The genetic basis of eyespot size in the butterfly *Bicyclus anynana*: an analysis of line crosses

PIETER J. WIJNGAARDEN* & PAUL M. BRAKEFIELD

Institute of Evolutionary and Ecological Sciences, Leiden University, PO Box 9516, NL-2300 RA Leiden, the Netherlands

The tropical butterfly *Bicyclus anynana* shows wide variation in the size of the eyespots on the ventral side of its wings. Differences in the environmental temperature during late larval and early pupal stages are a major source of this variation, but variation also exists within temperatures. Using lines selected at a single temperature for large and small eyespots, and a number of crosses derived from these lines, we studied the genetic basis of eyespot size in *B. anynana*. We applied Lande's modification of the Castle–Wright (C–W) estimator to estimate the minimum number of genes contributing to the difference between the two lines. Estimates indicated that at least five genes are involved. As the C–W estimator is based on a number of simplifying assumptions, we tested to what extent one of these assumptions (additive gene action) is actually met. Joint-scaling tests indicated that the assumption of additivity is not satisfied and that dominance and probably epistasis play a role. Because reciprocal crosses were available we looked for evidence for sex-linkage and cytoplasmic effects. No evidence for cytoplasmic effects was found, but the data were consistent with the presence of one or more loci on the X-chromosome. The results are discussed in the context of the current model of eyespot formation.

Keywords: *Bicyclus anynana*, butterfly eyespots, Castle–Wright estimator, cytoplasmic effects, sex-linkage.

Introduction

Butterfly wing patterns in general, and butterfly eyespots in particular, have proven very useful objects in the study of the genetic and developmental bases of phenotypic evolution. Although spectacularly diverse, butterfly wing patterns nonetheless possess an underlying simplicity: the nymphalid groundplan (Nijhout, 1991). This groundplan — the collection of homologies among pattern elements — has been modified in numerous ways to generate patterns that are involved in camouflage, deflection of predators, mimicry, and communication between the sexes. Experimental investigation of wing patterns is greatly facilitated by their structural simplicity: the colour patterns are mosaics of tiny, differently pigmented scales that cover the wings like tiles on a roof. Nijhout (1980) conducted a series of transplant and cautery experiments in early pupae of

Precis coenia that resulted in his formulation of the morphogen-sink model. In this model, cells at the centre (or focus) of the future adult pattern element (e.g. an eyespot) produce a chemical (the morphogen) that diffuses into surrounding cells. These surrounding cells respond to different concentrations of the morphogen by producing differently coloured scales (see French & Brakefield (1992) for an alternative model, in which the focus acts as a sink rather than a source).

Evolutionary change of wing patterns requires a genetic basis of pattern variation, which can be examined using spontaneous or induced mutations, or in artificial selection experiments. These approaches have been used extensively in the tropical butterfly *Bicyclus anynana* [see Brakefield (1998) and Brakefield & French (1999) for summaries]. Selection experiments have uncovered considerable amounts of genetic variation in various characteristics of eyespots in *Bicyclus*. Not all variation in this species has a genetic basis, however. Successive generations of *Bicyclus* show distinct phenotypes in response to wet–dry seasonality. In the wet season individuals are conspicuous because of the large

*Correspondence and Present address: Laboratory of Genetics, Wageningen University, Dreijenlaan 2, NL-6703 HA Wageningen, the Netherlands; E-mail: pieter.wijngaarden@genetics.dpw.wau.nl

eyespots on the ventral side of their wings, whereas in the dry season individuals are cryptic because their eyespots are greatly reduced in size. In the laboratory, rearing larvae at a high (27°C) or low (17°C) temperature results in the wet season form or dry season form, respectively. Intermediate temperatures yield intermediate phenotypes, indicative of a continuous reaction norm. This polyphenism is thought to have evolved as a balance between selection for crypsis in the cool, dry season and selection for large eyespots (that can act as deflection devices) in the warm, wet season. Results from field experiments support this hypothesis (N. Reitsma, G. Engelhard, and P. Brakefield, unpubl. data).

Here we report on a study of the genetic basis of eyespot size in *Bicyclus anynana*. A pair of lines was selected for large eyespots (the **HIGH** line) and for small eyespots (the **LOW** line) on the ventral side of the hindwings. These lines, and a number of derived line crosses, have been used by Brakefield *et al.* (1996) to produce initial estimates of the minimum number of loci contributing to the difference in eyespot size applying Lande's (1981) modification of the Castle–Wright (C–W) estimator (Castle, 1921; Wright, 1968). In this article we will provide additional (and more detailed) information to that reported by Brakefield *et al.* (1996). Because the C–W estimator assumes equal, additive allelic effects we examined whether an additive model actually suffices to explain the data. Finally, because reciprocal crosses were available for all line crosses it was also possible to look for both maternal effects and effects of loci on sex chromosomes.

