. The error component explained about 40 % of all variation. Because of this unequal contribution, several factors dropped out in the jack-knife procedure, and only three factors were robust. The species were clearly sexually dimorphic (*sex:* $F_{1,2804}$ = 3961.63, p < 0.0001, non-error variation explained = 95.98 %), leaving a mere 4 % to be explained by the other factors in the analysis. The two other robust factors were *'transect'* ($F_{1,2804}$ = 29.17, p < 0.0001, non-error variation explained = 0.71 %) and '*transect*original*experimental habitat'* ($F_{4,2804}$ = 5.37, p = 0.0003, non-error variation explained = 0.52 %, figure 4). The five factors which were not robust were the *'intercept'* ($F_{1,2804}$ = 4.45, p = 0.035, non-error variation explained = 0.11 %), two factors related to the experimental habitat (*experimental habitat*: $F_{2,2804}$ = 10.24, p < 0.0001, non-error variation explained = 0.50 %; *transect*experimental habitat*: $F_{2,2804}$ = 11.37, p < 0.0001, non-error variation explained = 0.55 %), two to the original habitat (*original habitat*: $F_{2,2804}$ = 4.60, p = 0.01, non-error variation explained = 0.22 %; *transect*original habitat*: $F_{2,2804}$ = 13.08, p < 0.0001, non-error variation explained = .63 %) and the last one to the interaction between experimental and original habitat (*original*experimental habitat*: $F_{4,2804}$ = 2.74, p = 0.027, non-error variation explained = 0.27 %). The contrast analysis (with superscripts indicating similarity groups) showed for *'experimental habitat'* the following order: intermediate ^{a} < grassland ^{b} < forest ^{b} and for the *'origin habitat'*: forest ^{a} < grassland $^{a, b}$ < intermediate ^{b}.

In the overall analysis, almost 41 percent of the variation were attributed to the error component of the model. Of the non-error variation, 95.98 % was explained by the differences between the sexes. Removal of this sex difference by using residuals for the analysis would increase the percentage variation explained by the different categories, but would not change the importance of the different categories relative to each other. About one percent was explained by each of the three categories (table 5): genetic (1.04%), environmental (1.14%) and GxE interactions (1.00%). The remaining variation, which could not be grouped into one of the main categories, was in the *'transect'* (0.71 %) and *'intercept'* (0.11 %) factors. When the factors were grouped according to whether they showed *'collection site'* or *'habitat'* specific differences, 1.97 percent is *'collection site'* related (e.g. showed transect specific variation between the habitats) while 1.21 percent was *'habitat'* related (e.g. habitat related variation which was similar between transects). The remaining variation was in the *'intercept'*, *'sex'* and *'transect'* components. Both *'habitat'* and *'collection site'* related variation is divided among these three categories: genetic, environmental and GxE interactions.

The different species showed clear responses to differences in their habitat. However, they showed hardly any similarity in those responses to the different *'habitats'* or *'collection sites'*. The two transects differed consistently from each other. Females are larger than males, and this was confirmed for all four species. The GxE interaction factor at the *'collection site'* level*,* was the only other factor in this analysis that was robust and significant.

Common environment experiment

The results of the basic and extended analyses are equivalent, so only the results of the extended analysis are presented. The sexual dimorphism in body size for all but one species showed a highly significant effect (table 6), and a strong combined and robust result (table 6). Furthermore, *'habitat'* (table 6: 9 out of 12), *'transect'* (table 6: 8 out of 11) and *'habitat*sex'* (table 6: 4 out of 12) had significant combined estimates, only *'habitat*sex'* was not robust. The remaining interaction factors did not have a significant combined estimate.

The overall analyses showed that 40.9% of the variation was in the error component. Of the non-error variation, nearly 99 % was attributable to the sexual dimorphism present in these species ($F_{1;9010}$ = 12818.6, p < 0.0001, non-error variation explained = 98.61 % (basic), 98.57 % (extended)).

In the basic analysis, all other factors were significant (*site*: $F_{5;9010}$ = 22.23, p < 0.0001, non-error variation explained = 0.85 %; *site*sex*: $F_{5;9010}$ = 5.14, p = 0.0001, non-error variation explained = 0.20 %; *'intercept'*: $F_{1;9010}$ = 44.22, p < 0.0001, non-error variation explained = 0.34 %), and the results were robust for all factors as removal of single species did not change the overall outcome of the test. The contrast analysis showed that Summit-Grassland and Maria-Forest individuals were smaller than the individuals of the remaining sites.

In the extended overall analysis, all but two factors (*'transect'* and *'transect*sex'*) were significant (*transect*: $F_{1;9010}$ = 0.035, p = 0.85, non-error variation explained < 0.01 %; transect*sex: $F_{2;9010}$ = 0.36, p = 0.55, non-error variation explained < 0.01 %). However, the jack-knife procedure showed that *'habitat*sex'* ($F_{2;9010}$ = 7.92, p < 0.0001, non-error variation explained = 0.12 %) and *'habitat*transect*sex'* ($F_{2;9010}$ = 6.41, p = 0.0017, non-error variation explained = 0.10 %) were not robust after removal of respectively *D. saltans* or *D. septentriosaltans*. The remaining two factors, '*habitat'* and *'habitat*transect'* were significant, and robust to jack-knifing the data (*habitat*: $F_{2;9010}$ = 6.21, p = 0.002, non-error variation explained = 0.10 %; *habitat*transect*: $F_{2;9010}$ = 50.27, p < 0.0001, non-error variation explained = 0.77 %). The contrast analysis showed that the grassland individuals were smaller than those from the forest or the intermediate habitat. The contrast analysis for the *'habitat*transect'* factor was consistent with the basic analysis.

The conclusion is that all but one species showed clear differentiation between the different populations. In the overall analysis, both a collection site effect as well as a habitat effect were present. Individuals from Summit-Grassland and Maria-Forest were smaller than individuals from other locations. Individuals collected in the grassland were genetically smaller than those collected in the other two habitats.

Table 6: P-values for the different factors as estimated in the extended analysis, for each species separately and based on body size residuals. The overall p-values were calculated with the Fisher-omnibus test. See Material & Methods for 'zeroed' values. The final column gives the number of individuals per species.

	Intercept	Habitat	Transect	Sex	Site	Hab*sex	Tran*sex	Hab*tran*sex	N
D. cardinoides	**p=0.0335**	**p=0.0002**	**p=0.0250**	**p<0.0001**	Zeroed	p=0.9253	Zeroed	Zeroed	248
D. equinoxialis	**p=0.0017**	**p<0.0001**	**p=0.0019**	**p<0.0001**	**p=0.0171**	p=0.6338	p=0.6898	p=0.1334	1586
D. melanogaster	**p=0.0001**	**p=0.0026**	**p<0.0001**	**p<0.0001**	p=0.5550	p=0.8658	**p=0.0103**	p=0.9155	1081
D. malerkotliana	p=0.4567	**p=0.008**8	p=0.1301	**p<0.0001**	Zeroed	p=0.2447	p=0.9929	Zeroed	136
D. nebulosa	p=0.7917	**p=0.0450**	**p=0.0423**	**p<0.0001**	Zeroed	p=0.3737	p=0.1482	Zeroed	400
D. neomorpha	p=0.7679	p=0.3388	Zeroed	p=0.6124	Zeroed	p=0.1088	Zeroed	Zeroed	10
D. saltans	p=0.2458	**p<0.0001**	p=0.7327	**p<0.0001**	p=0.1342	**p<0.0001**	p=0.2144	Zeroed	958
D. simulans	**p<0.000**1	p=0.5479	p=0.1995	**p<0.0001**	p=0.7157	p=0.8268	p=0.8756	p=0.2345	843
D. septentriosaltans	p=0.2523	**p<0.0001**	**p=0.0073**	**p<0.0001**	Zeroed	**p=0.0082**	**p=0.0404**	Zeroed	707
D. sturtevanti	**p=0.0216**	**p=0.0023**	**p=0.0001**	**p<0.0001**	p=0.4632	**p=0.0324**	**p=0.0064**	**p=0.0415**	395
D. tropicalis	p=0.1695	**p<0.0001**	**p=0.0001**	**p<0.0001**	Zeroed	p=0.0634	p=0.6773	Zeroed	1388
D. willistoni	p=0.2941	p=0.1642	**p<0.0001**	**p<0.0001**	Zeroed	**p=0.0062**	p=0.4744	Zeroed	1270
Chi-square value	36.26	97.31	99.4	352.43	6.75	36.54	15.58	5.85	0
Df	24	24	22	24	10	24	20	8	0
p-value	p=0.0518	**p<0.0001**	**p<0.0001**	**p<0.0001**	p=0.7491	**p=0.0486**	p=0.7426	p=0.6639	0

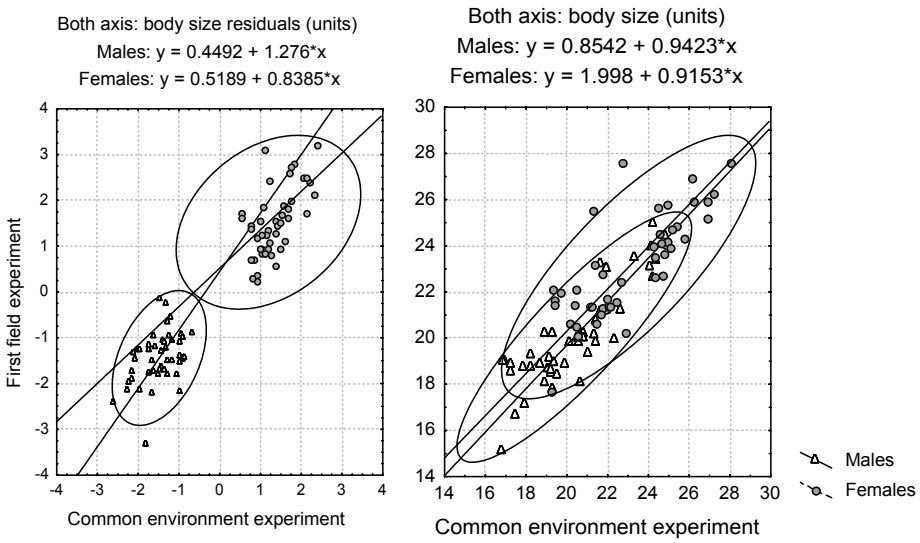


Figure 5: Common environment - first field experiment comparison for body size, full data and residuals (units). Population averages without weight factor, estimated using Reduced Major Axis regression (see Sokal & Rohlf 1981) as both variables are estimated with error. The residuals in the left panel are the same as in the analyses of the specific experiment, estimated over all individuals within a species. Ellipses indicate 95% confidence regions.

Inter-experiment comparison

The correlation across species between the first field experiment and the common environment experiment (figure 5, right panel) is highly significant (males: R^2 = 0.755, n = 47, p < 0.001; females: R^2 = 0.619, n = 47, p < 0.001) and the correlation coefficients do not differ from each other (p = 0.11). The slopes of the Reduced Major Axis (RMA) regression (see Sokal & Rohlf 1981) are close to 1, and are within the bootstrapped 95% confidence intervals (males: slope = 0.9423, range = 0.804 - 1.105; females: slope = 0.9153, range 0.778 - 1.089).

