Microarray Analysis of Replicate Populations Selected Against a Wing-Shape Correlation in *Drosophila melanogaster*

Kenneth E. Weber,*,[1] Ralph J. Greenspan,[†] David R. Chicoine,* Katia Fiorentino,*
Mary H. Thomas* and Theresa L. Knight*

*Department of Biological Sciences, University of Southern Maine, Portland, Maine 04104-9300 and [†]The Neurosciences Institute,
San Diego, California 92121

ABSTRACT

We selected bidirectionally to change the phenotypic correlation between two wing dimensions in *Drosophila melanogaster* and measured gene expression differences in late third instar wing disks, using microarrays. We tested an array of 12 selected lines, including 10 from a Massachusetts population (5 divergently selected pairs) and 2 from a California population (1 divergently selected pair). In the Massachusetts replicates, 29 loci showed consistent, significant expression differences in all 5 line-pair comparisons. However, the significant loci in the California lines were almost completely different from these. The disparity between responding genes in different gene pools confirms recent evidence that surprisingly large numbers of loci can affect wing shape. Our results also show that with well-replicated selection lines, of large effective size, the numbers of candidate genes in microarray-based searches can be reduced to realistic levels.

THE Drosophila wing is a convenient model system for shape genetics. Our approach has been to select against phenotypic correlations within the wing, using the metric of "angular offsets" (WEBER 1990). Angular offsets reduce shape to its simplest quantifiable aspect—the allometric relation between two interland-mark distances—by measuring each individual's deviation from the mean line of allometry of the base population (see MATERIALS AND METHODS). This converts variation that is orthogonal to the correlation into a univariate scale that is independent of size. Thus, angular offsets focus directly on the evolutionary constraints posed by correlations (LANDE 1979) and quantify the breakage of these constraints (*cf.* BELDADE *et al.* 2002).

Many genes affect wing shape. Quantitative trait locus (QTL) mapping found at least 20 genes affecting one wing-shape trait in divergently selected lines from a wild sample (WEBER *et al.* 1999, 2001). Experimental selection can change wing shape in many directions (WEBER 1990; MEZEY and HOULE 2005) and can affect small isolated parts (WEBER 1992). A screen of 50 random *P*-element insertions found 11 insertions with rigorously validated wing-shape effects (WEBER *et al.* 2005). In these cases, gene effects were in the range of 0.1–1 phenotypic standard deviations of the base population. These results show that the wing has a high short- and long-term potential to evolve new shapes based on many existing alleles with small effects and many more loci that could mutate to produce such alleles. This suggests that replicate lines under identical selection might use different genes to produce similar changes in wing shape.

Identical selection regimes often produce different genetic outcomes. In a classic example, GOODALE (1941, 1953) selected for heavier mice and created a large-boned line, while MACARTHUR (1949) selected mice in the same way from a different stock and obtained a line that was not large boned, but obese. In another case (SWALLOW *et al.* 1998; HOULE-LEROY *et al.* 2003), mice from a single stock were selected for wheel running in four replicate lines. All lines responded with comparable performance increases, but two distinct suites of physiological and morphological traits emerged. COHAN and HOFFMANN (1986) selected for ethanol tolerance in *Drosophila melanogaster* from five locations along the North American West Coast and found that response occurred in genetically different ways. In these and other cases (*e.g.*, GROMKO 1995; BULT and LYNCH 1996), replicate selection lines from different strains, or even from the same strain, produced the same selected phenotype by different genetic means.

On the other hand, identical selection often produces parallel genetic outcomes, even when starting with different strains or species. For example, in cereals like sorghum, rice, and maize, orthologous loci have been involved in parallel changes during domestication for traits such as large seeds and seasonal independence of flowering (PATERSON *et al.* 1995; DEVOS 2005). Experiments with bacteriophages adapting to high tempera-

[1]*Corresponding author:* Department of Biological Sciences, University of Southern Maine, 96 Falmouth St., Portland, ME 04104-9300.
E-mail: keweber@usm.maine.edu

ture (BULL *et al.* 1997; WICHMAN *et al.* 1999) or toxic chemicals (CUNNINGHAM *et al.* 1997) show that identical selection in different lines can lead to parallel and sometimes almost identical changes in the same genes. WOOD *et al.* (2005) review other cases of parallel genetic change in strains and species subjected to the same selective regime.

In adaptive radiations, parallel evolution often involves the same few major loci. Parallel morphological and genetic differences were found in independent cases of marine sticklebacks adapted to fresh water (SCHLUTER *et al.* 2004; SHAPIRO *et al.* 2004), in species of Drosophila that independently evolved similar pigmentation of wings (PRUD'HOMME *et al.* 2006) or abdomens (GOMPEL and CARROLL 2003), and in beak length in Darwin's finches (ABZHANOV *et al.* 2006). SCHLUTER *et al.* (2004) provide other examples both supporting and contradicting this principle.

In the most striking cases of genetic convergence, identical amino acid substitutions have occurred in the same proteins in unrelated groups (PATTHY 1999; CARROLL 2006). Yet when selection responses can be more fully dissected, genetic differences emerge. In cereal domestication, genetic parallelism is high in some traits but low in others (GALE and DEVOS 1998; MORRELL and CLEGG 2007). In the bacteriophage studies (WICHMAN *et al.* 1999), loci with parallel changes were not those with the largest effects. In Drosophila, the same gene caused wing spots in independent lineages, but the regulatory modules were different (PRUD'HOMME *et al.* 2006).

Genetic divergence or convergence during selection depends partly on population size and selection intensity and the nature of the selected trait. In the wheel-running study cited above, the alternative outcomes were attributed to drift in small populations, because they depended on the presence of a single allele with low frequency in the base population (HOULE-LEROY *et al.* 2003). Selection intensity may influence the relative recruitment of major or minor genes (LANDE 1983), as studies of the evolution of insecticide and herbicide resistance have emphasized (McKENZIE *et al.* 1992; GARDNER *et al.* 1998; NEVE and POWLES 2005). The primary determinant of genetic divergence or convergence may be the complexity of the trait. Features that utilize much information in development, and can be molded easily in many ways by selection, should permit alternative paths to similar phenotypes, and their responses in parallel selection lines should vary more. Wing shape seems to be such a trait.

In this study, we used microarrays to measure gene expression in the whole genome in a large panel of selection lines. The lines were created in different experiments, originated from separate populations, and included multiple replicates of one population, but all were created using the same selection regime and shape trait. Here we evaluate the data with two aims: (1)
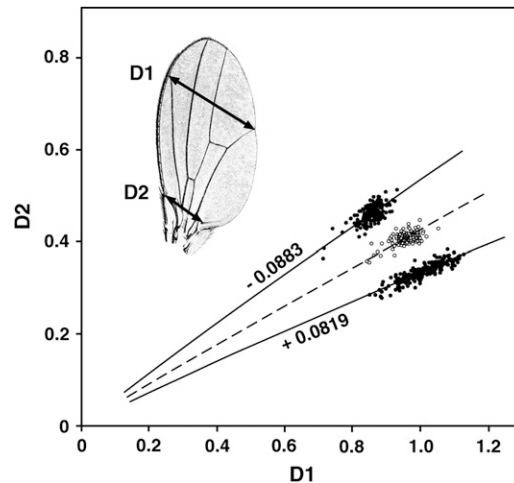

FIGURE 1.—Definition of the trait. The dashed line shows the mean allometric relationship of D1 and D2 in wild-type males. The phenotype of each wing is the angular offset of its point (D1, D2) from this baseline in radians of rotation about the origin. Selection on this angle produces antagonistic changes in D1 and D2. The figure shows samples of male wings from a wild-type population (open circles), with a mean angular offset of approximately zero, and from two divergently selected populations (solid circles) with mean offsets of -0.0883 and $+0.0819$ radians (K. E. WEBER, unpublished results). Environmental variables like temperature and culture density have large effects on body size (*i.e.*, the mean and variance of r), but little effect on angular offset. Offsets are normally distributed in control and selected flies.

to identify candidate wing-shape genes and (2) to assess variation in the outcome of identical selection regimes—between replicate lines from a single source and between lines from two geographically remote, local gene pools in Massachusetts and California.

MATERIALS AND METHODS

The trait and the selected lines: D1 and D2 are widths at the middle and base of the wing (Figure 1). The angular offset of a wing is the polar angle, in radians, between the point (D1, D2) and the point with equal radius (r) on the line of the polar equation $\theta = 0.4048r^{-0.043}$ (males) or $\theta = 0.4148r^{-0.134}$ (females). The polar equation is a curve approximating the mean allometric relation between D1 and D2, derived by regression of log θ on log r in wild-type flies (WEBER 1990). Clockwise and counterclockwise deviations from this baseline are called positive and negative, respectively. Angular offsets are independent of body size.

Lines H and L (WEBER 1990) were created by divergent selection of the most extreme 20% of 100 flies of each sex from a sample of a laboratory population established in 1981 from 350 isofemale lines captured in Lincoln, Massachusetts. The lines were selected for 20 generations, and all three major chromosomes were isogenized using balancer chromosomes. Lines H and L were then used to map QTL on chromosomes three (WEBER *et al.* 1999) and two (WEBER *et al.* 2001), using *in situ*-labeled insertion sites of the transposon *roo* as markers.