Materials and methods

The study species and selection procedure

The stock population was established from more than 80 gravid females collected in 1988 near Nkhata Bay in Malawi. It is maintained at an adult population size of 600–800 with some overlap of generations. Larvae feed on young maize plants; adult butterflies are given mashed banana.

The selection lines were established by truncation selection on the size of the largest (fifth) eyespot on the ventral hind wing. At least 40 females were used as selected parents in each line after being held in a mating cage with about 100 males with extreme large or small eyespots (under these conditions estimates of N_e will be at least 50; see Saccheri *et al.*, 1999). Selection for the first 10 generations was applied to butterflies reared at 20°C. The lines responded readily to selection due to the presence of considerable amounts of additive genetic variance (similar selection experiments yielded realized

heritabilities of about 0.4). In order to increase phenotypic variation, selection was then continued by rearing the **LOW** line at 23°C for 10 generations and the **HIGH** line at 18°C for six generations. In each generation 300–400 individuals of each sex were measured with high repeatability using an image analysis system.

Line crosses and measurements

All crosses were made at the end of selection with about 50 individuals of each sex. The rearing was carried out over two generations because it was logistically impossible to rear all lines/line crosses at the same time. Each line/line cross was reared in a separate cage at 23°C, 70% relative humidity, and under a 12 h:12 h light:dark cycle. 23°C is an intermediate temperature with respect to both seasonal forms so that temperature-effects on the lines/line crosses are minimized.

Eyespots in *Bicyclus* consist of a white centre, a black disc, and a gold outer ring. Eyespot size was measured as the diameter of the black disc of the fifth eyespot on the ventral hind wing. In the **LOW** line it sometimes occurred that the black ring was absent; in these cases the diameter of the white centre of the eyespot was measured. Forewing length was taken as a measure of overall body size. Measurements were made with high repeatability using a binocular microscope fitted with a micrometer.

The Shapiro–Wilk test and Levene's test were used to assess departures from normality and homogeneity of variances, respectively; Tukey's studentized range test was used to find out which line crosses differed from each other. The line crosses will be considered separately in the analyses of cytoplasmic effects and sex linkage, but they will be pooled in the analyses of the number of genes and of the composite effects.

When referring to line crosses the female parent will be given first. For example, the F_1 cross 'H × L' refers to the progeny of a cross between **HIGH** line females and **LOW** line males. Their offspring (= F_2) is referred to as 'HL × HL'. 'HL × H' is an example of a backcross, in this case between H × L (F_1) females and **HIGH** line males.

The Castle–Wright estimator

The C–W method assumes that loci are unlinked and that alleles have equal, additive effects both within and among loci. Furthermore, all alleles that increase the magnitude of the focal trait should be fixed in one parental (**HIGH**) line while all alleles that decrease the value of that trait should be fixed in the other parental (**LOW**) line. Violations of one or more of these assumptions will generally result in underestimating the number

of effective factors n_E (Cockerham, 1986; Zeng *et al.*, 1990; Zeng, 1992).

A consequence of linkage is that estimates of n_E should be smaller than the haploid number of chromosomes plus the mean number of recombination events per gamete. Assuming one or two recombination events per chromosome, estimates of n_E for *B. anynana* ($n=13$) cannot exceed 26–39. The claim by Turner & Sheppard (1975) that crossing over is absent in females of two *Heliconius* species (and some other butterfly species) is interesting in this context. As we have no data on recombination frequencies in *Bicyclus* we cannot use the modified C–W estimators given by Zeng (1992) and Zeng *et al.* (1990) that take linkage into account.

When the assumptions hold the number of genes responsible for the difference between the parental lines is given by:

$$n_E = \frac{(\mu_1 - \mu_2)^2}{8\sigma_s^2},$$

where μ_i is the mean of the parental line i and σ_s^2 is the segregational variance (the excess variance that appears in the F_2). A biased estimator of n_E is

$$\hat{n}_E = \frac{(\bar{x}_1 - \bar{x}_2)^2 + \frac{S_{\bar{x}_1}^2}{n_1} + \frac{S_{\bar{x}_2}^2}{n_2}}{8S_s^2},$$

where the extra terms in the numerator are the sampling variances of the means of the parental lines.

When data from the parental lines, the F_1 , the F_2 , and the backcrosses B_1 and B_2 are available four ways of estimating σ_s^2 are possible (see formulas 4a–d in Lande, 1981). Corresponding to the four ways of estimating there are four ways of estimating $\text{Var}[\sigma_s^2]$ (formulas 7 and 8a–d in Lande, 1981). Instead of these four interrelated estimates one can also combine all the information into one estimate (Cockerham, 1986). Because these sampling variances require that the trait values are normally distributed, empirical standard errors were obtained by a bootstrap procedure (Efron & Tibshirani, 1993; Manly, 1997) in which 1000 values of $\hat{n}_E$ were calculated by sampling with replacement from the pooled raw data.