In the inter-experiment comparison using residuals (figure 5, left panel), both sexes showed a significant match between the common environment and first field experiment (males: R^2 = 0.172, n = 46, p = 0.004; females: R^2 = 0.103, n = 47, p = 0.028). The two correlation coefficients did not differ significantly from each other (p = 0.74). The bootstrapped 95% confidence intervals around the slopes of the RMA regression included the x = y line (males: slope = 1.276, range = 0.877 - 1.680; females: slope = 0.8385, range = 0.486 - 1.598).

The comparison between the two experiments shows a high degree of similarity, which suggests that the underlying genetics are more important than the

environment for the realised body sizes. The match between the two experiments is especially striking at the between species comparison, and less so at the within species comparison. This suggests that extrapolation of the common environment experiment to the field situation is possible, especially when patterns across species are considered.

Overall conclusion for body size

Table 7: Overview of body size differences between habitats, regardless of whether the differences between the habitats were significant or not. Superscript indicate groups withn each row.

Experiment 1	Inconsistent
Experiment 2: experimental habitat	Intermediate [a] < Grassland [b] < Forest [b]
Experiment 2: original habitat	Forest [a] < Grassland [a, b] < Intermediate [b]
Common environment	Grassland [a] < Forest [b] < Intermediate [b]

Table 7 shows an overview of the patterns between the different habitats within the different experiments. The results of the second field experiment are split into original and experimental habitat. Within species, genetic and phenotypic variation between the populations was present. At an overall level, the phenotypic consistency between the species was weak and inconsistent. However, at the genotypic level, systematic variation was related to collection site as well as habitat (smaller individuals in the grassland habitat and in Summit-Grassland and Maria-Forest collection sites). This is confirmed by the close match between the realised body sizes in the common environment experiment and the first field experiment. The transplantation (second field) experiment showed that GxE effects played a role, and that this type of effect could explain at least in part the incomplete match in the inter-experiment comparison.

DEVELOPMENT TIME

First-field experiment

The extended analysis resulted in the same conclusions as the basic analysis. In this analysis *'collection site'* was spilt into the underlying *'habitat'* and *'transect'* factors and *'sex'* was significant in eight out of twelve species and in the overall estimate (table 8). Of the remaining factors, only the overall estimate for *'habitat'* was significant (table 8), with four of twelve species showing a significant result. However, this result was not robust, as removal of either *D. malerkotliana* or *D. nebulosa* resulted in a non-significant overall result. The remaining factors were all non-significant and in each case a limited number of species showed significant results.

In the overall analysis, over 91 % of the variation in the data was in the error component for both the basic as well as the extended analysis. Of the variation

Table 8: P-values for the different factors as estimated in the extended analysis, for each species separately and based on development time residuals. The overall p-values were calculated with the Fisher-omnibus test. The last column contains the number of individuals for each species.

	intercept	habitat	transect	sex	hab*tran	hab*sex	tran*sex	hab*tran*sex	N
D. cardinoides	p=0.6414	p=0.6656	p=0.8935	p=0.8355	P=0.0716	p=0.2524	p=0.9713	p=0.2989	97
D. equinoxialis	p=0.1567	p=0.8875	p=0.5891	**p=0.0002**	p=0.8613	p=0.5375	p=0.4522	p=0.2194	638
D. melanogaster	p=0.3696	p=0.1042	p=0.1832	**p<0.0001**	**p=0.0014**	p=0.4080	p=0.2257	p=0.3644	103
D. malerkotliana	**p<0.0001**	**p<0.0001**	**p<0.0001**	p=0.7139	All Zeroed	p=0.4329	p=0.5736	All Zeroed	290
D. nebulosa	p=0.5351	**p<0.0001**	p=0.6290	**p=0.0006**	p=0.1804	p=0.8128	p=0.3265	p=0.4923	621
D. neomorpha	p=0.4869	**p=0.0017**	p=0.6886	p=0.4282	One	p=0.5302	p=0.4144	One Zeroed	112
D. saltans	p=0.3087	p=0.0597	p=0.5002	**p<0.0001**	p=0.0715	p=0.1157	p=0.2576	p=0.3155	516
D. simulans	p=0.8974	p=0.6206	**p=0.0006**	**p<0.0001**	p=0.0804	**p=0.0258**	p=0.1734	**p<0.0001**	139
D. septentriosaltans	p=0.0617	p=0.0608	p=0.1657	p=0.1649	All Zeroed	p=0.6663	p=0.3681	All Zeroed	163
D. sturtevanti	p=0.4296	p=0.4619	**p=0.0345**	**p<0.0001**	p=0.0825	p=0.5048	p=0.8141	p=0.3182	253
D. tropicalis	p=0.0818	**p=0.0075**	p=0.2746	**p<0.0001**	One	p=0.0834	p=0.0819	One Zeroed	456
D. willistoni	p=0.6159	p=0.2843	p=0.2284	**p=0.0002**	All Zeroed	p=0.3614	p=0.3623	All Zeroed	364
Chi-square value	23.99	41.96	34.72	79.84	16.26	13.01	11.03	14.85	
Df	24	24	24	24	14	24	24	14	
p-value	p=0.4620	**p=0.0130**	p=0.0727	**p<0.0001**	p=0.2978	p=0.9660	p=0.9888	p=0.3882	

explained by either model, *'sex'* was the most important component, explaining over two-thirds of the non-error variation (extended: $F_{1;2756}$ = 191.61, p < 0.0001, non-error variation explained = 71.11 %; basic: $F_{1;2756}$ = 191.61, p < 0.0001, non-error variation explained = 71.80 %).

In the basic analysis, *'collection site'* had a significant effect ($F_{5;2756}$ = 7.83, p < 0.0001, non-error variation explained = 14.66 %) as had the interaction factor between *'collection site'* and *'sex'* ($F_{5;2756}$ = 5.77, p < 0.0001, non-error variation explained = 10.82 %) and the *'intercept'* ($F_{1;2756}$ = 7.27, p = 0.007, non-error variation explained = 2.73 %). The jack-knife procedure to examine whether any particular species had a disproportional impact on the basic analysis revealed that the *'intercept'* and the *'collection site*sex'* interaction factors were based on two species (*D. nebulosa* and *D. simulans*) and one species (*D. simulans*), respectively. *D. simulans* was the only species which had a significant result for the interaction factor *'collection site*sex'* (p < 0.0001) in the analysis of the individual species (table 8). A renewed jack-knife analysis with *D. simulans* excluded, showed a significant and consistent effect of *'sex'* ($F_{1;2159}$ = 171.06, p < 0.0001, non-error variation explained = 79.5 %) and *'collection site'* ($F_{5;2159}$ = 6.86, p < 0.0001, non-error variation explained = 15.9 %), while *'collection site*sex'* became non-significant ($F_{5;2159}$ = 1.23, p = 0.29). The *'intercept'* was now not significant (p = 0.056) but was sensitive to removal of four different species. The contrast analysis for *'collection site'* showed that the Maria-Forest populations had significantly longer development times than those from all other collection sites.

Most factors in the extended analysis were significant. After *'sex'*, the interaction factor *'habitat*transect*sex'* explained most of the non-error variation ($F_{2;2756}$ = 8.36, p = 0.0002, non-error variation explained = 6.20 %). The other significant factors explained, respectively, 4.98 % (*habitat*: $F_{2;2756}$ = 6.71, p = 0.0012), 5.61 % (*transect*: $F_{1;2756}$ = 15.13, p = 0.0001), 1.71 % (*habitat*transect*: $F_{2;2756}$ = 6.02, p = 0.0025) and 4.76% (*habitat*sex*: $F_{2;2756}$ = 6.41, p = 0.0017) of the non-error variation, while the non-significant '*transect*sex'* interaction factor explained 0.17 % of the non-error variation ($F_{1;2756}$ = 0.46, p = 0.50). Again, the jack-knife procedure showed a disproportional impact of *D. simulans*. Removal of this species resulted in elimination of all interaction factors with sex. However, the results without *D. simulans* were not robust, and subsequent removal of species resulted in elimination of all factors. The contrast analysis on *'collection site'* confirmed the picture of the basic analysis, in that the Maria-Forest populations had significantly longer development times than those from all other collection sites. The contrast analysis (with superscripts indicating similarity groups) for *'habitat'* revealed the following order: Grassland ^{a} < Intermediate $^{a, b}$ < Forest ^{b}.

This led to the conclusion that the realised development times in the field show a clear relation with *'collection site'*. Partitioning the *'collection site'* component into the underlying *'habitat'* and *'transect'* components did not yield stable results in either the individual analysis, nor in the overall analysis. Removal of *D. simulans* gave much more stable results in the *'collection site'* based analyses, but several species disrupted the overall picture for the extended analyses using *'habitat'* and *'transect'*.

Table 9: F-statistics and p-values for the different main and interaction factors as estimated in the second field experiment, each based on development time residuals and for all four species separately. The use of the 'Effective Hypothesis Decomposition' model causes the reduction to zero of the degrees of freedom (see Material & Methods).

Species	*D. cardinoides*			*D. equinoxialis*			*D. melanogaster*			*D. sturtevanti*		
	DF	F	p	DF	F	p	DF	F	p	DF	F	p
Intercept	0	0	-	1	0.08	0.7816	1	4.16	0.0417	0	0	-
transect (tran)	0	0	-	1	3.37	0.0668	**1**	**15**	**0.0001**	0	0	-
original habitat (or)	0	0	-	2	1.44	0.2381	**1**	**13.94**	**0.0002**	**1**	**3.99**	**0.0477**
Experimental habitat (ex)	0	0	-	**2**	**3.13**	**0.0441**	2	2.61	0.0740	1	0.19	0.6664
Sex	0	0	-	**1**	**114.73**	**<0.0001**	**1**	**56.6**	**<0.0001**	0	0	-
tran*or	0	0	-	2	0.85	0.4279	**1**	**4.64**	**0.0314**	1	0.01	0.9335
tran*ex	0	0	-	2	1.19	0.3033	**2**	**4.88**	**0.0078**	1	0	0.9579
or*ex	1	2.79	0.0966	**4**	**3.18**	**0.0129**	2	1.29	0.2750	3	1.94	0.1254
tran*sex	0	0	-	1	3.48	0.0623	1	0.69	0.4054	0	0	-
or*sex	0	0	-	2	0.23	0.7906	1	2.1	0.1472	1	0.52	0.4733
ex*sex	0	0	-	2	2.31	0.0993	2	2.38	0.0933	1	0.03	0.8534
tran*or*ex	1	0.09	0.7700	**4**	**5.45**	**0.0002**	**2**	**6.22**	**0.0021**	3	0.8	0.4955
tran*or*sex	0	0	-	**2**	**4.01**	**0.0183**	1	0.04	0.8379	**1**	**9.75**	**0.0021**
tran*ex*sex	0	0	-	**2**	**6.13**	**0.0023**	2	1.42	0.2430	1	2.57	0.1112
or*ex*sex	1	0.11	0.7365	4	2.14	0.0733	2	2.71	0.0667	3	0.58	0.6299
tran*or*ex*sex	1	2.08	0.1504	4	2.06	0.0843	2	2.51	0.0815	3	2.19	0.0917
Error	206			1203			1184			151		

Second-field experiment

As for body size, the four species in the second-field experiment were present in unequal numbers, which may influence the overall analysis. Therefore, the different species were first analysed independently.