Figure 2 outlines the creation of eight more selection lines. In September 1999, lines H and L were crossed in four separate crosses of 500 males and 500 virgin females to create
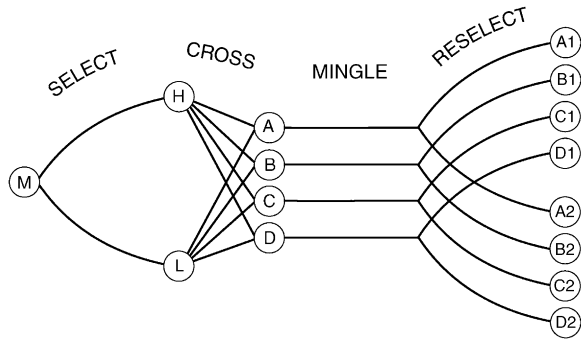
FIGURE 2.—Derivation of lines. A wild Massachusetts population sample (M) was selected divergently for 20 generations to produce lines H and L. H and L were isogenized and crossed to make populations A–D. After 34 generations of recombination in large populations ("Mingle"), hybrid lines were divergently selected for 25 generations. (Meanwhile, H and L were maintained with inbreeding, and a pair of divergent lines was selected from a California population.)

the hybrid lines A, B, C, and D. Before crossing we verified that both H and L were still isogenic for all the same *roo* markers that were used in QTL mapping on all three main chromosomes. Thus we can be certain that all hybrid populations began with identical fixed backgrounds and segregating alleles. Founder males and females were H and L, respectively, in A and B; C and D were the reciprocal cross. [About 10% of the phenotypic difference between H and L arises from the X chromosome (WEBER *et al.* 1999, 2001).] For 34 generations, each hybrid line was maintained in 30 vials with three female and five male parents per vial. In every generation, virgin offspring were collected for 9 days to include all eclosing flies, kept at 12° to prevent mating, and then randomly mated. Our aim was to mingle the H and L genomes in large panmictic populations with minimal selection.

After 34 generations of genetic mingling in these four hybrid lines, selection was reinitiated to derive a new pair of divergently selected lines from each hybrid line. The eight lines were designated A1, A2, B1, B2, and so on and were selected in the positive (1) or negative (2) direction. The parents were the most extreme 20% of 315 measured flies of each sex. In several generations fewer flies were measured, but the number of parents was always at least 63 of each sex per line. After 25 generations of selection, sublines were sib-mated for 10 generations to create inbred lines.

The California selected lines came from a base population that was derived from ∼40 *D. melanogaster* isofemale lines, captured in Davis, California, and kindly supplied by Michael Turelli (University of California, Davis, CA). In each generation, the most extreme 20% of 200 individuals of each sex were selected as parents. After 21 generations of selection, the lines were inbred by sib-mating for 10 generations, then maintained in vial cultures for several years, and then further inbred by sib-mating for 10 more generations, prior to these experiments.

Staging and dissection of larvae: Culture vials were set up with potato-flake medium plus 0.05% bromphenol blue to obtain semisynchronous larvae. The blue medium was retained by late third instar larvae but gradually excreted, permitting visual staging (MARONI and STAMEY 1983; ANDRES and THUMMEL 1994). We selected larvae with a median time to pupariation of 6.2 hr (data not shown). Larvae were dissected in cold PBS on siliconized slides using needle-tip tweezers, and wing disks were cleaned with tungsten needles. Each pair of disks was transferred to a tiny droplet of fresh PBS retained on the tip of a needle after dipping it in boiling deionized water

and then in ice-cold sterile PBS. Disks were transferred from this needle tip to the bottom of a 0.5-ml microcentrifuge tube in a covered gel-filled cooler (embedded in dry ice), where the droplet with disks was instantly frozen to the bottom. No damaged disks were used. Male and female larvae were not separated, so samples included both sexes about equally.

RNA was extracted (DIERICK and GREENSPAN 2006) by homogenization of 100 disks per sample, representing at least 50 larvae. Large samples of disks were used because individuals may vary greatly in expression levels (OLEKSIAK *et al.* 2002). For each sample, 90 culture vials were set up with blue medium and one mated female apiece. Only a few larvae came from each vial so any differences between cultures were averaged over many vials. Samples were analyzed using Affymetrix technology (Drosophila Genome Array Version 2) according to the manufacturer's protocols.

Statistical analysis of microarray data: The data set includes the scans of 36 Affymetrix Drosophila 2.0 microarray chips, representing six pairs of divergently selected lines and three samples per line. Results were analyzed using the software programs dChip (http://www.dchip.org; version of June 27, 2005), CyberT (Institute for Genomics and Bioinformatics, University of California, Irvine, CA), SAS and JMP5 (SAS Institute, Cary, NC), and Excel (Microsoft). Normalization of intensities and classification of probe sets as present, marginal, or absent by dChip was carried out anew within each subset of the data used in different tests. When line pairs were tested separately, the data for the six chips of each line pair were independently normalized using dChip and tested using a Bayesian analysis in CyberT after converting normalized intensities to natural logs. We used Bayesian methods for t-tests of individual line pairs because the sample sizes were small, with three chips for each selected line. For Bayesian statistical comparisons, variances were conditioned on the closest 50 probe sets above and below each probe set in order of their rank according to normalized intensity. Probabilities were corrected for the number of tests within each line pair by setting the significance threshold at $0.05/N$, where N was the number of tested probe sets for that pair. Probe sets were also tested using the combined data for all five Massachusetts line pairs in nested ANOVAs, with chips nested within lines and lines within treatments, using natural log-transformed intensities in SAS to provide overall P-values for each probe set, similarly corrected for the total number of tests.

The raw array data files are available at http://www.ncbi.nlm.nih.gov/projects/geo/ under accession no. GSE9107.

Plan of data analysis: Our analysis began with two preliminary steps. First, we examined the entire data set to assess the consistency and quality of the normalized intensity data and the performance of the analysis software. We then screened the data for genes with zero expression (null alleles) either in one direction of selection or in one gene pool. These preliminary analyses indicated that the data were of consistent quality and that the main analysis could be based solely on comparisons of expression levels among transcribed genes (see RESULTS for details).

In the main analysis, our first aim was to identify a small number of high-quality candidate genes. We assumed that the Massachusetts lines had the same contributory genes since they all came from the same original sample, but the contributory genes in Massachusetts and California lines might be different. Therefore, we derived our candidate gene list solely from the Massachusetts lines, since we had five pairs of them but only one pair of California lines.

We had two criteria for candidate genes: our list includes only genes that were (1) consistently significant in all five Massachusetts line pairs, when each line pair was tested independently, and that were also (2) significant in the nested

ANOVA of the combined Massachusetts data set. In the end, the candidate genes were completely decided by the first criterion, which was more stringent. This plan of analysis allowed us to use the same results to evaluate the incremental value of additional replicates, to investigate sources of variation in the outcome of selection among lines, and to compare outcomes between the Massachusetts and California lines.

Estimation of centimorgan values: We created a look-up table to estimate locations of genes on the genetic map by entering the experimental centimorgan values listed by Ashburner in LINDSLEY and ZIMM (1992, pp. 1117–1133) and interpolating values for intervening genes. The centimorgan values were plotted as a function of the starting nucleotide address for genes as given in the Affymetrix gene annotation list. Curves were fit to the data using the spline function in JMP5. The spline parameters of the best-fit curve for each major chromosome arm were:

$$X \quad \lambda = 5.00E + 16 r^2 = 0.9995$$

$$2L \quad \lambda = 3.00E + 16 r^2 = 0.9982$$

$$2R \quad \lambda = 3.00E + 17 r^2 = 0.9986$$

$$3L \quad \lambda = 4.00E + 16 r^2 = 0.9987$$

$$3R \quad \lambda = 2.00E + 16 r^2 = 0.9997.$$

Calculations of heritabilities and effective factors: Heritabilities and effective gene numbers were calculated for the four hybrid Massachusetts populations (A, B, C, and D) using the selection data (FALCONER and MACKAY 1996). We calculated realized heritabilities for each of the eight derived selection lines by the method of HILL (1972) on the basis of the first six generations of selection. We used the mean heritability of each pair of high and low lines to estimate the heritability of each hybrid population and the means of the two phenotypic standard deviations in the first generation of selection to estimate the standard deviation for each hybrid population. Total response was the difference between high and low lines after 25 generations of selection and 10 generations of inbreeding.

RESULTS

The phenotypic variances of isogenic lines H and L and of their four F_1 hybrids were all nearly equal (Table 1). In the F_2 generation, hybrid variance increased by a factor of ~6, as large blocks of high and low selected alleles began to segregate. During 34 subsequent generations of random mating in populations of an effective size of ~180, mean hybrid variance declined almost to the same value as base population variance, indicating approximate linkage equilibrium for alleles affecting the trait.

In generation F_{35}, we began selecting divergently on paired sublines from each hybrid line. After 25 generations of selection, these new lines diverged about as much as the parental H and L lines and more in some cases (Table 2). Thus no significant genetic variation for the trait was lost during the 34 generations of hybrid mingling. After selection, the 8 new lines were sib-mated

TABLE 1

Means and variances of parental and hybrid Massachusetts lines

	Mean	Variance ($\times 10^{-5}$)
Parental line H	+0.0559	8.9
Parental line L	−0.0756	6.9
F_1	+0.0013	9.3
F_2	−0.0069	62.1
Mean of F_{31}–F_{34}	−0.0133	14.2
Wild-type base	+0.0067	11.4

Means in radians of angular offset as defined in MATERIALS AND METHODS. For lines H and L, $N = 100$. For generations F_1–F_{34}, values are grand means of the means and variances of four H × L hybrid lines, with $N = 100$ for each line/generation. For wild-type base, values are means of the mean and variance of two lines of the unselected Massachusetts base population (CN1 and CN2) with $N = 150$ in each sample. Data are from males only.

for 10 generations to reduce genetic variation within lines. The two California lines were selected for 21 generations and then also inbred by sib-mating. Most of the phenotypic divergence between high and low lines remained after inbreeding. Tables 2 and 3 summarize the phenotypes of all 12 selected lines before and after inbreeding or, in the case of the H and L lines, before and after isogenization with balancers. The measurements in Table 3 were made just before microarray assays were performed on all lines.

Preliminary analysis of microarray data quality: On the Affymetrix chip, each probe is present in two versions on contiguous spots: a 25-base version (PM) that is a perfect genomic match and another 25-base version (MM) with a single mismatch at the 13th base. A probe set includes 14 different paired PM/MM probes from one transcript, and the chip includes probe sets representing most *D. melanogaster* transcripts. The dChip software normalizes hybridization intensities across chips and classifies probe sets as present, absent, or marginal. A few probe sets that were called present had negative or zero intensities when evaluated as the sum of PM–MM differences. We eliminated these as well as Affymetrix control sequences. We then looked at several fundamental aspects of the data after normalization by dChip across all 36 chips.