Estimation of composite effects

The C–W model depends critically on the assumption of additive allelic effects. The availability of line crosses enables one to obtain an impression of the relative contributions of additive, dominance, and epistatic effects to the differentiation of the two parental lines.

Here only models with additive and additive + dominance effects will be considered. A model that includes epistasis (containing six parameters) cannot be evaluated because data from only six lines/line crosses are available.

In a joint-scaling test, see Lynch & Walsh, (1998) for details, weighted least-square regression is used to estimate the parameters of an additive model (i.e. the expected mean phenotype of the F_2 μ_0 , and the composite additive effect α_i^c), or an additive + dominance model (μ_0 , α_i^c and the composite dominance effect δ_i^c). If the assumption of normality is met, then a χ^2 statistic for goodness-of-fit can be used to compare the estimates with the observed means. Starting with the simplest model, higher-order composite effects are added until the predictions are no longer significantly different from the observations. The test can be applied to variances in an analogous manner. For variances, only results for an additive model will be given because further partitioning of the segregational variance into dominance and epistatic components gives statistically questionable results (Lynch & Walsh, 1998).

Results

Individual lines and line crosses

The basic statistics for each line/line cross are given in Table 1. Most lines/line crosses show little ($0.05 < P < 0.01$) or no departures from normality; however, the LOW line (both sexes), $H \times L$ (females), $L \times H$ (females), and the stock (females) deviate markedly ($P < 0.01$ or $P < 0.001$) from normality. As the HIGH and LOW lines may be at the limits of their phenotypic expression, we might expect their phenotypic distributions to be skewed to the left and to the right, respectively. This expectation is borne out for the LOW line, but not for the HIGH line. Neither logarithmic transformations (cf. chapter 10 in Wright, 1968) nor Box–Cox transformations were able to remove these departures; in fact, these transformations made matters worse (data not shown).

The means of the reciprocal crosses of the F_1 differ in females ($F_{1,227} = 19.49$, $P < 0.0001$; a Mann–Whitney test – necessary because of non-normality – confirms this result) but not in males ($F_{1,204} = 1.56$, $P = \text{NS}$); both crosses have equal variances (Females: $F_{1,227} = 1.76$, $P = \text{NS}$; Males: $F_{1,204} = 0.59$, $P = \text{NS}$). The reciprocal crosses of the F_2 differ in their means in both females ($F_{1,283} = 4.54$, $P = 0.0341$) and males ($F_{1,300} = 15.99$, $P < 0.0001$); the variances do not differ (Females: $F_{1,283} = 0.30$, $P = \text{NS}$; Males: $F_{1,300} = 0.86$, $P = \text{NS}$). Highly significant differences occur among the means of the backcrosses to the HIGH line (Females:

Line	N	Mean	Variance	SE	Sk	Ku	P
(a)							
HIGH	137	2.03	0.0801	0.0242	-0.19	-0.31	*
Low	113	0.37	0.0194	0.0131	1.20	2.69	***
Stock	76	1.74	0.0523	0.0262	-0.27	-0.89	**
H × L	141	0.76	0.0595	0.0205	0.81	0.37	***
L × H	88	0.92	0.0780	0.0298	0.89	0.21	***
F ₁	229	0.82	0.0720	0.0177	0.86	0.49	***
HL × HL	155	1.45	0.0999	0.0254	-0.18	0.05	NS
LH × LH	130	1.53	0.1102	0.0291	-0.62	0.54	*
F ₂	285	1.49	0.1059	0.0193	-0.37	0.15	NS
HL × H	119	1.94	0.0600	0.0225	0.23	-0.19	NS
H × HL	136	1.81	0.0779	0.0239	-0.42	0.09	NS
LH × H	172	1.93	0.0549	0.0179	0.09	-0.19	NS
H × LH	154	1.79	0.0764	0.0223	-0.43	0.35	NS
B _H	581	1.87	0.0712	0.0111	-0.29	0.39	NS
HL × L	122	1.10	0.1354	0.0333	-0.09	-0.72	*
L × HL	90	1.12	0.1276	0.0377	-0.05	-0.22	NS
LH × L	112	1.17	0.0891	0.0282	-0.27	-0.16	NS
L × LH	120	1.18	0.0985	0.0287	-0.04	-0.70	NS
B _L	444	1.14	0.1128	0.0159	-0.15	-0.44	*
(b)							
HIGH	105	1.79	0.0557	0.0230	-0.05	-0.14	NS
Low	123	0.48	0.0174	0.0119	0.18	-0.78	**
Stock	75	1.53	0.0449	0.0245	-0.56	3.09	*
H × L	105	0.89	0.0371	0.0188	0.58	0.88	NS
L × H	101	0.93	0.0313	0.0176	0.05	0.02	NS
F ₁	206	0.91	0.0343	0.0129	0.33	0.40	NS
HL × HL	152	1.28	0.0758	0.0223	-0.28	-0.40	*
LH × LH	150	1.41	0.0658	0.0209	-0.33	-0.07	NS
F ₂	302	1.34	0.0743	0.0157	-0.32	-0.23	**
HL × H	154	1.60	0.0428	0.0167	-0.08	-0.41	NS
H × HL	173	1.63	0.0506	0.0171	-0.07	-0.42	NS
LH × H	140	1.69	0.0461	0.0181	0.23	0.30	NS
H × LH	140	1.67	0.0515	0.0192	0.20	0.29	NS
B _H	607	1.65	0.0486	0.0090	0.07	-0.02	NS
HL × L	155	0.95	0.0604	0.0197	0.12	0.01	NS
L × HL	118	0.95	0.0504	0.0207	0.51	0.46	NS
LH × L	110	1.00	0.0594	0.0232	-0.13	-0.25	NS
L × LH	133	1.00	0.0521	0.0198	0.31	0.21	NS
B _L	516	0.97	0.0561	0.0104	0.18	0.02	NS