The analysis of *D. cardinoides* suffered heavily from many empty cells (see Material & Methods), but neither the removal of highest interaction factor nor the use of type III sums-of-squares resulted in significant factors for this species.

The second species, *D. equinoxialis*, showed a clear sexual dimorphism (table 9). Five of the other factors were significant. Two factors were related to the experimental habitat (table 9: *experimental habitat*; *transect*experimental habitat*sex*), one to the original habitat (table 9: *transect*original habitat*sex*) and two for the interaction between original and experimental habitat (table 9: *original*experimental habitat*; *transect*original*experimental habitat*).

D. melanogaster also showed a clear sexual dimorphism (table 9). The impact of the transects was prominent (table 9). The subdivision of the remaining significant factors for this species showed that one factor is related to the experimental habitat (table 9: *transect*experimental habitat*), two factors to the original habitat (table 9: *original habitat*; *transect*original habitat*) and the last to the interaction factor between original and experimental habitat (table 9: *transect*original*experimental habitat*)

The final species, *D. sturtevanti* also suffered from empty cells in the analysis, but less so than *D. cardinoides*. Removal of the highest interaction factor resulted in more estimated factors but in lower estimates for both significant factors, while the factor for *'sex'* could now be estimated and was highly significant. The use of type III sums-of-squares did not give more insight either, other than that this species was also highly sexual dimorphic. The two significant factors were both associated with the original habitat (table 9: *original habitat*; *transect*original habitat*sex*).

Several factors in the overall analysis were significant and the robustness analysis left several factors untouched, despite the large and unequal contribution to the dataset by two species. Most variation was within the error component, (90.76 %). Overall, the species showed a clear sexual dimorphism ($F_{1,2830}$ = 170.0, $p < 0.0001$, non-error variation explained = 58.89 %). Two of the three remaining factors were related to the experimental habitat (*experimental habitat*: $F_{2,2830}$ = 8.36, p = 0.0002, non-error variation explained = 5.8 %; *transect*experimental habitat*sex*: $F_{2,2830}$ = 10.0, $p < 0.0001$, non-error variation explained = 6.94 %), the final one to the interaction between original and experimental habitat (*transect*original*experimental habitat*: $F_{4,2830}$ = 6.9, $p < 0.0001$, non-error variation explained = 9.58 %, figure 6). Three factors dropped out of the list as they were not robust, two related to the original habitat (*original habitat*: $F_{2,2830}$ = 4.61, p = 0.01, non-error variation explained = 3.2 %; *transect*original habitat*: $F_{2,2830}$ = 4.73, p = 0.0089, non-error variation explained = 3.28 %) and one related to the experimental

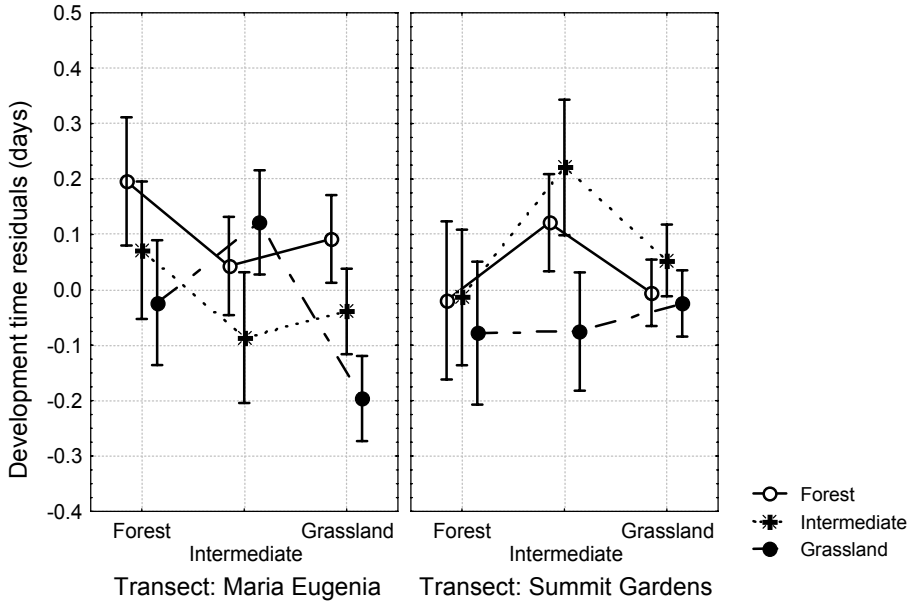


Figure 6: Transect-specific plot for development time residuals (with standard errors) with effects of experimental habitat (lines), original habitat (forest-grassland), and transect (Maria-Summit).

habitat (*transect*experimental habitat*: $F_{2,2830} = 4.70$, $p = 0.0092$, non-error variation explained = 3.26 %). The contrast analysis (with superscripts indicating similarity groups) showed for *'experimental habitat'*: grassland ^{a} < intermediate ^{b} < forest ^{b} and for the *'origin habitat'*: grassland ^{a} < forest $^{a, b}$ < intermediate ^{b}.

All factors except the *'intercept'*, *'transect'* and *'sex'* factors could be attributed to the genetic component, the environmental component, or the GxE interaction component (table 2). Most variation was in the error component, explaining 90.76% of all variation. When the remaining variation was divided over the different categories, 58.89 percent was explained by the sexual dimorphism, 7.91 percent by the genetic component, 16.37 percent by the environmental component and 15.77 percent was in the GxE interaction component. The remaining variation was in the *'transect'* and *'intercept'* factors. When the factors were grouped according to whether they showed differences between the transects or not, 27.56 percent was *'collection site'* related while 12.49 percent was *'habitat'* related (the remaining was in *'intercept'*, *'sex'* and *'transect'* components). Both *'habitat'* and *'collection site'* related variation was divided among all three components: genetic, environmental and GxE interactions.

Overall, it is clear that genetic, environmental and GxE interactions play a role in the expression of the development times between the different habitats. However, there was substantial variation between the different species. The similarity between the

two transects was low as about twice as much variation was related to *'collection site'* than to *'habitat'*.

Common environment experiment

In the extended individual species analysis, which resulted in the same conclusion as the basic analysis, several factors had a significant combined effect. *'Sex'* had the largest combined impact (table 10) and seven out of twelve species showed a significant effect. Half of the species showed a significant effect for *'habitat'*, and the combined estimate was significant and robust (table 10). Six out of eleven species showed a significant *'transect'* effect; the estimate for one species (*D. neomorpha*) was zeroed, as the remaining populations were all from one transect. The combined estimate was robust and significant (table 10). Seven of the twelve estimates for the *'habitat*transect'* factor were zeroed due to missing combinations between habitat and transect. Three out of the remaining five species showed a significant result, as was the robust combined result (table 10). Some of the estimates for the remaining factors were significant, but none of the factors had a significant combined estimate.

Most variation in the basic and extended overall analysis was pure error (basic: 94.79 %; extended: 95.09 %). Both analyses confirmed that females have a shorter development time than males, in both cases explaining 49.26 % of the non-error variation (both: $F_{1,9035} = 229.63$, $p < 0.0001$). *'Collection site'* was the only other significant factor in the basic analysis ($F_{5,9035} = 50.85$, $p < 0.0001$, non-error variation explained = 51.22 %), and the results were robust as none of the species had a disproportional impact on the outcome in the jack-knife procedure. The interaction factor *'site*sex'* and the *'intercept'* were both not significant (*site*sex*: $F_{5,9035} = 1.97$, $p = 0.08$, non-error variation explained: 1.98 %; *intercept*: $F_{1,9035} = 2.71$, $p = 0.10$, non-error variation explained: 0.55 %). The contrast analysis showed three groups as indicated by the superscripts: Maria-Grassland ^{a} < Maria-Intermediate ^{b} < Maria-Forest ^{b} < Summit-Grassland ^{b} < Summit-Intermediate ^{c} < Summit-Forest ^{c}.

In the extended analysis, both the *'habitat'* and *'transect'* factors were significant (*habitat*: $F_{2,9035} = 51.45$, $p < 0.0001$, non-error variation explained: 22.07 %; *transect*: $F_{1,9035} = 120.91$, $p < 0.0001$, non-error variation explained: 25.94 %). Both results were robust using the jack-knife method which showed that none of the species had a disproportional impact on the outcome of the analysis. The remaining (interaction) factors were all non-significant ($p < 0.10$), explaining less than one percent of the non-error variation. Contrast analysis showed that individuals collected in the grassland sites or in the Maria transect had shorter development times and formed a separate group versus the other two types of habitats or the summit transect.

The conclusion is that there is a clear habitat and transect related impact on the development times as measured in the laboratory under a common environment regime. Furthermore, this impact was consistent for all species.

Table 10: P-values for the different factors as estimated in the extended analysis, for each species separately and based on development time residuals. The overall p-values were calculated with the Fisher-omnibus test. Last column contains the number of individuals for each species.

	intercept	habitat	transect	sex	hab*tran	hab*sex	tran*sex	hab*tran*sex	N
D. cardinoides	p=0.7434	p=0.3446	p=0.6413	p=0.3247	Zeroed	p=0.0692	Zeroed	Zeroed	249
D. equinoxialis	p=0.7463	**p<0.0001**	**p=0.0011**	**p<0.0001**	**p=0.0242**	p=0.0850	p=0.1140	p=0.2294	1590
D. melanogaster	**p=0.0010**	**p<0.0001**	**p<0.0001**	**p<0.0001**	p=0.1322	p=0.2292	**p=0.0219**	p=0.8244	1084
D. malerkotliana	p=0.9746	**p=0.0481**	p=0.1197	p=0.4638	Zeroed	p=0.6832	p=0.2894	Zeroed	136
D. nebulosa	p=0.3671	p=0.5976	p=0.8118	**p=0.0396**	Zeroed	p=0.6080	p=0.4458	Zeroed	402
D. neomorpha	p=0.8691	p=0.8691	Zeroed	p=0.2739	Zeroed	p=0.2739	Zeroed	Zeroed	10
D. saltans	**p=0.0163**	**p<0.0001**	**p=0.0353**	p=0.1583	**p<0.0001**	p=0.9266	p=0.8876	Zeroed	959
D. simulans	p=0.7998	**p=0.0027**	p=0.6121	**p<0.0001**	p=0.1000	p=0.5039	p=0.1291	p=0.4216	844
D. septentriosaltans	p=0.3254	p=0.4501	p=0.2010	**p<0.0001**	Zeroed	p=0.9340	p=0.7708	Zeroed	710
D. sturtevanti	p=0.3774	p=0.1120	**p=0.0005**	**p<0.0001**	**p<0.0001**	p=0.7792	p=0.0872	p=0.7071	395
D. tropicalis	p=0.5207	**p=0.0002**	**p<0.0001**	**p<0.0001**	Zeroed	**p=0.0012**	p=0.2744	Zeroed	1393
D. willistoni	p=0.2174	p=0.0532	**p<0.0001**	p=0.2655	Zeroed	p=0.9371	p=0.9670	Zeroed	1275
Chi-square value	14.97	98.31	96.47	112.57	50.1	14.47	12.36	2.5	
Df	24	24	22	23	10	24	20	8	
p-value	p=0.9217	**p<0.0001**	**p<0.0001**	**p<0.0001**	**p<0.0001**	p=0.9352	p=0.9030	p=0.9618	