Figure 3A shows distributions of hybridization R-values, where $R = (PM - MM)/(PM + MM)$ for all PM/MM probe pairs on all 36 chips (~9.5 × 10^6 probe pairs) after separation into present, absent, and marginal probe sets. In absent probe sets, R is distributed symmetrically around zero, as expected for completely random hybridization. In the present probe sets, R has a primary mode on the positive side and a secondary mode showing a minor population of randomly hybridizing probe pairs within present probe sets. Figure 3B shows the distributions of probe sets on all 36 chips

TABLE 2

Mean phenotypes of all selected lines before inbreeding or isogenization

Up lines	Phenotypes	N	Down lines	Phenotypes	N	Ranges
H	0.0537 ± 0.0011	75	L	-0.0819 ± 0.0007	97	0.1356
A1	0.0706 ± 0.0006	328	A2	-0.0672 ± 0.0007	287	0.1378
B1	0.0688 ± 0.0007	298	B2	-0.0755 ± 0.0007	307	0.1443
C1	0.0506 ± 0.0006	284	C2	-0.0716 ± 0.0009	168	0.1222
D1	0.0609 ± 0.0007	269	D2	-0.0777 ± 0.0007	183	0.1386
Ca1	0.0516 ± 0.0006	200	Ca2	-0.0771 ± 0.0008	214	0.1287

Means and SE in radians of angular offset. Data are from selection generation 20 in H and L; 22 in Ca1 and Ca2, and 24 in line pairs A–D. N is the sample size. All data are from males.

($\sim 6.8 \times 10^5$ probe sets), according to the number of probe pairs (0–14) in each probe set showing the anomalous condition MM > PM for present, absent, and marginal probe sets. Again, present and absent calls are well differentiated, and the distribution for absent probe sets is that expected for random hybridization. Figure 3 shows that present and absent calls by dChip efficiently separated our probe sets into two distinct groups with appropriate distributions. Probe sets classified as marginal were much more similar to absent probe sets than to present ones. We therefore did not include the marginal probe sets in our analysis.

We also asked how consistently each transcript was called present across all 36 chips. Figure 4 shows the distribution of all $\sim 18,800$ probe sets on the Affymetrix chip, according to the number of times each one was called present in 36 chips. Almost 7000 probe sets were called present on every chip. At the other extreme, ~ 5000 probe sets were not called present even once. The remaining $\sim 36\%$ of probe sets were present on some but not all chips. Thus the majority were either always present or never present.

The dChip notes recommend special attention to any chip where >5% of probe sets are classified as outliers. The average percentage of outliers per chip, among all 36 chips, was only 0.49% according to dChip. A single atypical chip (H-3) had 4.08% outliers.

Absent genes associated with lines or treatments: Absent probe sets could be null expression alleles, which could be associated with trait variation. We first searched all 30 Massachusetts chips for probe sets that were called absent in one direction of selection but present in the other. We found only seven. Visual inspection of the PM/MM profiles in dChip showed that, in all cases, both the present and absent probe sets showed nearly identical profiles and essentially zero expression. These cases clearly represent random error in the evaluation of genes that are not actually expressed in the wing.

We next compared present and absent calls between Massachusetts and California to check for genes expressed in one population and not the other. There are 10 possible comparisons between Massachusetts line pairs, with a mean of 7187 shared present probe sets (Table 4). There are five possible comparisons between the California line pair and the Massachusetts line pairs, with a mean of 7160 shared present probe sets. Since the within-population comparisons must represent a single group of probe sets and the between-population comparisons show almost the same numbers, most present/absent differences are probably due to random error.

We conclude that null expression alleles may not be important in the genetic variance of this trait. By eliminating absent genes from further comparisons

TABLE 3

Mean phenotypes of all selected lines after inbreeding or isogenization

Up lines	Phenotypes	N	Down lines	Phenotypes	N	Ranges
H	0.0575 ± 0.0007	50	L	-0.0751 ± 0.0007	50	0.1326
A1	0.0711 ± 0.0010	50	A2	-0.0665 ± 0.0013	50	0.1376
B1	0.0664 ± 0.0009	50	B2	-0.0545 ± 0.0011	50	0.1209
C1	0.0534 ± 0.0016	50	C2	-0.0636 ± 0.0014	50	0.1170
D1	0.0686 ± 0.0010	50	D2	-0.0650 ± 0.0012	50	0.1336
Ca1	0.0449 ± 0.0010	50	Ca2	-0.0657 ± 0.0012	50	0.1106

Means and SE in radians of angular offset. H and L were isogenized after selection and then maintained by sib-mating. The other lines were inbred by sib-mating after selection. These measurements were made just before the microarray assays. All data are from males.
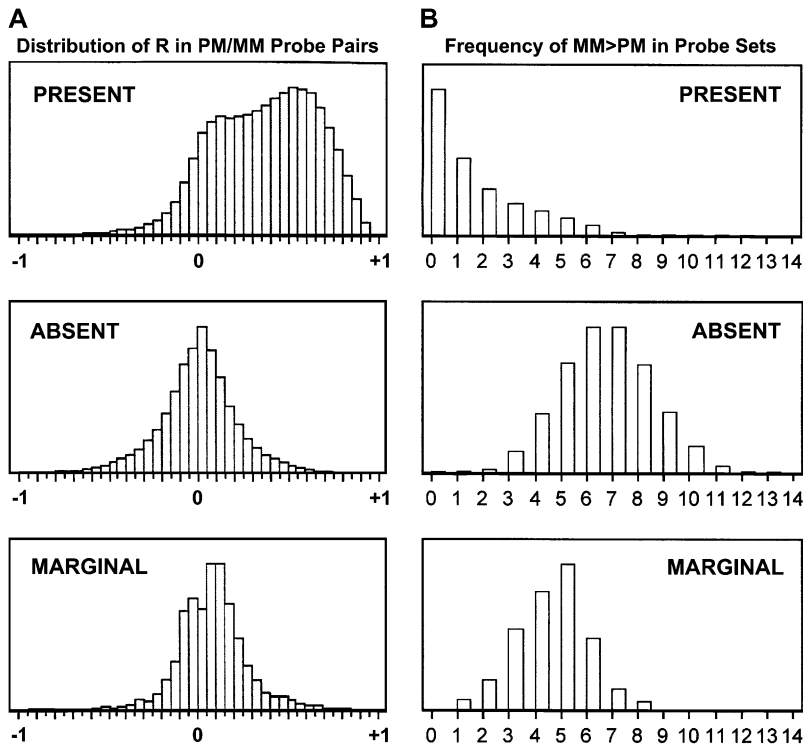
A

Distribution of R in PM/MM Probe Pairs

PRESENT

-1   0   +1

ABSENT

-1   0   +1

MARGINAL

-1   0   +1

B

Frequency of MM>PM in Probe Sets

PRESENT

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

ABSENT

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

MARGINAL

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

FIGURE 3.—Distributions of $\sim 9.5 \times 10^6$ probe pairs and $\sim 6.8 \times 10^5$ probe sets from 36 chips. (A) Values of R for probe pairs in probe sets called present, absent, and marginal, where $R = (PM - MM)/(PM + MM)$. (B) Values of X for probe sets called present, absent, and marginal, where X = probe pairs per probe set with MM > PM.

between populations or treatments, we simplified the rest of the analysis, which was focused on differences in expression intensity among genes called present on chips in both directions of selection or in both gene pools.

Transcripts associated with the trait in individual line pairs: We began our main analysis by testing each line pair separately. For each pair we tested probe sets for significant expression differences between high and low lines by *t*-test as explained in MATERIALS AND METHODS. Table 5 shows the number of testable probe sets for each line pair and the number with significant expression differences, treating each line pair as an independent experiment and correcting for the number of tests in each pair. Testable probe sets included all for which at least two of the three samples were present in both the high and low line. This number was consistent with a range of 8346–8435. By contrast, the number of significant probe sets was highly variable. We tested the numbers of nonsignificant *vs.* significant probe sets in Table 5 (R × C test of independence) to see whether the proportion of significant probe sets was independent of line pair. The *G*-value was ~ 268, with d.f. = 5 and $P < 10^{-55}$. The low *P*-value points to some factor or factors with a large effect on the numbers of significant probe sets.

The effect of population size in selection: Figure 5 shows that the number of significant probe sets in line pairs is not correlated with phenotypic divergence, but may be influenced by the effective population sizes of lines during selection. Larger lines have fewer significant transcripts ($r^2 = 0.872$; $P = 0.0067$, assuming equal

variance at all population sizes). Perhaps alleles with no effect on the trait are more likely to be fixed in opposite directions by drift in smaller lines, but are controlled more by fitness selection in larger lines. The slope of the line in Figure 5B does not estimate the isolated effect of population size, because different base populations are included and because some reduction in the association of hitchhiking alleles with the trait would be expected in the derived lines even at the same population size. In any case, the four derived Massachusetts line pairs show a large, concerted decrease in significant probe sets compared to their parental lines, H and L. Yet they have approximately the same phenotypic divergence (Table 3). We conclude that many allelic differences that do not affect the trait were fixed oppositely in the parental lines and that in the larger, derived lines these differences were more likely to be fixed for the same allele or to remain unfixed.