NS, not significant; * $P < 0.05$, ** $P < 0.01$, and *** $P < 0.001$.

$F_{3,577} = 13.08$, $P < 0.0001$. According to Tukey's studentized range test differences exist between HL × H and H × LH, HL × H and H × HL, LH × H and H × HL, and LH × H and H × LH; Males: $F_{3,603} = 4.67$, $P = 0.0031$. Tukey's studentized range test reveals a difference between LH × H and HL × H). No significant differences are found among the variances (Females: $F_{3,577} = 2.15$, $P = \text{NS}$; Males: $F_{3,603} = 0.58$, $P = \text{NS}$). No differences occur among the means of the backcrosses to the LOW line (Females: $F_{3,440} = 1.75$, $P = \text{NS}$; Males: $F_{3,512} = 1.94$, $P = \text{NS}$); their variances

show little (Females: $F_{3,440} = 2.94$, $P = 0.0331$) or no (Males: $F_{3,512} = 0.52$, $P = \text{NS}$) heterogeneity.

In both sexes only 4 out of 15 lines/line crosses (including the stock) show a significant linear relationship between eyespot diameter and forewing length. After 'controlling' for size effects using the eyespot diameter/forewing length ratio (as in Brakefield *et al.*, 1996), a significant linear relationship occurs in 6 lines/line crosses. Size effects appear to have little consequences for estimates of n_E , (however, see below).

Table 1 Means (mm), variances, standard errors (SE), skewnesses (Sk), and kurtoses (Ku) of eyespot diameters for both the individual and the pooled lines/line crosses (see text for nomenclature). Also given is the significance of the P -value of the Shapiro–Wilk test for departures from normality. (a) Females and (b) males

Cytoplasmic effects and sex-linked genes

Differences between reciprocal F_1 , F_2 , and backcross populations might be caused either by maternal effects or by sex-linked genes, or both. Taking into account the fact that in butterflies females are the heterogametic sex and males the homogametic sex, the preceding section will now be reconsidered to see if there is evidence for maternal effects and/or sex-linked genes affecting eyespot size. The sex chromosomes will be referred to as X and Y although Z (=X) and W (=Y) are frequently used for Lepidoptera. As female butterflies are XY, effects of loci on the Y-chromosome are confounded with cytoplasmic effects; the term 'cytoplasmic effect' therefore covers both effects.

The males of the reciprocal F_1 crosses have both an X-chromosome from the HIGH line and one from the LOW line, but they differ in the origin of their cytoplasm (females differ in the origin of both their X-chromosome and their cytoplasm). Because no significant difference occurs between the F_1 s in males, a cytoplasmic effect seems unlikely. X-linkage would show up as a difference between reciprocal crosses in F_1 females and a resemblance between the females and their fathers. $H \times L$ females are indeed significantly smaller than $L \times H$ females. The magnitude of this difference (0.155 ± 0.0351 mm) accounts for 9.3% of the difference between females of the parental lines (1.659 ± 0.0181 mm).

One half of the females of both F_2 populations contain X-chromosomes from the HIGH line, the other half X-chromosomes from the LOW line. One F_2 cross has 'HIGH line cytoplasm', however, while the other F_2 consists of 'LOW line cytoplasm'. A difference was found, but the evidence is weak. The males of the two F_2 populations differ in both their cytoplasm and their X-chromosomes. Ignoring the small difference between the reciprocal crosses in females (i.e. assuming no cytoplasmic effect) the highly significant difference between the crosses (which is in the predicted direction) in males might be a result of sex-linkage. The magnitude of this difference (0.123 ± 0.0306 mm) accounts for 9.3% of the difference between males of the parental lines (1.318 ± 0.0249 mm). Note that in this latter case the 9.3% difference is actually caused by two X-chromosomes rather than one, but that only 50% of the individuals of both F_2 populations differ in two X-chromosomes.