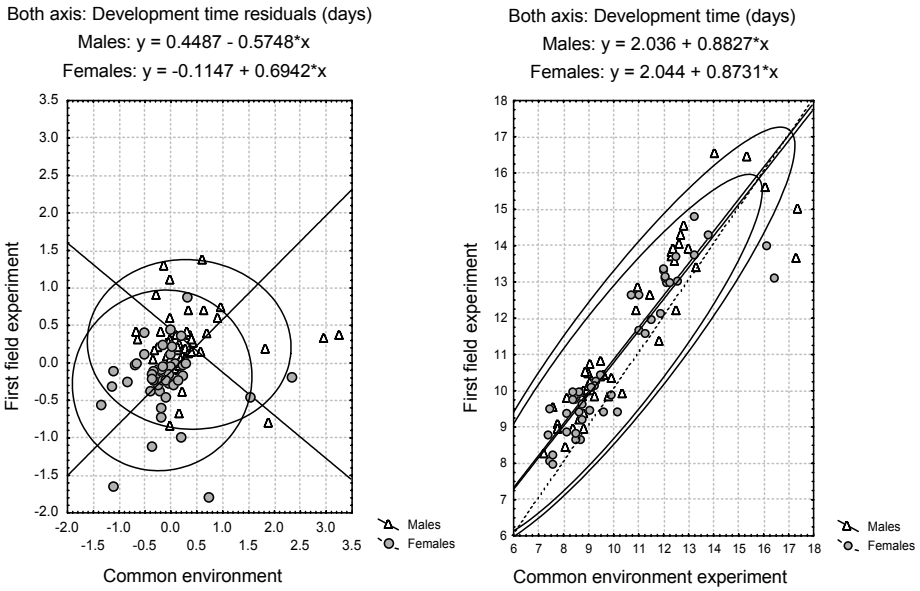


Figure 7: Common environment - first field experiment comparison for development times, full data and residuals (days). Population averages without weight factor, estimated using Reduced Major Axis regression (see Sokal & Rohlf 1981) as both variables are estimated with error. The residuals in the left panel are the same as in the analyses of the specific experiment. Ellipses indicate 95% confidence regions.

Inter-experiment comparison

The inter-experiment comparison on the averages of the raw data (figure 7, right panel) shows a clear correlation between the two experiments. The results of the two experiments were highly correlated, while there were no differences between the sexes (males: $R^2 = 0.831$, n = 47, p <0.0001; females: $R^2 = 0.826$, n = 47, p < 0.0001; comparison for difference: p = 0.47). The bootstrapped 95% confidence intervals of the slopes of the Reduced Major Axis (RMA) regression (see Sokal & Rohlf 1981) include the x = y line (males: slope = 0.8827, range = 0.733 - 1.083; females: slope = 0.8731, range = 0.720 - 1.030).

The same analysis with the averages from the residuals (figure 7, left part) showed a weak and non-significant correlation between the trait values of the common environment experiment and the first field experiment (males: R^2 =0.001, n = 47, p = 0.82; females: $R^2 = 0.002$, n = 47, p = 0.78). The bootstrapped 95% confidence intervals around the slopes of the RMA regression were extremely large, underlining the absence of any correlation between the two experiments.

This result led to the conclusion that the outcome of both experiments is highly correlated at the interspecific level, but not correlated at all at the intraspecific level. The first field experiment was carried out in three different environments, while the

common environment was similar for all populations. This makes the extrapolation of the results obtained in the common environment experiment to the field situation difficult.

Overall conclusion for development time

Table 11: Overview of development time differences between habitats, regardless of whether the differences between the habitats were significant or not.

Experiment 1	Grassland < Intermediate < Forest
Experiment 2: experimental habitat	Grassland < Intermediate < Forest
Experiment 2: original habitat	Grassland < Forest < Intermediate
Common environment	Grassland < Intermediate < Forest

There is evidence that local adaptation has occurred in most of the twelve species in this study. This local selection was *'habitat'* and *'transect'* specific with shorter development times for individuals from the grassland habitats and the Maria transect. However, development times as measured in the field do not show *'habitat'* or *'transect'* specific differentiation, but rather are *'collection site'* related. This difference between the common environment experiment and the first-field experiment suggest that GxE interactions might play a role. These GxE interactions were indeed found in the second-field experiment and explained a larger proportion of the variation than genetic differences alone. The lack-of-fit within species between the first field experiment and the common environment underlines that the realised development times are dependent on all three factors: environment, genetic and GxE interaction.

STARVATION RESISTANCE

First-field experiment

Due to the experimental set-up, it was impossible to match starvation resistance values to development times, body sizes, and sex of the same individual flies. However, the impact of sex on starvation resistance was estimated by regressing the residuals of the sample averages for starvation resistance as the independent variable and the sex ratio of the same sample as the dependent variable. The influence of sex ratio on starvation resistance was non-significant (R^2 = -0.0538, N = 162, p=0.5). Apparently, the variation in the sex ratios among the different samples appeared not to have influenced starvation resistances in the experiment and were not considered further.

The impact of the *'collection site'* was determined using the starvation resistance residuals with *'site'* as a categorical factor. The results in table 12 show that *'site'* was a significant factor in eight out of twelve species. The overall estimate showed that the overall effect was significant and none of the species had a disproportional impact on the combined outcome.

Table 12: P-values for the different factors as estimated in the basic and extended analysis, for each species separately and based on starvation resistance residuals. The final column gives the number of individuals per species. The overall p-values were calculated with the Fisher-omnibus test. (df=24 in each case; log(0) replaced with -16).

	Basic analysis		Extended analysis				
	Intercept	Site	Intercept	Habitat	Transect	Site	N
D. cardinoides	p=0.5413	**p=0.0131**	p=0.5413	**p=0.0117**	p=0.9966	**p=0.0239**	54
D. equinoxialis	p=0.8290	**p=0.0194**	p=0.8290	p=0.0867	p=0.0765	p=0.2983	283
D. melanogaster	p=0.5699	**p=0.0321**	p=0.5699	p=0.0597	p=0.5711	p=0.0921	464
D. malerkotliana	p=0.5552	**p=0.0001**	**p=0.0002**	**p=0.0043**	**p<0.0001**	Zeroed	129
D. nebulosa	p=0.9257	p=0.1128	p=0.9257	p=0.0570	p=0.2490	p=0.6313	297
D. neomorpha	p=0.5900	**p=0.0337**	p=0.5662	p=0.2076	p=0.0960	Zeroed	60
D. saltans	p=0.9057	**p<0.0001**	p=0.9057	**p=0.0002**	p=0.0813	**p=0.0043**	259
D. simulans	p=0.6602	**p=0.0004**	p=0.6602	**p=0.0182**	p=0.0643	**p=0.0317**	597
D. septentriosaltans	p=0.5960	p=0.1167	p=0.7285	p=0.0637	p=0.6509	Zeroed	83
D. sturtevanti	p=0.5205	p=0.0996	p=0.5205	p=0.2217	p=0.1453	**p=0.0220**	127
D. tropicalis	p=0.4735	**p<0.0001**	p=0.3360	**p<0.0001**	p=0.4702	Zeroed	225
D. willistoni	p=0.8851	p=0.2923	p=0.3881	p=0.1193	p=0.2828	Zeroed	178
Chi-square value	4.44	58.81	12.24	48.22	24.23	17.81	
Df	24	24	24	24	24	14	
p-value	p=1.0000	**p=0.0001**	p=0.9771	**p=0.0024**	p=0.4483	p=0.2154	

The extended analysis showed that the starvation resistance of five out of twelve species was significantly affected by *'habitat'*, as was the combined estimate (table 12). However, this combined result was not robust, as removal of *D. tropicalis* yielded a non-significant result (after removal: χ^2 = 33.67, df = 22, p = 0.053). Populations of *D. malerkotliana* showed significant differentiation between the two transects, while four out of the twelve species showed a significant *'habitat*transect'* interaction. However, the combined estimates for '*transect'*, *'habitat*transect'* and intercept were not significant (table 12).

In the overall analysis, most variation in the data was in the error component (basic and extended: 98.69 %). In the basic analysis, *'collection site'* had a significant effect ($F_{5;2921}$ = 7.58, p < 0.0001, non-error variation explained = 97.99 %), while the *'intercept'* was non-significant ($F_{1;2921}$ = 0.78, p = 0.38, non-error variation explained = 2.01 %). The jack-knife procedure showed that none of the species had a disproportional impact on the outcome of the analysis. The contrast analysis revealed three groups (with superscripts indicating different groups): Maria-Grassland ^{a} < Summit-Intermediate $^{a, b}$ < Summit-Grassland $^{b, c}$ < Maria-Intermediate $^{b, c}$ < Summit-Forest $^{b, c}$ < Maria-Forest ^{c}.

In the extended overall analysis, both the *'habitat'* and *'habitat*transect'* factors were significant (habitat: $F_{2;2921}$ = 8.64, p = 0.0002, non-error variation explained = 51.2 %; habitat*transect: $F_{2;2921}$ = 8.60, p = 0.0002, non-error variation explained = 47.56 %). The result was, however, not robust as *D. simulans* had a disproportional impact on the *'habitat*transect'* factor (after removal: $F_{2;2298}$ = 1.57, p = 0.21). The renewed analysis with *D. simulans* omitted showed that *D. tropicalis* now had a disproportional impact on the outcome of the *'habitat'* factor (after removal of both species: $F_{2;2065}$ = 2.81, p = 0.06). The two remaining factors were non-significant (*transect*: $F_{1;2921}$ = 0.001, p = 0.97; *intercept*: $F_{1;2921}$ = 0.78, p = 0.38). The contrast analysis showed two groups (with superscripts indicating different groups): grassland ^{a} < intermediate $^{a, b}$ < forest ^{b}.

The overall conclusion is that flies show clear differences in starvation resistance. Flies from the forest habitats had higher starvation resistance than the intermediate populations, and both had a higher starvation resistance than the grassland populations, although the variation was not consistent at the *'habitat'* level unlike the *'collection site'* level.

Second-field experiment

As with body size and development time, the different species were first analysed independently because of the large variation in numbers of individuals for the different species.

The analysis of *D. cardinoides* suffered from many empty cells (see Material & Methods) and none of the factors that could be estimated was significant. However,

Table 13: F-statistics and p-values for the different main and interaction factors as estimated in the second field experiment, each based on starvation resistance residuals and for all four species separately. The use of the 'Effective Hypothesis Decomposition' model causes the reduction to zero of the degrees of freedom (see Material & Methods).