Consistency of significant probe sets in Massachusetts lines: We next compared the five Massachusetts line pairs to see how often the same probe sets were significant in different line pairs. Table 6 shows that 35 probe sets were significant in all five cases. Many more probe sets were significant in fewer than five. These were most numerous in the H/L lines, which were the parents of the other four pairs. For example, of 449 probe sets that were significant only once, almost half (218) were significant in the H/L comparison. Again, the greater abundance of significant transcripts in the H/L line pair is likely to be due to alleles that do not affect the trait. These alleles might have been separated from selected alleles by recombination in the hybrid
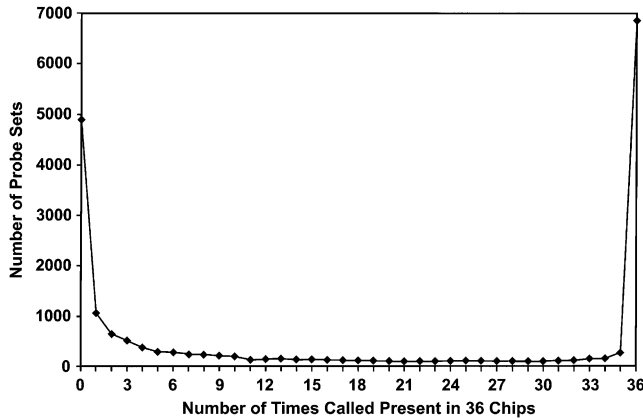
FIGURE 4.—Distribution of probe sets according to the number of chips on which each probe set was called present. Includes all Drosophila probe sets and Affymetrix control probe sets. Most probe sets were either always present or always absent, indicating accurate detection.

populations with their larger size. This recombination could have occurred either during the 34 generations of mingling in the hybrid lines or during the 25 subsequent generations of selection.

Some probe sets that were significantly different in multiple line pairs showed expression differences that were in opposite directions in individual line pairs. That is, their expression was significantly higher in the high line *and* significantly higher in the low line in different line pairs. The final column in Table 6 shows the number of such cases. For example, of 164 probe sets that were significant in just two line pairs, 21 were significant in opposite directions. Overall, $\sim$10% (39/375) of all probe sets that were significant more than once were expressed in opposite directions in different line pairs. For simplicity, we refer to these probe sets as "flip-flops." They might be genes that are expressed oppositely in different epistatic complexes to produce equivalent phenotypic effects. A simpler explanation would be that they represent randomly fixed expression polymorphisms with no relation to the trait.

The value of additional replicates: We asked how the number of consistently significant probe sets declines as

TABLE 4

Consistently present probe sets within and between pairs of lines

	Ca1/Ca2	H/L	A1/A2	B1/B2	C1/C2	D1/D2
Ca1/Ca2	7527					
H/L	7051	7346				
A1/A2	7225	7111	7616			
B1/B2	7151	7065	7258	7504		
C1/C2	7190	7084	7290	7216	7524	
D1/D2	7181	7099	7280	7233	7229	7547

Table includes probe sets called present in all 12 chips in each possible comparison of two line pairs.

TABLE 5

Testable and significant probe sets for each line pair

Line pair:	H/L	A1/A2	B1/B2	C1/C2	D1/D2	Ca1/Ca2
Testable probe sets	8350	8435	8367	8376	8405	8346
Significant probe sets	534	335	253	198	238	450

Probabilities were computed with log-transformed intensities using Cyber-T with Bayesian methods and compared to $P = 0.05/$(testable probe sets in each line pair).

more replicates are added to the analysis. For this purpose we treated all five Massachusetts line pairs as replicates, representing approximately the same set of alleles and selection history. The average number of significant probe sets for individual Massachusetts line pairs in Table 5 is 312. Figure 6 shows that the addition of a second replicate line pair eliminated about two-
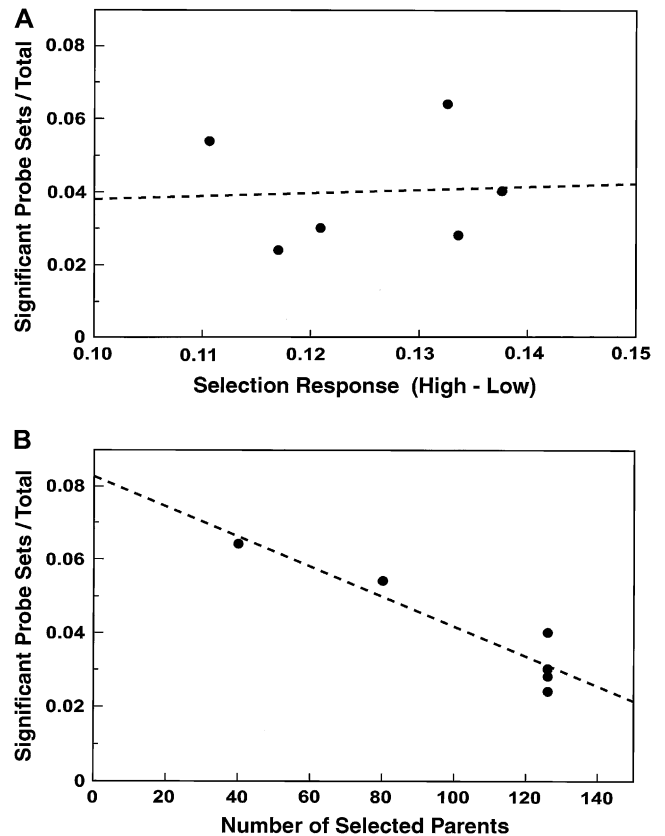

FIGURE 5.—The proportion of significant probe sets within each line pair (Table 5) is not correlated with phenotypic divergence, but with population size. Each point represents one line pair. (A) Proportion of significant probe sets as a function of the phenotypic divergence between high and low lines from Table 3; $r^2 = 0.003$. (B) Proportion of significant probe sets as a function of the number of parents per generation during selection; $r^2 = 0.872$. Lines H and L had 40 parents, lines Ca1 and Ca2 had 80, and the derived Massachusetts lines had 126.

TABLE 6

Repeated significance of the same probe sets in Massachusetts replicates

Times significant	Total	H/L	A1/A2	B1/B2	C1/C2	D1/D2	Opposite
5/5	35	35	35	35	35	35	2
4/5	78	74	66	57	54	61	6
3/5	98	89	60	46	35	64	10
2/5	164	118	78	45	31	56	21
1/5	449	218	96	70	43	22	
Total		534	335	253	198	238	
0/5	8541						
Total probe sets tested: 9365							

The first column gives the number of times a probe set was significant in the five comparisons. The final column gives the numbers of probe sets, included in the totals, that were significant multiple times but not always in the same direction ("flip-flops").

thirds of significant probe sets, on average, by the criterion of consistent significance. The addition of a third replicate eliminated about half of the remaining probe sets. With the addition of more and more replicates, probe sets that remain consistently significant become increasingly interesting, but all probe sets would eventually be eliminated by the occurrence of occasional false negatives. According to Figure 6, in retrospect, the critical or minimum number of replicate pairs of lines in our experiment was three. The addition of the fourth and fifth replicates was still possibly worth the effort in terms of the increasing resolution of probe sets. By extrapolation, a sixth replicate would be of marginal value.
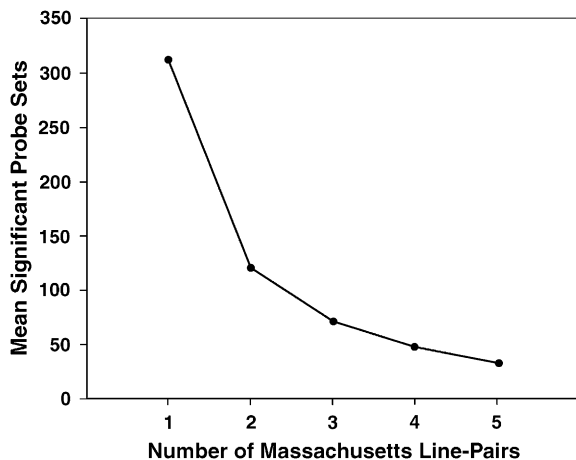

FIGURE 6.—As the number of replicate line pairs increases, the mean number of consistently significant probe sets falls rapidly and then levels off. The first ordinate value (311.6) is the mean of significant probe sets in individual Massachusetts lines from Table 5. The final value (33) is the number of probe sets that were consistently significant in all five Massachusetts line pairs (Table 6), minus the two cases with contradictory signs ("flip-flops"). The values for two, three, and four line pairs were calculated as the mean numbers of consistently significant probe sets in all possible two-way, three-way, and four-way combinations, minus cases with contradictory signs.

Transposons: Transposons are disproportionately frequent among the significant probe sets in Table 6, especially among those with self-contradictory expression differences, described above as "flip-flops." The total of 39 flip-flop probe sets in the final column of Table 6 includes at least 11 transposons. The two flip-flops that were significant in all five line pairs are both transposons. The large number of transposon flip-flops probably arises in part because transposons are present in numerous copies in the genome and most active insertion sites are not fixed. This allows their copy number to diverge between selection lines. The divergence may be random or could be affected by hitchhiking due to associated genes for the trait. But not only flip-flops are transposons: among the 35 consistently significant probe sets in Table 6, 4 are transposons that show significant expression differences in the same direction of expression in all five line pairs. These four transposons, with their lengths and approximate copy numbers per genome according to FlyBase (CROSBY *et al.* 2007), are *hopper* (1435 bp; ~15), *Rt1b* (~5100 bp; ~40), *3S18* (6126 bp; ~20), and *S* (1736 bp; ~40). One interpretation of such an association might be that the transposon itself affects the trait. It is also possible that these are cases where actively transcribed transposons are inserted in genes that affect the trait in such a way that they create selectable alleles. This would mean that the insertion is a marker of a trait allele, but one that is difficult to utilize in gene hunting because of the existence of many other copies of the same transposon. We removed all six transposons (two flip-flop, four non-flip-flop) from the list of 35 consistently significant probe sets, leaving 29 candidate genes for the trait. These are listed in Table 7 with some annotation according to FlyBase.