Two of the backcrosses provide an opportunity for assessing cytoplasmic effects in females. The female offspring of the cross between $H \times L$ females and LOW line males and between $L \times H$ females and LOW line males both received their X-chromosome from their LOW line father, but they received different cytoplasm.

The means of the two crosses do not differ, however, again supporting the conclusion that cytoplasmic effects are absent. The same line of reasoning can be applied when considering the female offspring of crosses between $L \times H$ and $H \times L$ females and HIGH line males. These crosses, too, do not differ significantly.

Pooled crosses

The basic statistics are given in Table 1. When the variances of the lines/line crosses are plotted against the means and the assumptions of the C-W method are met, this should give a triangular pattern with F_1 and backcross populations at the midpoints of the edges connecting the parental and F_2 populations. Figure 1(a) shows substantial deviations from this pattern; log-transforming the data removes some dependence of the variances on the means, but still does not yield the triangle (Fig. 1b).

Fig. 1 The relationship between means and variances of eyespot size for lines/line crosses and the stock in males (—▲—) and females (—■—) for (a) raw data (mm) and (b) log-transformed data. H, HIGH line; L, LOW line; B_H , backcross to the HIGH line; B_L , backcross to the LOW line; S, unselected stock.

	Sex	1	2	3	4
I	F	10.16 (3.36)	7.65 (1.69)	12.40 (8.83)	5.53 (1.09)
	M	5.42 (0.96)	5.57 (0.97)	4.93 (1.46)	6.41 (1.33)
II	F	10.87 (4.18)	8.75 (2.39)	9.06 (4.96)	8.47 (2.73)
	M	6.01 (1.13)	6.52 (1.27)	5.24 (1.56)	8.63 (2.36)
III	F	9.19 (3.10)	7.55 (1.87)	11.38 (8.34)	5.65 (1.30)
	M	5.27 (0.94)	5.59 (1.01)	5.04 (1.57)	6.26 (1.34)
IV	F	6.57 (1.47)	6.17 (1.22)	9.92 (6.23)	4.48 (0.76)
	M	4.87 (0.80)	5.19 (0.88)	4.66 (1.36)	5.86 (1.16)
V	F	11.56 (12.45)	8.09 (2.25)	13.70 (112.77)	5.73 (1.27)
	M	5.62 (1.06)	5.77 (1.00)	5.65 (2.49)	6.59 (1.45)

F, females; M, males.

The females in the F_1 and the males in the F_2 show marked departures from normality; the females of the backcross to the LOW line differ only slightly from normality. Neither logarithmic transformations nor Box-Cox transformations were able to remove these departures (data not shown). The variances of the backcrosses are significantly different in females ($F_{1,1023} = 28.10$, $P < 0.0001$), but not in males ($F_{1,1121} = 2.83$, $P = \text{NS}$).

Estimates of n_E

Estimates of n_E are given in Table 2. Estimates based on pooled raw data indicate that at 23°C at least 6–12 loci in females and 5–6 loci in males contribute to the difference in eyespot size between the HIGH and the LOW line. Using a combined estimate of σ_s^2 (Cockerham, 1986) yields estimates (SE) of 6.51 (1.01) and 5.73 (0.51) for females and males, respectively.

Dividing the diameter of the black ring by the length of the forewing gives estimates and associated standard errors that are comparable to those based on raw data (see also table 1 in Brakefield *et al.* 1996; note that the values reported by Brakefield *et al.* are slightly different because the sampling variances of the means of the parental lines were not taken into account). Another way of removing the effect of wing size is to use the residuals of a linear regression of eyespot diameter on forewing length. Both estimates and standard errors based on these residuals are very close to the estimates based on uncorrected diameters. Box-Cox transformations (Males: $\lambda = 1.124$; Females: $\lambda = 1.236$) yield the smallest estimates.

The estimates of n_E and their standard errors obtained from the bootstrap procedure are somewhat higher than the raw data values in males. In females the standard errors of estimates 1 and 3 are considerably higher than the raw data standard errors.

Table 2 Estimates and their standard errors of the minimum number of genes n_E contributing to the difference between the HIGH line and the LOW line. Columns 1–4 correspond to the four ways of estimating n_E (see text). Data used for estimations: (I) Raw data (II) Eyespot diameter/forewing length ratio (III) Residuals of a linear regression of eyespot size on forewing length (IV) Box-Cox transformed data, and (V) Bootstrap data

So far, estimates of n_E have been obtained by computing the segregational variance as a linear function of the observed phenotypic variances within lines. However, the joint-scaling test applied to variances (see below) yields least-square estimates of σ_s^2 and its sampling variance that can be used to calculate $\hat{n}_E$ and its standard error. These estimates (SE) are 5.79 (0.84) for females and 5.15 (0.50) for males.