Species	*D. cardinoides*			*D. equinoxialis*			*D. melanogaster*			*D. sturtevanti*		
	DF	F	p	DF	F	p	DF	F	p	DF	F	p
Intercept	0	0	0	1	1.99	0.1591	1	2.7	0.1007	1	0.03	0.8741
Transect (Tran)	0	0	0	1	66.2	0	1	27.11	0	1	0.2	0.6518
Original habitat (Or)	1	2.09	0.1495	2	1.09	0.3379	1	8.46	0.0037	2	0.86	0.4265
Experimental habitat (Ex)	0	0	0	2	87.7	0	2	19.93	0	2	2.1	0.1253
Tran*Or	1	0.49	0.4837	2	14.6	0	1	0.98	0.3235	2	0.42	0.6592
Tran*Ex	0	0	0	2	11.14	0	2	5.02	0.0067	2	1.33	0.2676
Or*Ex	2	1.54	0.2172	4	7.42	0	2	1.95	0.1423	4	1.48	0.2114
Tran*Or*Ex	2	0.7	0.4964	4	4.72	0.0009	2	11.68	0	4	0.29	0.8855
Error	218	0	0	1247	0	0	1234	0	0	171	0	0

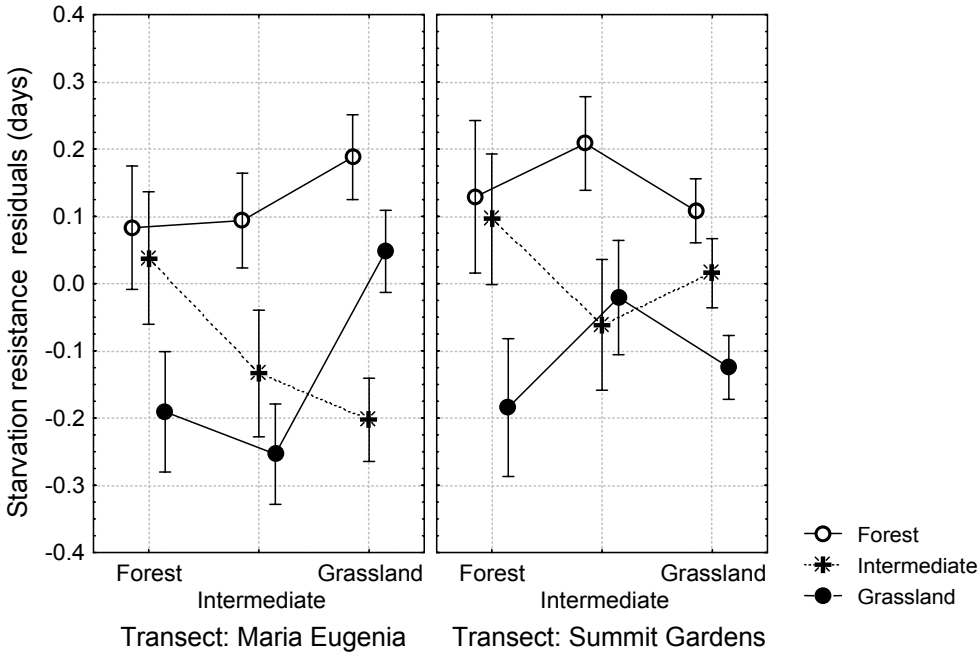


Figure 8: Transect-specific plot for starvation resistance residuals (with standard errors) with effects of experimental habitat (lines), original habitat (forest-grassland), and transect (Maria-Summit).

when the highest interaction factor was removed, *'experimental habitat'* became significant ($F_{2;220}$ = 4.11, p = 0.018). Using type III sum-of-squares gave the similar results. *D. equinoxialis* showed significant effects of all but two factors in the analysis (table 13). The number of significant results for *D. melanogaster* was lower than for the previous species (table 13). For *D. sturtevanti*, none of the factors was significant.

In the overall analysis, 92.25% of the variation was in the error component and five of the factors were significant. After the jack-knife procedure to test whether any of the species had a disproportional impact on the outcome, three factors remained. The most important factor was *'experimental habitat'* ($F_{2;2915}$ = 66.16, p < 0.0001, non-error variation explained: 54.04 %). The other two factors were *'transect'* ($F_{1;2915}$ = 6.73, p =0.01, non-error variation explained: 2.75 %) and *'transect*original*experimental habitat'* ($F_{4;2915}$ = 9.02, p < 0.0001, non-error variation explained: 14.73 %, figure 8). Two interaction factors were not significant after removal of *D. equinoxialis* which is in line with the individual species analysis in which only this species had significant effects for these two factors (*transect*original habitat*: $F_{2;2915}$ = 8.49, p = 0.0002, non-error variation explained: 6.94 % ; *original*experimental habitat*: $F_{4;2915}$ = 10.49, p < 0.0001, non-error variation explained: 17.33 %). The remaining factors were not significant. The

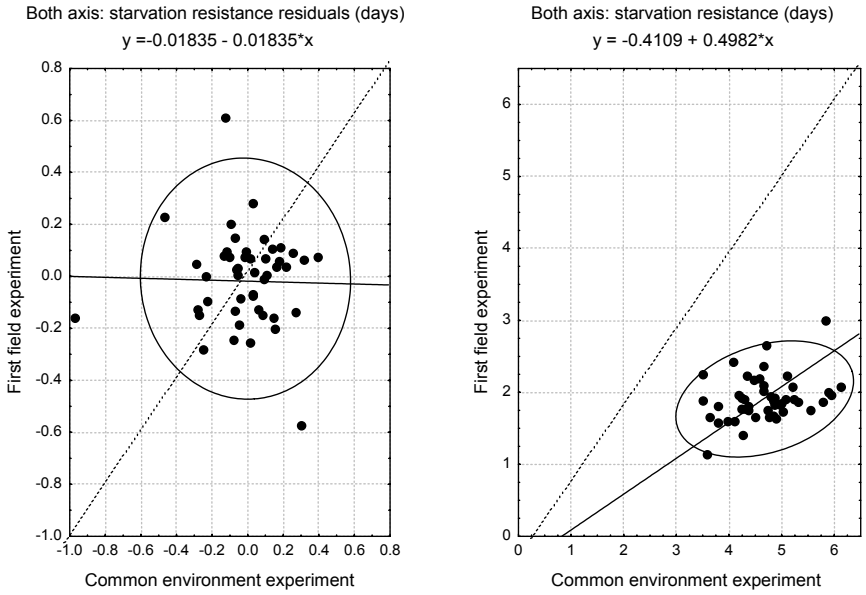


Figure 9: Common environment - first field experiment comparison for starvation resistance, full data and residuals (days). Population averages without weight factor, estimated using Reduced Major Axis regression (solid lines, see Sokal & Rohlf 1981) as both variables are estimated with error. The dotted line indicates when estimated values within both experiments would be equivalent. The residuals in the left panel are the same as in the analyses of the specific experiment. Ellipses indicate 95% confidence regions.

contrast analysis (with superscripts indicating similarity groups) showed for *'experimental habitat'*: grassland ^{a} < intermediate ^{b} < forest ^{c}.

A partitioning of the explained variation into genetic, environment and GxE fractions (table 2) showed that the fractions explained 8.70, 56.36 and 32.06 percent of the variation, respectively. A similar subdivision of the explained variation in a *'transect'*, *'habitat'* and *'collection site'* fraction showed that these fractions explained 73.13, 23.99 and 2.75 percent, respectively.

We can make the following conclusions for starvation resistance. The *'experimental habitat'* had a larger impact on the realised values than the original habitat, as it explained most of the variation in the overall analysis as well within three of the four species. Grassland populations had the lowest starvation resistance, while forest populations survived for the longest periods. The origin of a population seemed to be relatively unimportant in this experiment, but a GxE interaction was clearly present.

Table 14: P-values for the different factors as estimated in the basic and the extended analysis, for each species separately and based on body size residuals. The final column gives the number of individuals per species. The overall p-values were calculated with the Fisher-omnibus test. (df=24 in each case; log(0) replaced with -16).

	Basic analysis		Extended analysis				
	Intercept	Site	Intercept	Habitat	Transect	Site	N
D. cardinoides	p=0.6356	**p=0.0003**	p=0.3898	p=0.1779	p=0.5738	Zeroed	249
D. equinoxialis	p=0.1176	**p<0.0001**	p=0.3967	**p<0.0001**	p=0.8653	**p<0.0001**	1588
D. melanogaster	p=0.3654	**p=0.0001**	p=0.3654	**p=0.0032**	**p=0.0013**	**p=0.0057**	1082
D. malerkotliana	p=0.6750	p=0.4536	p=0.5720	p=0.2340	p=0.2662	Zeroed	136
D. nebulosa	p=0.9861	**p=0.0262**	p=0.5070	p=0.2471	p=0.4743	Zeroed	402
D. neomorpha	p=0.8743	p=0.6938	p=0.8743	p=0.6938	Zeroed	Zeroed	10
D. saltans	**p=0.0214**	**p<0.0001**	**p=0.0098**	**p=0.0285**	p=0.7161	p=0.7269	956
D. simulans	p=0.1777	**p<0.0001**	p=0.1777	**p<0.0001**	p=0.5966	**p=0.0007**	844
D. septentriosaltans	p=0.5216	**p=0.0113**	p=0.8576	p=0.0549	p=0.9379	Zeroed	710
D. sturtevanti	p=0.5328	p=0.7677	p=0.5328	p=0.6902	p=0.5364	p=0.9083	386
D. tropicalis	p=0.3927	**p<0.0001**	p=0.1630	p=0.1758	**p<0.0001**	Zeroed	1393
D. willistoni	p=0.0802	**p=0.0045**	**p=0.0095**	**p=0.0035**	**p=0.0272**	Zeroed	1270
Chi-square value	12.55	97.8	15.5	47.28	35.08	21.52	
Df	24	24	24	24	22	10	
p-value	p=0.9729	**p<0.0001**	p=0.9051	**p=0.0031**	**p=0.0380**	**p=0.0177**	

Common environment experiment

The basic analysis with only *'collection site'* as an explaining variable showed that nine out of the twelve species had clear differences between the populations (table 14) and removal of any of the species did not result in a different overall outcome. The extended analysis showed significant results for five out of twelve species for *'habitat'*, three out of eleven species for *'transect'*, and three out of five species for *'collection site'* (table 14). However, all three factors were not robust. *D. equinoxialis* determined the effect within *'habitat'*, while *D. melanogaster* and *D. tropicalis* had a disproportional impact on *'transect'* and *D. equinoxialis* and *D. simulans* on *'collection site'* .

Both the basic and the extended analysis showed that most variation was present in the error component of the analysis (basic: 99.50 %; extended: 99.43 %). The overall basic analysis showed that *'collection site'* had a significant impact on the realised starvation resistances in the laboratory ($F_{5;9040}$ = 9.0, p < 0.0001), and the jack-knife procedure showed that the outcome was robust. The *'intercept'* was not significant (table 17: $F_{1;9040}$ = 0.004, p = 0.95, non-error variation explained: 0.01 %). The contrast analysis for *'collection site'* showed the following order (with superscripts indicating similarity groups): Maria-Forest ^{a} < Maria-Intermediate ^{b} < Maria-Grassland ^{b} < Summit-Forest ^{b,c} < Summit-Intermediate ^{b,c} < Summit-Grassland ^{c}.