A test of the combined data: We regarded independent significance in all five of five Massachusetts line pairs as a pragmatic first criterion for choosing candidate genes. Genes that show such a consistent effect may be easier to validate. For a second, overall perspective,

TABLE 7

Candidate genes from the Massachusetts lines

Gene	Arm	Band	cM	Fold	California?	Function
CG3625	2L	21B7	0.4	+1.8	−	Unknown
CG4291	2L	21D4	1.5	−1.7	0	RNA binding (spliceosome)
CG3036	2L	25B1-2	14.2	+2.5	0	Na/phosphate symporter
CG31918	2L	25C1-3	15.1	+2.1	0	Unknown
CG10570	2L	37A4	54.0	+2.8	0	Unknown
CG31617 (*His1*)	2L	39E1	54.6	−4.6	−	Histone
GH15538	2L	40F6	54.9	+2.1	0	Unknown (Jensen *et al.* 2006)
CG40293 (*Stlk*)	2R	41C1-6	55.0	−1.5	−	Ste20-like kinase, cytokinesis
CG10067 (*Act57B*)	2R	57B5	95.4	+4.2	+	Actin
CG17090	3L	61C3-5	0.1	−2.6	0	ser/thr kinase (apoptosis)
CG13895	3L	61C8	0.2	+1.9	0	Transcription factor
CG2469	3L	61F5-6	0.4	−1.6	0	Unknown
CG9186	3L	61F6	0.4	−2.1	0	Unknown
CG2211	3L	61F6	0.4	+2.2	−	Unknown
CG10359	3L	63E5	6.8	+4.2	+	Innate immunity
CG32373	3L	66A8-12	23.0	+4.1	−	Unknown
CG6776	3L	66D5	26.0	−3.3	0	Glutathione transferase
CG13053	3L	72D10	43.6	+3.3	0	Unknown
CG13023	3L	73D6	44.4	−1.7	0	Unknown
CG4319 (*rpr*)	3L	75C6	46.1	−2.9	0	Apoptosis
CG4144 (*GNBP2*)	3L	75D6	46.4	+2.2	0	Gram-negative bacteria-binding protein 2
CG31522	3R	82B2-3	47.2	−2.7	0	Long-chain fatty acid acyltransferase
CG1163 (*Rpll18*)	3R	83A1	47.5	+1.5	0	RNA polymerase II 18-kDa subunit
CG10287 (*Gasp*)	3R	83D4	47.7	+2.0	+	Chitin-binding protein
CG18213	3R	89E10	59.1	−4.0	0	Unknown
CG5184 (*mRpS11*)	3R	89E11	59.2	+1.9	0	Mitochondrial ribosomal protein
CG6439	3R	93F13	73.9	+1.9	0	Isocitrate dehydrogenase
CG1471	3R	99F5-6	101.6	−1.9	0	Ceramidase (lipid metabolism)
CG32017	4	102F8	3.1	+9.1	0	Unknown

Band locations and gene functions are according to FlyBase (Crosby *et al.* 2007). Centimorgans (cM) are estimated as described in materials and methods. Fold differences from Massachusetts line means with sign indicating greater expression in high lines (+) or low lines (−). Entries under the column heading "California?" indicate whether the gene was not significant (0) in California lines, was significant in the same direction of expression (+), or was significant in the opposite direction (−). FlyBase reports GH15538 only as a cDNA with an exact match on 2L coding for a putative protein. BLAST searches with the 14 Affymetrix probes for this detected transcript all give unique hits at this site. Jensen *et al.* (2006) report differential expression of GH15538 after insecticide treatment.

we combined the data from all five line pairs for all 7542 probe sets that were called present on at least two of three chips for every line and performed a nested ANOVA on the natural log-transformed intensities. Using a probability threshold corrected conservatively for the number of tests ($P = 0.05/7542$), there were 792 probe sets with significant expression differences, or just >10% of the genes commonly expressed in the wing

disk. Figure 7A shows a "volcano plot" of probability *vs.* fold difference for the pooled Massachusetts data. The solid circles represent the 29 candidate genes from Table 7. Figure 7B shows the same probe sets and probabilities, plotted against log-transformed mean intensities. The candidate genes have reasonably strong intensities and fold differences, with absolute fold differences of 1.5–9.1 and a mean of 2.8. In the ANOVA-based
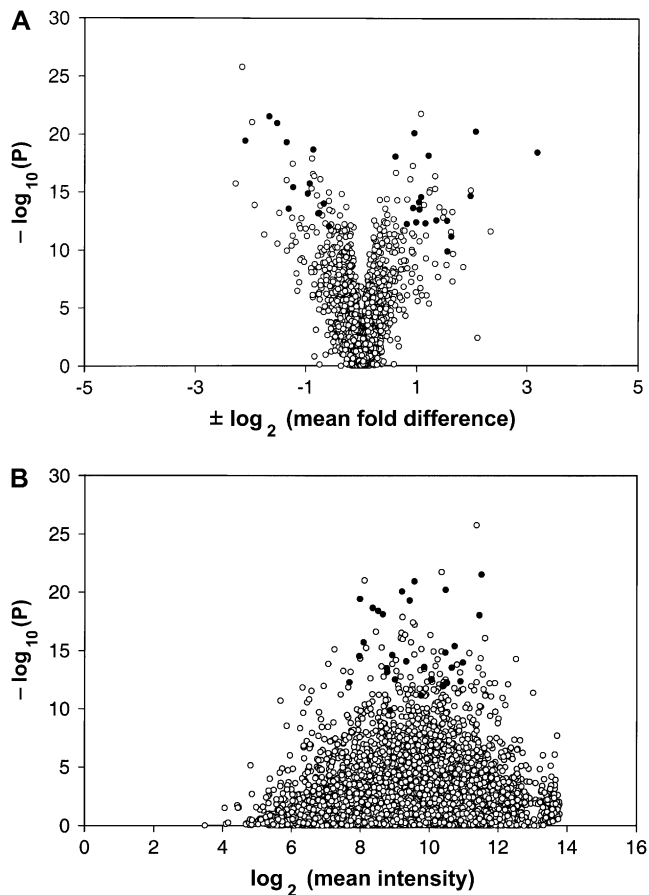
FIGURE 7.—Probability *vs.* fold difference and intensity in combined Massachusetts data. Solid circles show the 29 non-transposon probe sets that were significant in the same direction in all five line pairs (Table 7). (A) Volcano plot for data from all testable probe sets in all five Massachusetts line pairs. *y*-Axis is the −log (base 10) of *P*. *P*-values were calculated by nested ANOVA of log-transformed intensities of high and low lines. *x*-Axis is the ±log (base 2) of average fold differences. Fold differences were calculated by comparing high-line mean intensities (*h*) to low-line mean intensities (*l*). Fold differences with *l* > *h* were quantified as −log (base 2) (*l/h*); those with *h* > *l* were quantified as +log (base 2) (*h/l*). *P*-values were not corrected for the number of tests. (B) Plot of *P* *vs.* mean intensity; same probe sets with same *P*-values. Intensities were calculated as log (base 2) of mean intensity of all present samples for each probe set.

probability ranking of Figure 7, the 29 candidate genes are all highly significant, but some other probe sets look interesting as well. However, our immediate aim was to define a small preliminary set of the highest quality candidates that would be the most likely to show detectable effects in a variety of tests. Our 29 candidate genes represent the intersection of two ranking methods. The first method emphasizes consistent association with the trait, while the second (Figure 7) allows for genes that may be less consistent but show significant overall association with the trait. Each method may have pitfalls, but should be better than the other sometimes. Where both agree, the inference seems especially reliable.

Clustering among candidate genes: Table 7 indicates the probable importance of hitchhiking. Five genes fall in an interval of 0.3 cM on the left end of chromosome three in division 61. Table 7 also includes two triplets and two pairs of closely linked genes (≤0.5 cM). Clusters of linked, coexpressed genes have been found in Drosophila and other organisms (reviewed in HURST *et al.* 2004). However, every cluster in Table 7 includes genes that are expressed oppositely—some more in high lines and some more in low lines. Only two genes in one cluster are side by side (CG9186 and CG2211), but even these two are expressed oppositely with respect to the trait. The others are not adjacent, and most occur in regions of reduced recombination. For example, the 0.3-cM interval with five candidate genes covers 792 kbp containing ∼90 protein-coding genes. Despite their highly significant associations with the trait, some of the 29 candidate genes in tight clusters are likely to represent hitchhiking alleles that do not affect the trait.

We asked whether we would find larger clusters of genes associated with the trait if we relaxed our criteria to include more loci. We looked at the 200 most-significant probe sets in the list of 792 from the ANOVA of the Massachusetts lines, after eliminating all transposons. In addition to the 29 candidate genes, this larger group probably includes all expression differences that contribute to the phenotypic divergence of the Massachusetts lines, and undoubtedly many that do not. Figure 8 shows how these 200 genes are distributed on the five main chromosome arms in bins of width 0.5 cM. For comparison, Figure 8 also shows the distribution of all *D. melanogaster* genes. In regions of low recombination, the density of genes is lower on the physical map (BARTOLOMÉ *et al.* 2002), but higher on the genetic map. The 200 candidate genes are not strongly clustered but are distributed over many sites, often as single, isolated genes. To some extent, their clustering reflects the overall distribution of gene densities.