Estimates of composite effects

Means. Tables 3 and 4 shows the results of joint-scaling tests applied to the means of the line crosses. The additive model is clearly insufficient to explain the data in both males ($\chi_4^2 = 26.69$, $P < 0.0001$) and females ($\chi_4^2 = 57.78$, $P < 0.0001$). Including dominance still gives a poor fit (Males: $\chi_3^2 = 15.71$, $P = 0.0015$; Females: $\chi_3^2 = 33.75$, $P < 0.0001$), although the improvement is considerable [Males: $\Lambda = 10.98$, $\text{Pr}(\chi_1^2 \geq 10.98) = 0.0005$; Females: $\Lambda = 24.03$, $\text{Pr}(\chi_1^2 \geq 24.03) < 0.0001$].

Variances. The results of a joint-scaling test applied to the variances of the line crosses are given in Tables 5 and 6. Estimates (SE) of $\text{Var}(P_1)$, $\text{Var}(P_2)$, and the

Table 3 Means and their standard errors (mm) of lines/line crosses for females. (I) Observed values (II) Predicted values in an additive model, and (III) Predicted values in an additive + dominance model

		H	L	F_1	F_2	B_H	B_L
I	$\bar{z}_i$	2.0263	0.3671	0.8206	1.4856	1.8675	1.1432
	$\text{SE}(\bar{z}_i)$	0.0242	0.0131	0.0177	0.0193	0.0111	0.0159
II	$\hat{z}_i$	2.2066	0.4375	1.3220	1.3220	1.7643	0.8798
	$\text{SE}(\hat{z}_i)$	0.0123	0.0110	0.0063	0.0063	0.0085	0.0075
III	$\hat{z}_i$	2.3317	0.4883	1.2051	1.3076	1.7684	0.8476
	$\text{SE}(\hat{z}_i)$	0.0180	0.0122	0.0138	0.0065	0.0085	0.0083

Table 4 Means and their standard errors (mm) of lines/line crosses for males. (I) Observed values (II) Predicted values in an additive model, and (III) Predicted values in an additive + dominance model

		H	L	F ₁	F ₂	B _H	B _L
I	$\bar{z}_i$	1.7931	0.4754	0.9103	1.3448	1.6470	0.9729
	SE($\bar{z}_i$)	0.0230	0.0119	0.0129	0.0157	0.0090	0.0104
II	$\hat{z}_i$	1.9254	0.5151	1.2202	1.2202	1.5728	0.8677
	SE($\hat{z}_i$)	0.0105	0.0092	0.0050	0.0050	0.0070	0.0061
III	$\hat{z}_i$	2.0596	0.5777	1.1135	1.2161	1.5866	0.8456
	SE($\hat{z}_i$)	0.0155	0.0106	0.0104	0.0053	0.0071	0.0063

Table 5 Variances and their standard errors of lines/line crosses for females. (I) Observed values, and (II) Predicted values in an additive model

		H	L	F ₁	F ₂	B _H	B _L
I	v_i	0.0801	0.0194	0.0720	0.1059	0.0712	0.1128
	SE(v_i)	0.0096	0.0026	0.0067	0.0088	0.0042	0.0076
II	$\hat{v}_i$	0.0673	0.0241	0.0457	0.1051	0.0862	0.0646
	SE($\hat{v}_i$)	0.0064	0.0079	0.0034	0.0063	0.0033	0.0034

Table 6 Variances and their standard errors of lines/line crosses for males. (I) Observed values, and (II) Predicted values in an additive model

		H	L	F ₁	F ₂	B _H	B _L
I	v_i	0.0557	0.0174	0.0343	0.0743	0.0486	0.0561
	SE(v_i)	0.0054	0.0022	0.0034	0.0060	0.0028	0.0035
II	$\hat{v}_i$	0.0396	0.0203	0.0300	0.0721	0.0559	0.0462
	SE($\hat{v}_i$)	0.0029	0.0014	0.0016	0.0028	0.0016	0.0015

segregational variance $\text{Var}(S)$ are 0.0396 (0.0029), 0.203 (0.0014), and 0.0421 (0.0037), respectively, for males, and 0.0673 (0.0064), 0.0241 (0.0079), and 0.0595 (0.0084) for females. The results give strong support to the conclusion that the additive model should be rejected (Males: $\chi^2_3 = 58.86$, $P < 0.0001$; Females: $\chi^2_3 = 347.01$, $P < 0.0001$).