In the extended analysis, *'habitat'* and *'transect'* effect were significant (habitat: $F_{2;9040}$ = 10.9, p < 0.0001, non-error variation explained: 41.9 %; transect: $F_{1;9040}$ = 25.09, p < 0.0001, non-error variation explained: 48.4 %). However, the jack-knife procedure showed that the *'transect'* effect was not robust and solely attributable to *D. tropicalis* (after removal: $F_{1;7642}$ = 0.66, p = 0.41), whilst the effect of *'habitat'* was robust. The interaction between the two factors was not significant ($F_{2;9040}$ = 2.5 p = 0.08, non-error variation explained: 9.68 %) as was the *'intercept'* ($F_{1;9040}$ = 0.004, p = 0.95, non-error variation explained: 0.01 %). The contrast analysis for the interaction factor before removal of *D. tropicalis* showed exactly the same pattern as in the basic analysis, but after removal, the pattern was as follows: Maria-Forest ^{a} < Summit-Forest $^{a, b}$ < Maria-Intermediate ^{a,b,c} < Summit-Intermediate ^{a,b,c} < Summit-Grassland ^{b,c} < Maria-Grassland ^{c}. The clear impact of the different habitats was also found in the contrast analysis after removal of *D. tropicalis* for *'habitat'* alone: Forest ^{a} < Intermediate ^{a} < Grassland ^{b}. Before removal of *D. tropicalis*: Forest ^{a} < Intermediate $^{a, b}$ < Grassland ^{b}.

The overall conclusion is that populations showed clear differentiation between the different habitats. The grassland populations had the longest starvation resistances while the forest populations had the shortest.

Inter-experiment comparison

The inter-experiment comparison using the starvation resistance averages from the raw data (figure 9, right panel) showed a large difference between the two

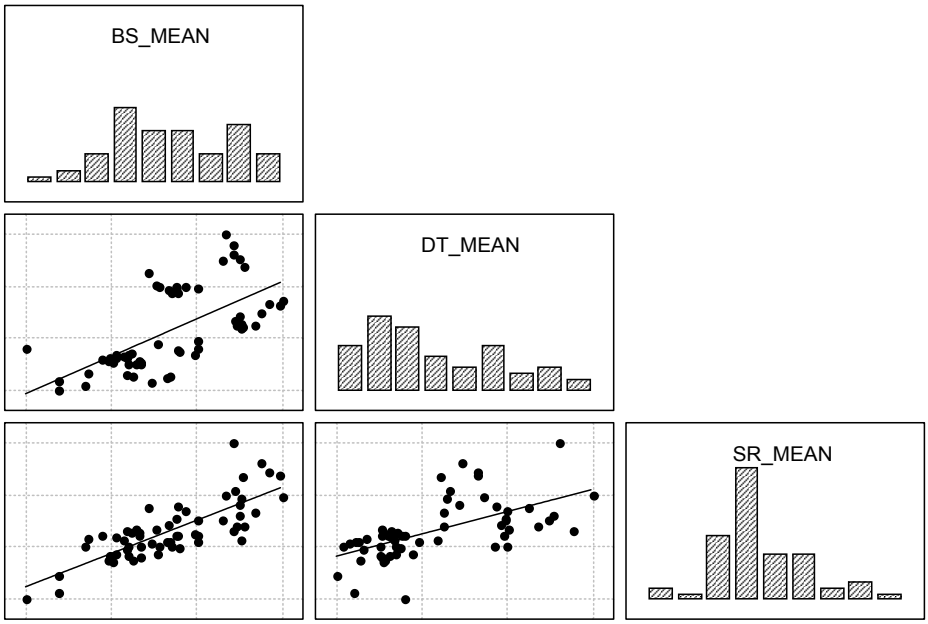


Figure 10: Correlations across species based on population averages from the first field experiment, for body size, development time and starvation resistance. Plots on the complete diagonal axis indicate distribution of the values within a trait, while off-diagonal plots are scatterplots between two traits. Data along the x-axis in each scatterplot correspond to the histogram above the plot, while the data along the y-axis data correspond to the histogram at the right of the plot. BS_MEAN: body size means; DT_MEAN: development time means; SR_MEAN: starvation resistance means.

experiments. The unweighted data showed a barely significant correlation (R^2 = 0.09, n = 48, p = 0.038). The bootstrapped 95% confidence intervals around the estimated Reduce Major Axis (RMA) regression slopes were large, but did not include the unity (x = y) line (slope: 0.4982; range: -0.4041 - 0.6852). The correlation between the two experiments using residuals showed no correlation (R^2 = 0.0004, n = 47, p = 0.89). Fitting a regression line was meaningless as the bootstrapped 95% confidence intervals were extremely large and included both x = y and x = -y lines (figure 9, left panel). The lack of fit between both the intraspecific as well as the interspecific comparison makes it impossible to extrapolate the common environment results to the field situation.

Overall conclusion for starvation resistance

The overall conclusions are straightforward. In the first-field experiment, the differences between the species were clearly related to the habitat, and the populations from the forest had the longest starvation resistances while those from the grassland had the shortest. The *'experimental habitat'* component in the second

field experiment showed the same pattern as in the first field experiment, the grassland population had the shortest, while the forest populations had the longest starvation resistance. In contrast, the *'original habitat'* component showed an opposite trend, indicating that genetically, grassland populations were more highly adapted to starvation resistance than forest populations. The common environment experiment supported this picture as the pattern was similar to the *'original habitat'* component of the second field experiment.

Table 15: Overview of starvation resistance differences between habitats, regardless of whether the differences between the habitats were significant or not.

Experiment 1	Grassland < Intermediate < Forest
Experiment 2: experimental habitat	Grassland < Intermediate < Forest
Experiment 2: original habitat	Forest < Intermediate < Grassland
Common environment	Forest < Intermediate < Grassland

The pattern as described fits neatly with the idea of countergradient variation (Conover & Schultz 1995). The grassland populations have the highest genetically determined starvation resistance, but the starvation resistance values as measured directly in the field are the lowest. This pattern fits the idea well, in that forests represent a more favourable environment for these species, while the grassland is the least favourable environment. This could then explain the increased starvation resistance in the grassland. However, the presumed adaptation is apparently not complete, as the realised starvation resistances in their own habitat are still lower for the grassland populations than for the forest populations. When this mechanism underlies the pattern in starvation resistance, and there is sufficient genetic variation, adaptation is expected to continue towards flies with even higher starvation resistances. The absence of any meaningful correlation on the residuals in the inter-experiment comparison does fit within this pattern. However, the variation between populations is large and the lack of a better fit could be attributable to GxE interactions that explained about one-third of the non-error variation in the second field experiment. The large deviation between the averages based on the raw data is discussed under 'Discussion'.

BETWEEN SPECIES CORRELATIONS

All correlations between two traits were positive and highly significant ($p<0.0001$; $N=59$; R^2 body size-development time: 0.41; R^2 body size-starvation resistance: 0.63; R^2 development time-starvation resistance: 0.36). The results of the corresponding homogeneity-of-slopes model showed that the independent trait was significant in all cases while the interaction factor was never significant. Development time versus body size: intercept: $F_{1,47} = 0.47$, $p = 0.5$; collection site: $F_{5,47} = 0.19$, $p = 0.96$; body size: $F_{1,47} = 33.1$, $p < 0.0001$; collection site * body size: $F_{5,47} = 0.17$, $p = 0.97$. Starvation resistance versus body size: intercept: $F_{1,47} = 4.24$, $p = 0.045$; collection site: $F_{5,47} = 2$, $p = 0.096$; body size: $F_{1,47} = 100.85$, $p < 0.0001$; collection site * body size: $F_{5,47} = 2.29$, $p = 0.061$. Starvation resistance versus development time: intercept: $F_{1,47} = 16.92$, $p = 0.0002$; collection site: $F_{5,47} = 0.37$, p

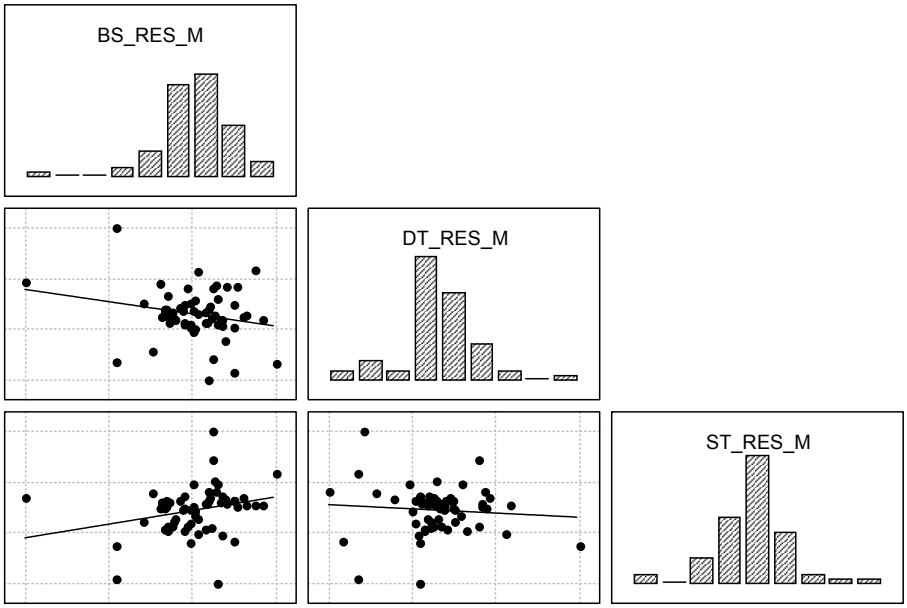


Figure 11: Correlations across species based on population averages from the first field experiment, for residual of body size, development time and starvation resistance. Plots on the complete diagonal axis indicate distribution of the values within a trait, while off-diagonal plots are scatterplots between two traits. Data along the x-axis in each scatterplot correspond to the histogram above the plot, while the data along the y-axis data correspond to the histogram at the right of the plot. Means are based on residuals. BS_RES_M: body size means; DT_RES_M: development time means; SR_RES_M: starvation resistance means.

= 0.87; development time: $F_{1,47}$ = 27.88, p < 0.0001; collection site * development time: $F_{5,47}$ = 0.56, p = 0.7298. The analysis for the common environment data gave a similar overall picture. This homogeneity in the slopes of the regressions within the different collection sites leads to the conclusion that the pattern from the correlations is robust and was little changed by differences between the habitats.

WITHIN SPECIES CORRELATIONS

The results of the homogeneity-of-slopes model were very inconsistent between the first-field experiment and the common environment experiment (table 21). However, the number of data points within each collection site is limited and that could obscure any underlying pattern. The correlations between the traits using the whole data set showed that there is a significant positive correlation between body size and starvation resistance, while the other two correlations are negative and non-significant (body size - starvation resistance: R^2 = 0.071, n = 59, p = 0.041; body size - development time: R^2 = 0.052, n = 59, p = 0.081; development time - starvation resistance: R^2 = 0.0074, n = 59, p = 0.52). This suggests that the link

between body size and starvation resistance is stronger than for the other combinations, but also that all relations between traits are weak and variable.

Table 16: Comparison between the first-field experiment and common environment experiment. Bold cells indicate significant effects.