Comparison of Massachusetts candidate genes to significant California genes: We next compared the 29 final candidate genes from the Massachusetts data to the 450 probe sets that were significant in the single pair of divergent California lines (Table 5). The two groups had only eight genes in common, and five of these were flip-flops—significant in both the California and the Massachusetts lines, but in opposite directions (column 6 in Table 7). Visual inspection of expression profiles in dChip confirmed that the California–Massachusetts flip-flop transcripts all show large expression differences that are associated consistently with the trait in all chips, but in opposite directions in the two populations. Again, various interpretations are possible for these flip-flop genes. They may affect the trait in a context-dependent way that is opposite in the two genetic backgrounds. They may also be cosmopolitan expression polymorphisms that do not affect the trait in either population.
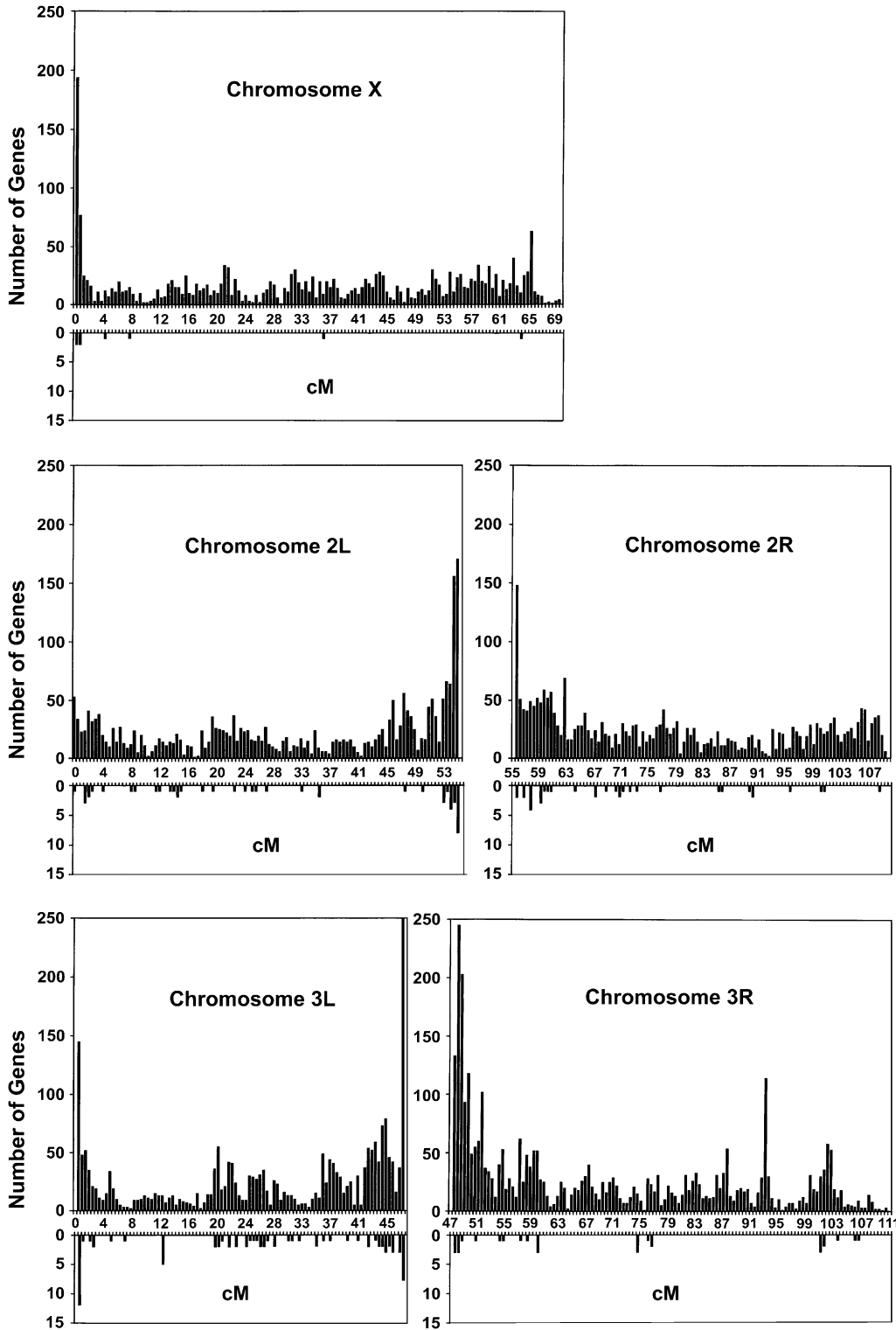
FIGURE 8.—Genetic map locations of the 200 most-significant nontransposon probe sets ranked by ANOVA using combined Massachusetts data. The downward histograms show the distribution of the 200 genes in 0.5-cM bins. The upward histograms show the total numbers of genes in the same bins. Map locations were derived as explained in MATERIALS AND METHODS.

Three of them occur in gene clusters in Table 7, strongly suggesting fixation by hitchhiking in the Massachusetts lines. Overall, the California data give virtually no support to the candidate genes that we would choose by the criteria that we applied to the Massachusetts data. The simplest explanation is that the wing-shape genes segregating in these populations were very different.

Visualizing the differences between gene pools: The Massachusetts and California data sets showed large differences in loci associated with the trait. To study these differences further, we pooled and ranked both data sets separately by t-test probabilities, using the ~7000 probe sets called present on every one of the 36 chips, and then compared the 1000 probe sets with the
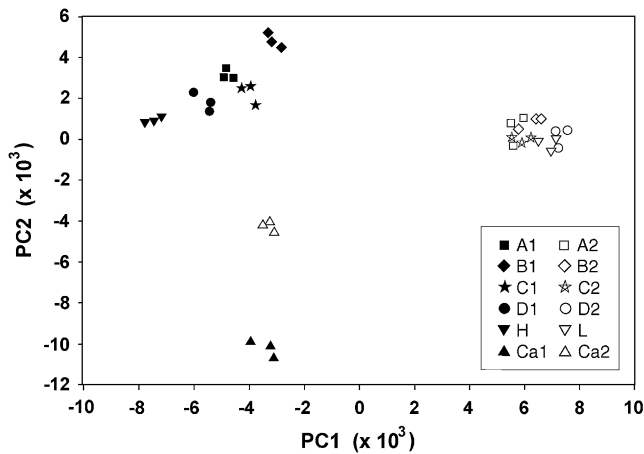
FIGURE 9.—Principal components analysis showing clustering of all 36 chips along the first two components of variance, according to the 1792 probe sets with lowest the *P*-values. Solid symbols, high lines; open symbols, low lines. For each line, three replicate samples (individual chips) are tightly clustered. The five Massachusetts low lines are clustered more closely than the five high lines. Massachusetts and California lines diverge orthogonally for the probe sets associated most closely with the trait by their expression differences.

lowest *P*-values for each population. Of these, only 208 were the same for both populations. (In random draws of 1000 from 7000, 143 would be the same.) We combined these probe sets from both populations for a total of $2000 - 208 = 1792$. The chips from both populations were renormalized as a group in dChip, so that we could perform a principal components analysis (PCA) using these 1792 probe sets, with the probe sets as objects and the 36 chips as variables, to visualize the associations between chips. The first two components of variation explained 49.8% of the variance. In Figure 9, each line is represented by three replicate chips that form a tight cluster. The 10 Massachusetts lines form distinct clusters of high lines and low lines, differentiated along the first principal component, and the high lines are more dispersed than the low lines. The two California lines are differentiated along the second principal component. California and Massachusetts genes clearly show different associations with the trait.

Differences between replicate lines within treatments: The derived Massachusetts line pairs came from repli-

cate crosses of two isogenic lines (H and L), thus creating hybrid lines with identical segregating alleles and identical fixed genetic backgrounds. We asked how much genetic differentiation occurred between them in each direction of selection. These lines are represented by $8 \times 3 = 24$ chips and 7048 probe sets that were present on every chip. For each probe set, we tested the variance in expression among lines over the variance among chips within lines, within each direction of selection, by ANOVA. We found significant variances (*i.e.*, $P \leq 0.05$, without correcting for the total number of ANOVAs) among high-line means for 34.4% of all probe sets and among low-line means for 21.9% of all probe sets. This result agrees with the PCA analysis (Figure 9), showing chips highly clustered within differentiated lines and greater differentiation among high lines than low lines. The percentages seem huge compared to the expectation of 5% of all probe sets. However, they evidently reflect the random fixation of many segregating effects on gene expression that do not affect the trait. Alleles that do not affect a selected trait will often be fixed inconsistently among lines selected in the same direction.

Effective number of contributory loci: We calculated realized heritabilities and effective gene numbers for the four hybrid Massachusetts lines from the selection data. The formula assumes that all allele frequencies were 0.5 when selection began—a reasonable approximation since the lines were hybrids of two isogenic lines and were maintained at large size. The formula also assumes that alleles were fixed in the high and low selected lines, so we used the phenotypes of the selected lines after inbreeding. The estimates of gene number were fairly consistent. The closeness between the mean estimated gene number of Table 8 (26.5) and the number of candidate genes in Table 7 (29) owes much to chance. But the result may indicate that our previous estimate of ~20 detectable genes for this trait, from QTL mapping of the Massachusetts H and L lines, was not too far off (WEBER *et al.* 1999, 2001).

DISCUSSION

We identified candidate genes for a wing-shape trait by microarray expression analysis of divergent selection

TABLE 8

Calculation of effective gene numbers for the derived Massachusetts lines

Line pair	$h^2 \pm$ SE	Response (radians)	S_p	R_p = response/S_p	$n_e = R_p^2/8h^2$
A1/A2	0.59 ± 0.05	0.1376	0.012	11.47	27.9
B1/B2	0.38 ± 0.03	0.1209	0.013	9.30	28.5
C1/C2	0.45 ± 0.10	0.1170	0.013	9.00	22.5
D1D2	0.49 ± 0.05	0.1336	0.013	10.28	27.0

Realized heritability (h^2) was calculated as explained in MATERIALS AND METHODS. "Response" is the phenotypic difference from Table 3. S_p is the square root of mean phenotypic variance in the first generation of selection. R_p is the response in standard deviations. n_e is the effective number of loci.

lines. The lines were replicated within one population and also between two distant locations. The genetic outcomes of selection were variable both within and between gene pools. The results confirm the highly polygenic and degenerate basis of wing shape and shed light on the use of microarrays with selection lines to find quantitative trait genes.

Our data suggest that large population size usefully reduces the number of genes identified as candidates in selected lines by reducing the fixation of alleles with no effect on the trait. The parental lines (H and L) showed 534 significant loci (Table 5), but the hybrid lines derived from them had far fewer (198–335), despite having almost the same phenotypic divergence (Table 3). We attribute this difference to reduced fixation of noncontributory alleles due to recombination in the hybrid lines. This was probably enhanced by their larger size: lines H and L had 40 selected parents per line during selection, whereas the hybrid lines had 126. Selection is more deterministic in larger populations, and more efficient in sorting additive effects. This should reduce variation among outcomes and the number of significant genes.