Discussion

Using the C–W estimator we find that at least 5–14 loci in females and 5–9 loci in males contribute to the difference in mean eyespot size between the HIGH line and the LOW line at 23°C. These numbers are equal to or smaller than the haploid number of chromosomes in *B. anynana* ($n = 13$). The estimates and their standard errors are fairly stable across the different ways of estimating n_E in males. In females, however, especially

estimates 1 and 3 and their standard errors are considerably higher than estimates 2 and 4. Given the assumptions of the C–W estimator the figures in Table 2 may seriously underestimate n_E , although the large sample sizes may reduce the consequences of these violations to some extent. Furthermore, Lynch & Walsh (1998) emphasize ‘that each estimate only applies to the specific pair of parental lines and that substantial differences would be likely if other parental stocks were used’ (p. 238).

Even a brief inspection of Fig. 1 makes clear that additive and additive + dominance models are insufficient to explain the data. With an additive model we would expect the means of both the F₁ and F₂ to lie halfway between the parental means. The mean of the F₁ lies actually closer to the mean of the LOW line, thus suggesting partial dominance. (Note that the LOW line may be truncated by the lower limit for the expression of eyespot size). The mean of the F₂, however, lies closer to the mean of the HIGH line. This shift might be a result of the breakdown of linkage disequilibrium or to differences in rearing conditions (e.g. temperature) because rearing of the F₁ and F₂ could not be performed simultaneously. A third explanation might be that the F₂ is more sensitive to temperature (i.e. shows more phenotypic plasticity) than either the parental lines or the F₁. The parental lines (especially the LOW line) are less plastic than the stock (cf fig. 6 in Brakefield *et al.*, 1996), and the reaction norms of the F₂ differ markedly from those of the stock (P.J. Wijngaarden and P.M. Brakefield, unpubl. obs.); unfortunately, the reaction norms of the F₁ have never been assessed.

Joint scaling tests confirm the insufficiency of additive and additive + dominance models. Including epistasis might give an adequate fit between model and data, but our number of lines did not allow an evaluation of such a model. Despite this lack of additivity considerable amounts of additive genetic variance were available to create the difference between the HIGH and LOW line. It is well known that dominance and epistasis can contribute to the additive genetic variance (Lynch & Walsh, 1998). Unfortunately, these different sources of additive variance are hard to disentangle experimentally.

In Nijhout’s (1980) model of eyespot formation the individual scale cells that make up the eyespots produce only one pigment depending on the morphogen level relative to certain threshold values in these cells. It should therefore not have come as a surprise that simple additive or additive + dominance models do not work. Transplant experiments showed that the response to selection on the size of eyespots on the dorsal wing surfaces was mainly caused by changes in the activity of the focal cells and, to a lesser extent, by changes in the sensitivity of the surrounding cells (Monteiro *et al.*,

1994). Unless the sensitivities of these surrounding cells can be kept constant eyespot size will have an epistatic component. Similar results were obtained by Nijhout & Paulsen (1997) when they explored the behaviour of a one-dimensional diffusion gradient and threshold model that might apply in many developmental systems (including butterfly eyespots). The model contained six parameters, each controlled by a single locus with two alleles. Although the alleles were assumed to have additive effects, dominance and a substantial amount of epistasis showed up as emergent properties.

The difference in mean eyespot size in the HIGH and LOW lines is the result of artificial selection, initially at a single intermediate temperature. A similar difference in phenotype can be obtained by rearing larvae at different temperatures (i.e. through phenotypic plasticity). Investigations of the physiological basis of this environmentally induced variation in eyespot size have shown that the dynamics of ecdysteroid hormones shortly after pupation play an important role in the response to rearing temperature (see Koch *et al.*, 1996; Brakefield *et al.*, 1998). In addition, individuals carrying the mutant *Bigeys* allele show substantially larger ventral eyespots than wild-type individuals at any particular temperature, an effect that probably involves the response to focal signalling (Brakefield *et al.*, 1996). It seems likely that alleles of genes, which influence both of these classes of change in eyespot size, will contribute to the difference between the selected lines analysed in this report. A challenge for future research is to map and identify at least some of these genes using both a candidate gene approach for developmental genes such as those of the *hedgehog* signalling pathway (Carroll *et al.*, 1994; Brakefield *et al.*, 1996; Keys *et al.*, 1999), and for genes involved in the hormonal regulation of phenotypic plasticity in eyespot size (Brakefield *et al.*, 1998).

We found no evidence for cytoplasmic effects (i.e. maternal effects or effects of loci on the Y-chromosome). Differences between reciprocal crosses were found that are consistent with effects of one or more X-linked loci. In both cases where X-linkage might become apparent the differences between the reciprocal crosses accounted for 9.3% of the difference between the parental lines. That these differences account for the same percentage is consistent with the hypothesis that dosage compensation is absent in butterflies (Johnson & Turner, 1979) because they are caused by different numbers of X-chromosomes. X-linkage is common for quantitative traits in Lepidoptera and is thought to play an important role in species differences (Sperling, 1994). Its implications for *B. anynana* are difficult to assess as long as we do not know to what extent differences in eyespot size can be attributed to differences in source activity and differences in threshold values.