Dependent variable	Factor	Common environment	1st field experiment
Development time	Intercept	$F_{1,37} = 0.31$, $p = 0.5789$	$F_{1,47} = 0.03$, $p = 0.8528$
	Site	$F_{5,37} = 1.34$, $p = 0.2696$	$F_{5,47} = 3.41$, $p = 0.0105$
	Body size	$F_{1,37} = 2.12$, $p = 0.1538$	$F_{1,47} = 4.39$, $p = 0.0416$
	Site*Body size	$F_{5,37} = 1.93$, $p = 0.1121$	$F_{5,47} = 1.64$, $p = 0.1689$
Starvation resistance	Intercept	$F_{1,37} = 0.30$, $p = 0.5842$	$F_{1,47} = 0.03$, $p = 0.8635$
	Site	$F_{5,37} = 0.86$, $p = 0.5167$	$F_{5,47} = 0.93$, $p = 0.4680$
	Body size	$F_{1,37} = 0.04$, $p = 0.8398$	$F_{1,47} = 4.06$, $p = 0.0497$
	Site*Body size	$F_{5,37} = 4.38$, $p = 0.0031$	$F_{5,47} = 1.21$, $p = 0.3213$
Starvation resistance	Intercept	$F_{1,37} = 0.01$, $p = 0.9346$	$F_{1,47} = 0.00$, $p = 0.9823$
	Site	$F_{5,37} = 1.38$, $p = 0.2559$	$F_{5,47} = 2.02$, $p = 0.0936$
	Development time	$F_{1,37} = 8.80$, $p = 0.0052$	$F_{1,47} = 2.53$, $p = 0.1184$
	Site*Development time	$F_{5,37} = 6.91$, $p = 0.0001$	$F_{5,47} = 2.04$, $p = 0.0898$

Discussion

BODY SIZE

The analysis for body size revealed habitat-related phenotypic and/or genotypic variation between populations within most species, but this variation was not consistent over all species at the overall level. At the collection site level, flies from populations collected in the Summit-Grassland and Maria-Forest *'sites'* were generally significantly smaller than individuals from the other *'collection sites'*. However, at a phenotypic level, the variation between the collection sites was inconsistent and varied with the inclusion or exclusion of a particular single species.

The genetic variation (as measured in the common environment experiment) was significantly correlated with the phenotypic variation (as measured in the first-field experiment). However, the correlations did not explain all the variation. The second-field experiment showed that GxE interactions were important in explaining the variation between populations. The common environment as I used in the laboratory roughly matched the natural environment, but large differences remained. For example, the temperature in the laboratory was 25 ºC, which was close to the average temperature in the field, but the daily temperature fluctuations were much greater in the field compared to the laboratory. The sensitivity of the flies for changes in the natural environment suggest that changes in the environment when the flies are transferred from the field to the laboratory could be of great importance and lead to adaptation to the novel laboratory environment (Matos *et al.* 2000b, Matos *et al.* 2002, Schlichting & Smith 2002, Service & Rose 1985).

DEVELOPMENT TIME

Development time showed clear habitat-related variation in the common environment experiment. Grassland individuals had a shorter development time than individuals from the intermediate habitat, which in turn had a shorter development time than the forest individuals. The two transects also differed significantly from each other: individuals from the Maria transect had shorter development times than those from the Summit transect. The first field experiment data showed a similar pattern for the habitats, but the result for the transects was opposite, with Summit individuals having the shorter development times. The same applies to the *'experimental habitat'* factor in the transplantation (second field) experiment, while the order in the *'original habitat'* factor was different, placing the forest habitat between the two others. This concordance, although with exceptions, between the experiments suggests that the genotypic and phenotypic variation was determined by the same single underlying cause, and that this cause was related to the habitats as they had the same order within the two transects.

What could this selective factor be? Temperature is an unlikely candidate. The average temperatures in the grasslands were higher than in the forest and higher environmental temperatures are associated with shorter development times (Azevedo *et al.* 1996, James *et al.* 1997, Zwaan *et al.* 1992). This was indeed observed for the *'experimental habitat'* component of the second field experiment, while the results of the first field experiment were roughly in line with the expected pattern also. However, the common environment experiment and the *'original habitat'* component of the second field experiment, were expected to show the opposite pattern comparable with the low temperature selection lines. These selection lines are comparable to the forest habitat and have a short development time in comparison to high temperature lines (Anderson 1966, James & Partridge 1995, Partridge *et al.* 1994a, b). This is in sharp contrast with the data, in which the forest individuals have the longer development times. The difference in average temperatures in my experiment is limited to one degree Celsius. This difference is much smaller than the difference in the temperatures used in the selection experiments. However, this difference, whilst expected to give less dramatic results, is not likely to result in opposite outcomes.

Relative humidity varied also with habitat and was lowest in the grassland with the lowest humidity around midday. However, there are no published results on the effect of relative humidity for development time for *Drosophila* and results for other species than *Drosophila* contradict each other (Krasnov *et al.* 2001, Smith 1993). The design of the experiment was such that effects of desiccation on the larvae were unlikely to occur, as the pieces of banana were located on a layer of moist vermiculite that was kept moist. However, genetic differentiation due to variation in relative humidity among the habitats can not be excluded.

Krijger (2000) found in his study on *Drosophila* species in Panama that mean resource abundance increased with disturbance of the habitat. This was consistent with the expectation based on the life-history model of Sevenster & van Alphen

(1993a, 1993b) that an increase in mean resource abundance would lead to a decrease in mean community development time. Furthermore, it was expected that this change would be accomplished by a relative change in the community composition, e.g. the replacement of slow species by fast species. Contrary to the expectations, Krijger (2000) did not find this negative correlation between resource abundance and mean community development time. However, the calculations of the mean community development time were based on a single estimate for the species-specific development times, regardless of the habitat.

In the present study, grassland individuals have the shortest development times while forest individuals have the longest. Based on the results of Krijger (2000) for the average resource abundance in relation to disturbance, forest habitats had the lowest mean resource abundance while the intermediate habitats had a higher mean resource abundance. No data on the mean resource abundance of the grassland habitats were available. His data were obtained in the same area as my own. Moreover, the pattern found in my study fits the prediction based on the life-history model of Sevenster & van Alphen (1993a, 1993b) in which mean resource abundance is negatively correlated with mean community development time. The main difference is that the variation in mean-community development time was not achieved by the replacement of the slow species by fast species, but by a community wide adaptation to the changed environment.

STARVATION RESISTANCE

Starvation resistance shows high levels of phenotypic plasticity. The transplantation experiment showed that flies from the same population, but reared in different habitats realise a higher starvation resistance in the forest habitat compared to the grassland habitat. This difference in expression suggests that the grassland environment is harsher than the forest environment. The same experiment also showed that the *'experimental habitat'* factors, i.e. the environmental component, were more important in explaining the observed pattern than were the *'original habitat'* factors, i.e. the genetic component. This dominance of the environment over the genetics was reflected in the pattern in the first field experiment, which was similar to the *'experimental habitat'* related factors of the second field experiment. Furthermore, the pattern in the common environment experiment was similar to the *'original habitat'* related factors.

Thus, the patterns within the different experiments point towards an overall picture in which the environment becomes increasingly harsh when it is degraded from primary forest to grassland. Such a trend would then suggest a need for the grassland populations to adapt to the changed environment, which has indeed happened. That the realised starvation resistances remains lower in the grassland than in the forest, indicates that the adaptation is incomplete and, if selectable genetic variation remains (Blows & Hoffmann 1993, Hoffmann *et al.* 2003a, Roff 2003), the populations would be expected to evolve further and become even better adapted to the changed environment.

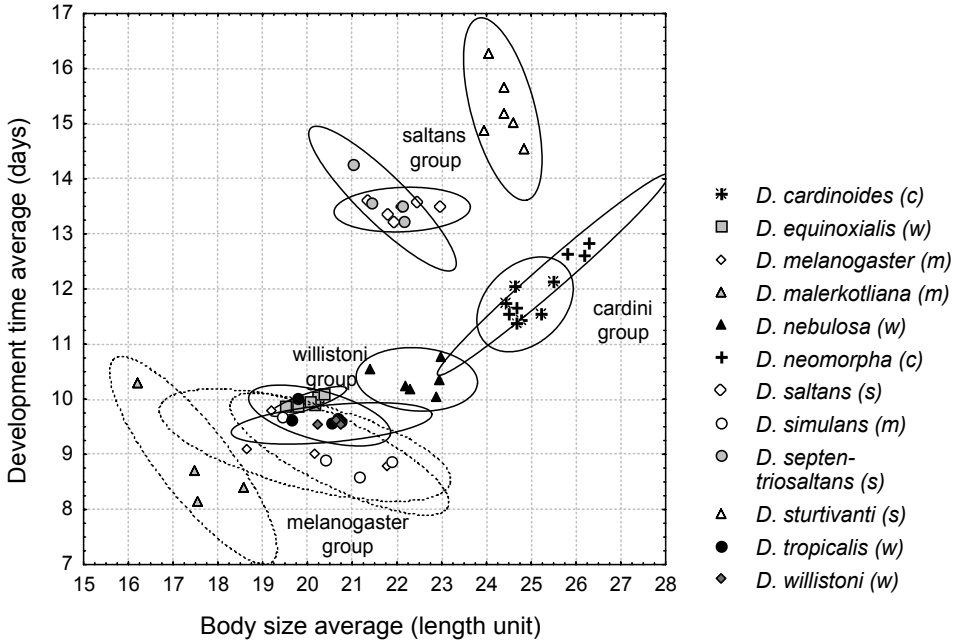


Figure 12: Body size versus development time based on population averages for males and females combined. Ellipses indicate the 95% range for the different species. Line patterns indicate phylogenetic relatedness at the level of species groups. Similar points are only of the same species when they are within the same ellipse.

The differences in the realised starvation values between the first field experiment and the common environment experiment are remarkable (figure 9, right panel). Survival times under desiccation stress are generally much shorter than under starvation stress, and depending on the species, vary from just a few hours for the smaller species like *D. bipectinata* up to 48 hours for the larger species like *D. repleta* (Parkash & Munjal 1999). Estimates for *D. melanogaster* vary between nine (Hoffmann *et al.* 2001b, Hoffmann *et al.* 2001a) and 24 hours (Parkash & Munjal 1999), with most estimates not exceeding 15 hours. The data of the first field experiment on *D. melanogaster* showed an average survival time of 43.7 hours (range 39 - 48 hours) for this species. This is much lower than some laboratory measured starvation resistances on freshly established stocks (105 - 130 hours (Parkash & Munjal 1999)), but within the range reported by other authors (40-80 hours (Hoffmann *et al.* 2001a)). E. Baldal (in preparation) observed in his base line that the variation between generations covered the whole range of reported starvation resistances. This observation underlines the sensitivity of this trait to environmental variation. Furthermore, if repeated measurements of a single stock under constant conditions already result in such a variable outcome, measurements obtained in different environments are likely to be even more variable. This was confirmed in the comparison of the first field experiment with the common environment experiment (figure 9, right panel).

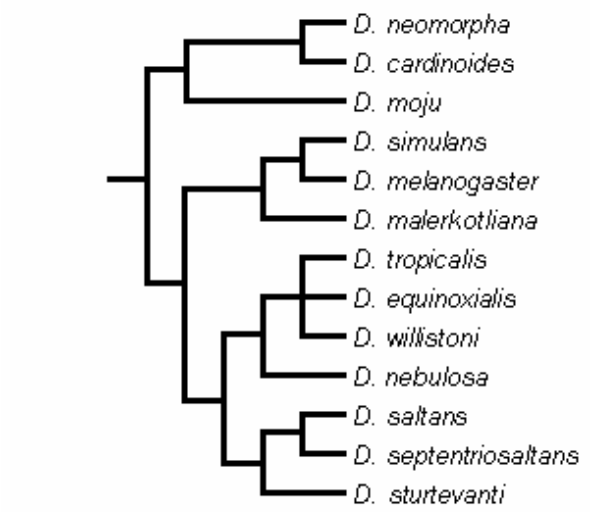
Another reason for why desiccation is unlikely to explain the differences between the experiments, is that the flies were provided with water in the form of water-agar that remained moist for many days and was never visibly dry by the time the last fly in the cohort died. Furthermore, if desiccation had played a role, it would have been likely to affect the grassland populations more severely than the forest populations since the humidity in the forest was always very high and often near saturation. To examine this, I divided the estimate of the first field experiment by the estimate of the common environment experiment and tested whether the ratios differed between the three different habitats. This showed that the ratios did not differ between habitats ($F_{2, 44} = 1.92$, $p = 0.16$), and therefore it is unlikely that desiccation explains the differences between the two experiments.