Resolution was also improved by the replication of selection lines (Figure 6). When tested separately, the individual line pairs all showed unrealistically high numbers of significant probe sets (Table 5). But many probe sets were significant in only one line pair or in several (Table 6). Genes that were significant in some comparisons, but not in all, in some cases would be genes of lesser effect that were not consistently fixed, but many of them must simply be randomly fixed genes with no effect on the trait. Random fixation and hitchhiking are inevitable in finite selection lines. As many as four or five pairs of replicate selection lines may be useful in eliminating these false positives and genes with marginal or inconsistent effects.

A problem with our method of replication may be that the derived Massachusetts replicates represented too few recombinant haplotypes, because of the way in which they were constructed from crosses of isogenic lines. This is almost certainly the reason that we found some tightly linked clusters of genes showing the same repeated association with the trait. Even after 34 generations of hybrid mingling, and even after the trait variance declined to nearly base population levels, the originally isogenic H and L genomes would not dissolve into global linkage equilibrium. Many tight linkages persisted in the same phase in all lines. The same problem will occur whenever replicates come from a restricted source. Residual linkage disequilibrium will preserve numerous blocks of alleles in the same phase that have no effect on the trait, and this will inflate the number of significant genes, even in the combined analysis. Thus replicates for gene hunting should arise from wide sampling within a gene pool. Yet our results also suggest a potential dilemma: if sampling comes from locations that are too far apart, results may be inconsistent because different contributory alleles are present.

Nevertheless, selection lines from distant gene pools may also play a role when looking for genes in highly polygenic traits. We found five genes that were significant in opposite directions in the Massachusetts and California lines. These may be ubiquitous expression polymorphisms, unrelated to the trait. Even after coming up "significant" five times in the Massachusetts lines, an allele might be associated with the trait only because it is closely linked to some other allele (not necessarily an expression polymorphism) and always in the same phase. The California results flag the flip-flop candidate genes as potential false positives, although worthy of investigation.

More interesting than these flip-flops is the finding that virtually no genes with significant expression differences were associated with the trait in the *same* direction of expression in both Massachusetts and California flies. This result is credible because the replicated Massachusetts lines do agree with each other, showing a substantial core group of consistently significant genes. This repeatability proves the reliability of the microarray measurements. Equally convincing is the close agreement between chips within lines, since each chip represents an independent sample of disks. By the same token, the striking disparity between the Massachusetts and California lines—all tested simultaneously—must be a real phenomenon. This result argues that most expression differences with important effects on this trait are different in the two gene pools. Several recent studies support this interpretation.

In a screen of 50 random homozygous-viable *P*-element insertions, 11 (22%) had significant and repeatable effects on wing shape (Weber *et al.* 2005). All 11 were in sites that would be likely to cause expression differences in associated genes. The sample is small, but if it is even approximately representative, then the total number of genes in the genome that can affect wing shape by changes in expression is vastly greater than the number that could be segregating in any local population sample. This implies that, in population samples from widely separated locations, most genes that contribute to selection response for wing-shape traits might be different, as we report here.

QTL analyses of genetic variation for wing shape are also consistent with this conclusion. Wing-shape QTL have now been mapped in three different labs using different genetic stocks and trait definitions (Weber *et al.* 1999, 2001; Zimmerman *et al.* 2000; Mezey *et al.* 2005). A comparison of all three studies concluded that, although each one detected many wing-shape genes, the loci in each study were probably mostly different (Mezey *et al.* 2005).

On the other hand, it is hard to say whether any present candidate genes confirm previous QTL assignments.

Our QTL map for lines H and L (WEBER *et al.* 1999, 2001) was a continuum of broad overlapping probability peaks, with QTL assigned at high points. Some points fall within 0.5 cM of the candidate genes in Table 7, but locations are only approximate, and, in regions of average gene density, 0.5 cM can include 25 genes. Moreover, genes that could affect wing shape are not obvious from their known features (WEBER *et al.* 2005).

The candidate genes that we identified in the Massachusetts lines show a notable lack of coincidence with any of the 213 genes previously associated with non-neural aspects of wing development (BRODY 2005), including genes implicated in regulating wing shape through cell proliferation such as *vestigial* (KIM *et al.* 1996; BAENA-LOPEZ and GARCIA-BELLIDO 2006), *Notch*, or *Delta* (BAONZA and GARCIA-BELLIDO 2000). They include none of the 43 genes in the transforming growth factor-β and epidermal growth factor signaling pathways tested by DWORKIN and GIBSON (2006), the majority of which have small effects on wing shape. Of the 11 genes affecting wing shape in our *P*-element screen (WEBER *et al.* 2005), one (*hephaestus*) coincided with a known wing developmental mutant and one (*stet*) showed shape effects in the survey of DWORKIN and GIBSON (2006).

There could be many reasons for the differences between our results and those of previous mutation studies. Perhaps genes previously identified as major mutants were involved in our experiments but their expression differences were too small to be significant. Or perhaps some classical developmental genes affect the timing or location of transcription without changing mean expression levels in disks in the stage at which we assayed them. It is clear that some morphological traits are influenced by many genes with small effects (NORGA *et al.* 2003; MEZEY *et al.* 2005; WEBER *et al.* 2005), and this fact alone may account for the lack of extensive overlap of wing-shape genes identified by different approaches.

Several of the genes in Table 7 affect functions that could be relevant to wing development, including the well-studied cell death gene *reaper* (WHITE *et al.* 1996), the *CG17090* gene with homology to serine/threonine protein kinases associated with apoptosis, and the Ste20 kinase-like gene *Stlk* associated with cytokinesis in yeast (CVRCKOVA *et al.* 1995). Three other genes are associated with the regulation of gene expression: the RNA polymerase subunit *Rpll18*, the histone gene *His1*, and a putative RNA-binding protein component of the spliceosome *CG4291*. Others with possible developmental relevance include the cytoskeletal actin gene *Act57B* and two genes associated with lipid biosynthesis (*CG31522* and *CG1471*).

There is one striking correspondence between the present results and the 11 genes with shape effects caused by *P*-element insertions in WEBER *et al.* (2005). None of those 11 genes were significant in this study, but 10 of them were among the 6843 transcripts that were called present on all 36 chips. This result is extremely nonrandom. It makes sense because WEBER *et al.* (2005) was a loss-of-function screen, where insertions affecting wing shape should occur in genes that are normally expressed in wing disks. Shape changes may also arise by ectopic expression of genes not normally active in wings. But from this coincidence of results, and other results reported here, it appears that selection has ample scope to act simply by modulating the expression of active genes. These genes are expressed at different intensities in localized regions of wing disks (BUTLER *et al.* 2003), but most are expressed not only in wings but also in all imaginal disks (KLEBES *et al.* 2002).

In our view, selection penetrates an ever-expanding field of potential variation, with incremental improvements of trait alleles (WEBER *et al.* 2005) and progressive transfer of selection to modifier genes (WEBER 1996). This is enhanced by the degenerate and network nature of ways in which genes influence phenotypes (GREENSPAN 2001, 2004; DWORKIN and GIBSON 2006). In some traits, alleles may affect a trait inconsistently in different genetic backgrounds (VAN SWINDEREN and GREENSPAN 2005). In such cases, microarray analysis of replicated selection lines might be used to seek Wrightian peaks in the genetics of quantitative traits. However, the amount of replication and validation that may be required to achieve definitive results is still daunting. On the other hand, if any of our flip-flop genes prove to have opposite effects in different genetic backgrounds, we will have a case worth further study.

Microarray analysis of divergently selected lines is an established method of gene hunting (TOMA *et al.* 2002; MACKAY *et al.* 2005; DIERICK and GREENSPAN 2006; EDWARDS *et al.* 2006). The use of more replicate lines can greatly increase resolution by eliminating many false candidates caused by random fixation and hitchhiking. Replicates should have the same genetic base, but also the likelihood of high recombinational diversity among replicates. Larger populations may also help improve resolution. The consistency of recent results shows that microarray expression measurements are reliable. With further development of methods, microarray testing of replicated selection lines should become increasingly efficient in the search for quantitative trait genes.

LITERATURE CITED

ABZHANOV, A., W. P. KUO, C. HARTMANN, B. R. GRANT, P. R. GRANT *et al.*, 2006 The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature **442:** 563–567.

ANDRES, A. J., and C. S. THUMMEL, 1994 Methods for quantitative analysis of transcription in larvae and prepupae, pp. 565–573 in *Drosophila melanogaster: Practical Uses in Cell and Molecular Biology*

(Methods in Cell Biology, Vol. 44), edited by L. GOLDSTEIN and E. FYRBERG. Academic Press, San Diego.

BAENA-LOPEZ, L. A., and A. GARCIA-BELLIDO, 2006 Control of growth and positional information by the graded vestigial expression pattern in the wing of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **103:** 13734–13739.

BAONZA, A., and A. GARCIA-BELLIDO, 2000 Notch signaling directly controls cell proliferation in the Drosophila wing disc. Proc. Natl. Acad. Sci. USA **97:** 2609–2614.

BARTOLOMÉ, C., X. MASIDE and B. CHARLESWORTH, 2002 On the abundance and distribution of transposable elements in the genome of Drosophila melanogaster. Mol. Biol. Evol. **19:** 926–937.

BELDADE, P., K. KOOPS and P. M. BRAKEFIELD, 2002 Developmental constraints versus flexibility in morphological evolution. Nature **416:** 844–847.

BRODY, T., 2005 *The Interactive Fly.* http://rail.bio.indiana.edu/allied-data/lk/interactive-fly/aimain/1aahome.htm.