Acknowledgements

We are grateful to Russ Lande for his comments on an earlier version of the manuscript.

References

- BRAKEFIELD, P. M. 1998. The evolution–development interface and advances with the eyespot patterns of *Bicyclus* butterflies. *Heredity*, **80**, 265–272.
- BRAKEFIELD, P. M. AND FRENCH, V. 1999. Butterfly wings: the evolution and development of colour patterns. *Bioessays*, **21**, 391–401.
- BRAKEFIELD, P. M., GATES, J., KEYS, D., KESBEKE, F., WIJNGAARDEN, P. J., MONTEIRO, A. *ET AL.* 1996. Development, plasticity and evolution of butterfly eyespot patterns. *Nature*, **384**, 236–242.
- BRAKEFIELD, P. M., KESBEKE, F. AND KOCH, P. B. 1998. The regulation of phenotypic plasticity of eyespots in the butterfly *Bicyclus anynana*. *Am. Nat.*, **152**, 853–860.
- CARROLL, S. B., GATES, J., KEYS, D., PADDOCK, S. W., PANGANIBAN, G. F., SELEGUE, J. *ET AL.* 1994. Pattern formation and eyespot determination in butterfly wings. *Science*, **265**, 109–114.
- CASTLE, W. E. 1921. An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. *Science*, **54**, 223.
- COCKERHAM, C. C. 1986. Modification in estimating the number of genes for a quantitative character. *Genetics*, **114**, 659–664.
- EFRON, B. AND TIBSHIRANI, R. J. 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- FRENCH, V. AND BRAKEFIELD, P. M. 1992. The development of eyespot patterns on butterfly wings: morphogen sources or sinks? *Development*, **116**, 103–109.
- JOHNSON, M. S. AND TURNER, J. R. G. 1979. Absence of dosage compensation for a sex-linked enzyme in butterflies (*Heliconius*). *Heredity*, **43**, 71–77.
- KEYS, D. N., LEWIS, D. L., SELEGUE, J. E., PEARSON, B. J., GOODRICH, L. V., JOHNSON, R. J. *ET AL.* 1999. Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science*, **283**, 532–534.
- KOCH, P. B., BRAKEFIELD, P. M. AND KESBEKE, F. 1996. Ecdysteroids control eyespot size and wing color pattern in the polyphenic butterfly *Bicyclus anynana* (Lepidoptera: Satyridae). *J. Insect Physiol.*, **42**, 223–230.
- LANDE, R. 1981. The minimum number of genes contributing to quantitative variation between and within populations. *Genetics*, **99**, 541–553.
- LYNCH, M. AND WALSH, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
- MANLY, B. F. J. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*, 2nd edn. Chapman & Hall, London.
- MONTEIRO, A. F., BRAKEFIELD, P. M. AND FRENCH, V. 1994. The evolutionary genetics and developmental basis of wing pattern variation in the butterfly *Bicyclus anynana*. *Evolution*, **48**, 1147–1157.

- NIJHOUT, H. F. 1980. Pattern formation on Lepidopteran wings: determination of an eyespot. *Dev. Biol.*, **80**, 267–274.
- NIJHOUT, H. F. 1991. *The Development and Evolution of Butterfly Wing Patterns*. Smithsonian Institution Press, Washington, DC.
- NIJHOUT, H. F. AND PAULSEN, S. M. 1997. Developmental models and polygenic characters. *Am. Nat.*, **149**, 394–405.
- SACCHERI, I. J., WILSON, I. J., NICHOLS, R. A., BRUFORD, M. W. AND BRAKEFIELD, P. M. 1999. Inbreeding of bottlenecked butterfly populations: estimation using the likelihood of changes in marker allele frequencies. *Genetics*, **151**, 1053–1063.
- SPERLING, F. A. H. 1994. Sex-linked genes and species differences in Lepidoptera. *Can. Entomol.*, **126**, 807–818.
- TURNER, J. R. G. AND SHEPPARD, P. M. 1975. Absence of crossing-over in female butterflies. *Heredity*, **34**, 265–269.
- WRIGHT, S. 1968. *Evolution and the Genetics of Populations. I. Genetic and Biometric Foundations*. University of Chicago Press, Chicago, IL.
- ZENG, Z.-B. 1992. Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. *Genetics*, **131**, 987–1001.
- ZENG, Z.-B., HOULE, D. AND COCKERHAM, C. C. 1990. How informative is Wright's estimator of the number of genes affecting a quantitative character? *Genetics*, **126**, 235–247.