The differences in average temperature between the laboratory and the field are minimal, but the daily temperature variation in the field is up to 7 ºC, much higher than in the laboratory. Higher temperatures reduce starvation resistances (Da Lage *et al.* 1989, Karan & David 2000), which is thought to be related to an increased metabolism. High temperatures can also induce protection mechanisms (Hoffmann *et al.* 2003c), but starvation resistance might not be increased by this mechanism (Minois 2001) although non-induced flies (e.g. without prior heat-shock treatment) had a longer starvation time than induced flies. Based on the available literature on desiccation resistance, metabolic rates and heat-induced protection mechanisms, no interpretations about the causes of the difference between the experiments can be made.

The life-history model of Sevenster & van Alphen (1993a, 1993b) is based on an ecological trade-off between development time and starvation resistance. Individuals with a long starvation resistance have a better chance of finding a new patch, favouring a higher starvation resistance. The field data of Krijger (2000) showed that the forests with the highest temporal heterogeneity indeed have the lowest mean resource abundance. This favours slow species with long development times and correspondingly longer starvation times. The observed phenotypic pattern for starvation times is consistent with this prediction. However, the genetic pattern is opposite to the expectations. Apparently, the abiotic selection pressure is more important in shaping starvation resistance.

PHYLOGENETIC DEPENDENCE

A component of phylogenetic history is clearly visible within the data (see figure 12 for an example of body size versus development time; members within species groups tend to resemble one another). (Pagel 1999a, b) developed a method for estimating the phylogenetic dependence within such data. The estimated λ ranges between 0 (phylogenetic independence) and 1 (species' traits co-vary in direct proportion to their shared evolutionary history), and it is possible to estimate the phylogenetic dependence for several traits together. For this analysis, I used the data from the first field experiment, for all three traits, all species and all populations. The populations within a species were considered to originate from the same node, so these first nodes correspond with the different species. The higher



species in the study reported here (Bock 1980, Rodriguez-Trelles *et al.* 2000, Val *et al.* 1981, Vilela 1983).

order nodes were based on the phylogenetic classification of Bock (1980), Rodriguez-Trelles *et al.* (2000), Val *et al.* (1981) and Vilela (1983).

This analysis showed that phylogenetic history explained the pattern within the three traits completely ($\lambda = 1$; 95% confidence interval: $0.856 < \lambda <$ larger than one (not estimatable as λ is between 0 and 1)). A similar analysis for each trait separately showed that $\lambda = 1$ for body size (95% confidence interval: $0.837 < \lambda <$ larger than one) and development time (95% confidence interval: $0.833 < \lambda <$ larger than one), and $\lambda = 0.891$ for starvation resistance (95% confidence interval: $0.497 < \lambda <$ larger than one). The clear relation between the phylogenetic history and the interspecific variation shows that a part of the underlying genetic architecture is fundamental. This footprint of the past is neither easily changed, nor related to the current day local adaptation as observed. The λ for starvation resistance is the lowest, which is noteworthy since selection in the field is most obvious for this trait.

INTRASPECIFIC AND INTERSPECIFIC CORRELATIONS

All three interspecific correlations between two traits were positive, and the principal component analysis (data not shown) showed that variation among all three traits could be reduced to a single significant principal component explaining about 75% of the variation among the species. This reduction to one principal component could either indicate that one main cause underlies much of the interspecific variation in these three traits, or that selection on multiple underlying mechanisms has resulted in a consistent simultaneous selection of the traits. In the phylogenetic analysis, as presented above, variation in body size and development time matched perfectly the phylogenetic history of the group indicating that the linkage at the phenotypic level is common, and of ancient origin. Analysis of molecular evolution data sets frequently splits the major groups within the *Drosophila* genus at between 50 and 100 million years ago (Beverley & Wilson 1984). This underlines that the tight linkage between the traits among species is embedded strongly within the *Drosophila* genus. The most likely explanation for my results is that a single set of highly conserved genes and genetic pathways are primarily responsible for the co-variation of all three traits. The alternative explanation, that different selection

pressures independently targeted different traits, is less likely, as that would require the co-occurrence of those selection pressures over at least several millions of years.

For the interspecific correlations, it is likely that underlying genetic correlations were producing the phenotypic correlation, as the phylogenetic history is reflected in all three traits and in the principal component factor. For the intraspecific correlations, the use of phenotypic correlations as a surrogate for genetic correlation is still debated, but review studies on morphological and life-history traits show that for most estimates of two morphological traits, or a morphological and a life-history trait, the sign and magnitude of the phenotypic correlations were similar to the genetic correlations (Cheverud 1988, 1995, Roff 1995, 1996, 1997, 2000). However, exceptions have been reported in which the estimates for the phenotypic and genetic correlations differed in sign (Roff & Mousseau 1987) or magnitude (Hebert *et al.* 1994).

Both the literature and the results presented here were not conclusive about the sign of the different genetic or phenotypic correlations. At the interspecific level, the three correlations were positive, similar to the results from the selection experiments reported in the literature (see **chapter 3** for a literature overview). In contrast, at the intraspecific level, only the correlation between body size and starvation resistance is positive, the other two were negative. This is for the correlation between body size and development time in line with the findings in studies of latitudinal clines, but inconsistent with most selection experiments (Cortese *et al.* 2002, Gu & Barker 1995, Nunney 1996b, Partridge & Fowler 1993, Partridge *et al.* 1999, Reeve 1954, Robertson 1957, 1960a, b, 1963, Roper *et al.* 1996, Santos *et al.* 1992, 1994, Zwaan *et al.* 1995a). The limited published results for the other two correlations were not consistent (see **chapter 3**).

INTER-EXPERIMENT COMPARISONS

The inter-experiment comparisons at the intraspecific level for development time and body size showed a close fit between the common environment experiment and the first field experiment. The differences between the larger and smaller species have increased between the two experiments, while the development times tended to become a little shorter. In contrast, the correspondence between the two experiments for starvation resistance was poor (see discussion on this under starvation resistance). At the intraspecific level, only body size showed a significant correlation between the two experiments. To the contrary, the comparisons for development time and starvation resistance showed no fit between the two experiments.

Potentially, several sources can contribute to this variation between experiments. The genetic variation did not change, but reports of rapid laboratory adaptation suggest that the populations could have changed between the two experiments during the months the stocks were maintained in the open-air laboratory (Hoffmann *et al.* 2001b, Matos *et al.* 2000a, Matos *et al.* 2000b, Matos *et al.* 2002, Partridge *et*

al. 1995, Sgro & Partridge 2000). Furthermore, environmental differences are another potential source for variation. Some aspects of the environment might have changed in a consistent manner, but most of them must have changed to a different degree for populations from different collection sites, as the common environment was the same for all populations. These differences in direction and the extent of the changes are potentially magnified if Genotype-by-Environment interactions exist. The final source of variation is the random variation always present in experiments.

When the data from two experiments, carried out under different environmental conditions, yield closely similar interpretations (i.e. body size), it suggests that the underlying genetics are dominating. It is also an indication that rapid laboratory adaptation is absent. In contrast, a complete lack of fit in such a comparison (i.e. for starvation resistance), underlines that the contribution of the underlying genetics to the realised phenotypes is only small, or that rapid laboratory adaptation has taken place. Based on the inter-experiment comparisons in this chapter, I conclude that the extrapolations of results obtained in a different environment are at least to be interpreted with caution, especially for development time and starvation resistance.

GENOTYPE-BY-ENVIRONMENT INTERACTIONS

Genotype-by-Environment (GxE) interactions arise when different genotypes respond in different ways to variation between environments. In this study, the existence of GxE interactions at the level of populations from different locations was tested using the natural variation between different habitats in the transplantation experiment. The results of the experiment showed that GxE interactions exist at the population level for all three traits and that a part of the GxE interaction variation was consistent over the four species in the experiment. For body size, the GxE interaction component explained 31.4% of the variation explained by genetic, environment, and GxE interactions, while this was 39.4% and 24.5% for development time and starvation resistance, respectively. The consistency of the GxE interaction over the four species may indicate that selection favours similar patterns of GxE interactions across the different species. Furthermore, it showed that GxE interactions are likely to be ubiquitous for those types of key life-history traits in natural populations.

FIELD VERSUS LABORATORY

My primary aim of this study was to measure life-history traits directly in the field to test the extent to which laboratory-based *Drosophila* life-history theory applies to natural conditions. The results presented in this chapter show that measuring life-history traits directly in the field is possible and that it gives additional insight about life-history evolution.

This chapter shows clearly that extrapolating the results obtained in a common environment towards the field situation is not easy. Comparisons across experiments often showed little correspondence, and genotype-by-environment

interactions often explained more of the variation present than the genetic component. Despite this, the patterns within the common environment matched those within the *'original habitat'* component of the transplantation experiment indicating that an accurate prediction of the field pattern is possible based on the common environment experiment.

CONCLUSIONS

This study is the first to measure at the same time the expression of three different life history traits directly in the field. The wealth of information from this approach provides insights into the evolution of the life-history traits in the field. The comparison of the results of the three experiments revealed that the variation within the three traits and the correlations between the traits show different patterns. Both the reported variation between laboratory and field studies and my comparative results stress the ubiquity of GxE interactions.

Starvation resistance shows a pattern in which the adaptation to an environmental stress is not yet completed. Populations from the grassland (high stress) have the shortest starvation times but are genetically more resistant to that same stress. For development time, this direct response to an environmental stress is less clear as the genetic patterns were opposite to those expected pattern based on temperature selection in the laboratory. However, the pattern is consistent with the expectations from the life-history model of Sevenster & van Alphen (1993a, 1993b). Body size seem to be relatively unaffected by the differences among the habitats, or it is less consistently affected than are the other traits.

The comparison between the three traits showed that the interspecific covariance between the three traits was high. At the interspecific level, all correlations between any two traits were positive and the variation between the species shows a clear impact of the phylogenetic history. At the intraspecific level, only the correlation between body size and starvation resistance is positive, the two other correlations of starvation resistance with development time and with body size were both negative. This result contrasts with those found in selection experiments but matches in part the results from other studies such as on latitudinal clines.

The presence of considerable genotype-by-environment interactions at the population level, which is similar across the different species, may indicate that selection favours similar patterns of GxE interactions across the different species. The GxE interactions (for all three traits) and the lack-of-fit between the field experiment and the common environment experiment (for development time and starvation resistance), make the extrapolation of laboratory results to the field challenging. The integration of laboratory work with field-based experiments clearly has an important contribution to make, over and above that of more traditional, laboratory-based studies.