BULL, J. J., M. R. BADGETT, H. A. WICHMAN, J. P. HUELSENBECK, D. M. HILLIS *et al.*, 1997 Exceptional convergent evolution in a virus. Genetics **147:** 1497–1507.

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

BUTLER, M. J., T. L. JACOBSEN, D. M. CAIN, M. G. JARMAN, M. HUBANK *et al.*, 2003 Discovery of genes with highly restricted expression patterns in the *Drosophila* wing disc using DNA oligonucleotide microarrays. Development **130:** 659–670.

CARROLL, S. B., 2006 *The Making of the Fittest.* W. W. Norton, New York.

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

CROSBY, M. A., J. L. GOODMAN, V. B. STRELETS, P. ZHANG, W. M. GELBART *et al.*, 2007 FlyBase: genomes by the dozen. Nucleic Acids Res. **35:** D486–D491.

CUNNINGHAM, C. W., K. JENG, J. HUSTI, M. BADGETT, I. J. MOLINEUX *et al.*, 1997 Parallel molecular evolution of deletions and nonsense mutations in bacteriophage T7. Mol. Biol. Evol. **14:** 113–116.

CVRCKOVA, F., C. DE VERGILIO, E. MANSER, J. R. PRINGLE and K. NASMYTH, 1995 Ste20-like protein kinases are required for normal localization of cell growth and for cytokinesis in budding yeast. Genes Dev. **9:** 1817–1830.

DEVOS, K. M., 2005 Updating the "crop circle." Curr. Opin. Plant Biol. **8:** 155–162.

DIERICK, H. A., and R. J. GREENSPAN, 2006 Molecular analysis of flies selected for aggressive behavior. Nat. Genet. **38:** 1023–1031.

DWORKIN, I., and G. GIBSON, 2006 Epidermal growth factor receptor and transforming growth factor-β signaling contributes to variation for wing shape in *Drosophila melanogaster*. Genetics **173:** 1417–1431.

EDWARDS, A. C., S. M. ROLLMANN, T. J. MORGAN and T. F. C. MACKAY, 2006 Quantitative genomics of aggressive behavior in Drosophila melanogaster. PLoS Genet. **2**(9): e154.

FALCONER, D. S., and T. F. C. MACKAY, 1996 *Introduction to Quantitative Genetics.* Longman, Harlow, UK.

GALE, M. D., and K. M. DEVOS, 1998 Plant comparative genetics after 10 years. Science **282:** 656–659.

GARDNER, S. N., J. GRESSEL and M. MANGEL, 1998 A revolving dose strategy to delay the evolution of both quantitative vs. major monogene resistances to pesticides and drugs. Int. J. Pest Manage. **44:** 161–180.

GOMPEL, N., and S. B. CARROLL, 2003 Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. Nature **424:** 931–935.

GOODALE, H. D., 1941 Progress report on possibilities in progeny-test breeding. Science **94:** 442–443.

GOODALE, H. D., 1953 Appendix: third progress report on breeding larger mice, in A study of selection limits in the mouse, D. S. FALCONER and J. W. B. KING. J. Genet. **51:** 561–581.

GREENSPAN, R. J., 2001 The flexible genome. Nat. Rev. Genet. **2:** 383–387.

GREENSPAN, R. J., 2004 E pluribus unum, ex uno plura: quantitative- and single-gene perspectives on the study of behavior. Annu. Rev. Neurosci. **27:** 79–105.

GROMKO, M. H., 1995 Unpredictability of correlated responses to selection: pleiotropy and selection interact. Evolution **49:** 685–693.

HILL, W. G., 1972 Estimation of realized heritabilities from selection experiments. II. Selection in one direction. Biometrics **28:** 767–780.

HOULE-LEROY, P., H. GUDERLEY, J. G. SWALLOW and T. GARLAND, JR., 2003 Artificial selection for high activity favors mighty mini-muscles in house mice. Am. J. Physiol. Regul. Integr. Comp. Physiol. **284:** 433–443.

HURST, L. D., C. PÁL and M. J. LERCHER, 2004 The evolutionary dynamics of eukaryotic gene order. Nat. Rev. Genet. **5:** 299–310.

JENSEN, H. R., I. M. SCOTT, S. R. SIMS, V. L. TRUDEAU and J. T. ARNASON, 2006 The effect of a synergistic concentration of a *Piper nigrum* extract used in conjunction with pyrethrum upon gene expression in *Drosophila melanogaster*. Insect Mol. Biol. **15:** 329–339.

KIM, J., A. SEBRING, J. J. ESCH, M. E. KRAUS, K. VORWERK *et al.*, 1996 Integration of positional signals and regulation of wing formation and identity by *Drosophila vestigial* gene. Nature **382:** 133–138.

KLEBES, A., B. BIEHS, F. CIFUENTES, and T. B. KORNBERG, 2002 Expression profiling of *Drosophila* imaginal discs. Genome Biol. **3**(8): RESEARCH0038.1–0038.16

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

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

LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster.* Academic Press, San Diego.

MACARTHUR, J. W., 1949 Selection for small and large body size in the house mouse. Genetics **34:** 194–209.

MACKAY, T. F. C., S. L. HEINSOHN, R. F. LYMAN, A. J. MOEHRING, T. J. MORGAN *et al.*, 2005 Genetics and genomics of *Drosophila* mating behavior. Proc. Natl. Acad. Sci. USA **102:** 6622–6629.

MARONI, G., and S. C. STAMEY, 1983 Use of blue food to select synchronous, late third-instar larvae. Dros. Inf. Serv. **59:** 142–143.

MCKENZIE, J. A., A. G. PARKER and J. L. YEN, 1992 Polygenic and single gene responses to selection for resistance to diazinon in *Lucilia cuprina*. Genetics **130:** 613–620.

MEZEY, J. G., and D. HOULE, 2005 The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. Evolution **59:** 1027–1038.

MEZEY, J. G., D. HOULE and S. V. NUZHDIN, 2005 Naturally segregating quantitative trait loci affecting wing shape of *Drosophila melanogaster*. Genetics **169:** 2101–2113.

MORRELL, P. L., and M. T. CLEGG, 2007 Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proc. Natl. Acad. Sci. USA **104:** 3289–3294.

NEVE, P., and S. POWLES, 2005 Recurrent selection with reduced herbicide rates results in the rapid evolution of herbicide resistance in *Lolium rigidum*. Theor. Appl. Genet. **110:** 1154–1166.

NORGA, K. K., M. C. GURGANUS, C. L. DILDA, A. YAMAMOTO, R. F. LYMAN *et al.*, 2003 Quantitative analysis of bristle number in Drosophila mutants identifies genes involved in neural development. Curr. Biol. **13:** 1388–1396.

OLEKSIAK, M. F., G. A. CHURCHILL and D. L. CRAWFORD, 2002 Variation in gene expression within and among natural populations. Nat. Genet. **32:** 261–266.

PATERSON, A. H., Y.-R. LIN, Z. LI, K. F. SCHERTZ, J. F. DOEBLEY *et al.*, 1995 Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science **269:** 1714–1718.

PATTHY, L., 1999 *Protein Evolution.* Blackwell Science, Oxford.

PRUD'HOMME, B., N. GOMPEL, A. ROKAS, V. A. KASSNER, T. M. WILLIAMS *et al.*, 2006 Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature **440:** 1050–1053.

SCHLUTER, D., E. A. CLIFFORD, M. NEMETHY and J. S. MCKINNON, 2004 Parallel evolution and inheritance of quantitative traits. Am. Nat. **163:** 809–822.

SHAPIRO, M. D., M. E. MARKS, C. L. PEICHEL, B. K. BLACKMAN, K. S. NERENG *et al.*, 2004 Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature **428:** 717–723.

SWALLOW, J. G., P. A. CARTER and T. GARLAND, JR., 1998 Artificial selection for increased wheel-running behavior in house mice. Behav. Genet. **28:** 227–237.

Toma, D. P., K. P. White, J. Hirsch and R. J. Greenspan, 2002 Identification of genes involved in *Drosophila melanogaster* geotaxis, a complex behavioral trait. Nat. Genet. **31:** 349–353.

Van Swinderen, B., and R. J. Greenspan, 2005 Flexibility in a gene network affecting a simple behavior in *Drosophila melanogaster*. Genetics **169:** 2151–2163.

Weber, K. E., 1990 Selection on wing allometry in *Drosophila melanogaster*. Genetics **126:** 975–989.

Weber, K. E., 1992 How small are the smallest selectable domains of form? Genetics **130:** 345–353.

Weber, K. E., 1996 Large genetic change at small fitness cost in large populations of *Drosophila melanogaster* selected for wind-tunnel flight: rethinking fitness surfaces. Genetics **144:** 205–213.

Weber, K., R. Eisman, L. Morey, A. Patty, J. Sparks *et al.*, 1999 An analysis of polygenes affecting wing shape on chromosome 3 in *Drosophila melanogaster*. Genetics **153:** 773–786.

Weber, K., R. Eisman, S. Higgins, L. Morey, A. Patty *et al.*, 2001 An analysis of polygenes affecting wing shape on chro-

mosome 2 in *Drosophila melanogaster*. Genetics **159:** 1045–1057.

Weber, K., N. Johnson, D. Champlin and A. Patty, 2005 Many P-element insertions affect wing shape in *Drosophila melanogaster*. Genetics **169:** 1461–1475.

White, K., E. Tahaoglu and H. Steller, 1996 Cell killing by the Drosophila gene *reaper*. Science **271:** 805–807.

Wichman, H. A., M. R. Badgett, L. A. Scott, C. M. Boulianne and J. J. Bull, 1999 Different trajectories of parallel evolution during viral adaptation. Science **285:** 422–424.

Wood, T. E., J. M. Burke and L. H. Rieseberg, 2005 Parallel genotypic adaptation: when evolution repeats itself. Genetica **123:** 157–170.

Zimmerman, E., A. Palsson and G. Gibson, 2000 Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. Genetics **155:** 671–683.

Communicating editor: L. Harshman