DOI: 10.1534/genetics.103.021766

Association Between Nucleotide Variation in *Egfr* and Wing Shape in *Drosophila melanogaster*

Arnar Palsson[1] and Greg Gibson[2]

Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695

Manuscript received August 29, 2003
Accepted for publication December 31, 2003

ABSTRACT

As part of an effort to dissect quantitative trait locus effects to the nucleotide level, association was assessed between 238 single-nucleotide and 20 indel polymorphisms spread over 11 kb of the *Drosophila melanogaster Egfr* locus and nine relative warp measures of wing shape. One SNP in a conserved potential regulatory site for a GAGA factor in the promoter of alternate first exon 2 approaches conservative experimentwise significance ($P < 0.00003$) in the sample of 207 lines for association with the location of the crossveins in the central region of the wing. Several other sites indicate marginal association with one or more other aspects of shape. No strong effects of sex or population of origin were detected with measures of shape, but two different sites were strongly associated with overall wing size in interaction with these fixed factors. Whole-gene sequencing in very large samples, rather than selective genotyping, would appear to be the only strategy likely to be successful for detecting subtle associations in species with high polymorphism and little haplotype structure. However, these features severely limit the ability of linkage disequilibrium mapping in Drosophila to resolve quantitative effects to single nucleotides.

THREE of the major challenges in efforts to dissect complex traits to the nucleotide level relate to disentangling of population stratification, linkage disequilibrium, and the correlated effects of underlying genetic mechanisms. To date, association of morphological with genotypic variation has been investigated largely in relation to simple traits, namely those that are measured by a single unambiguous descriptor, such as enzyme activity, bristle number, or longevity (Laurie *et al.* 1991; Lai *et al.* 1994; Long *et al.* 1998, 2000; Lyman and Mackay 1998; Robin *et al.* 2002; de Luca *et al.* 2003). Compound traits are those for which the whole is assembled from multiple parts whose individual contributions are not straightforward to identify, such as shape or psychological disease (Atchley and Hall 1991; Peralta *et al.* 1997; Shastry 1999). They present novel challenges for association mapping.

The shape of the *Drosophila melanogaster* wing is an example of a complex morphological trait that has proven difficult to describe (Whitlock and Fowler 1999; Gilchrist *et al.* 2000; Klingenberg and Zaklan 2000). No matter how it is measured, it has consistently been shown to have high heritability due to a large number of genetic factors of small effect (Weber *et al.* 1999, 2001; Zimmerman *et al.* 2000). We have advocated the use of geometric morphometric techniques for description of orthogonal components of variation, as these capture the major components of variation and are independent of observer bias (Bookstein 1991; Rohlf 1996; Dryden and Mardia 1998). Relative warp analysis (Rohlf 1993) of the shapes of intervein regions defined by landmarks at the junctions between veins and the wing margin shows that shape is stable to size and environmental variation (Birdsall *et al.* 2000). This observation suggests that the placement and length of the veins is a major determinant of shape and hence that genes affecting venation are candidate modifiers of wing shape.

Consistent with this hypothesis, *EGF Receptor* has been implicated as a candidate gene for wing shape by two quantitative approaches. Quantitative trait locus (QTL) mapping in two different crosses segregating variation affecting different regions of the wing identified >30 significant QTL intervals, and although resolution of this approach was too low to implicate individual genes, a significant excess of peaks was found in the vicinity of loci involved in epidermal growth factor receptor (EGFR) signaling (Zimmerman *et al.* 2000). One of these peaks covered the *Egfr* locus itself, and follow-up quantitative complementation mapping was consistent with the possible contribution of segregating variation in the locus to wing shape variation (Palsson and Gibson 2000). A large deletion, *Df(2R)Pu-D17*, that also eliminates function of three other potential regulators of EGFR signaling and vein development (*MESK-2*, *PTP-ER*, and *cv-2*), showed interactions with wild-type alleles in all three intervein regions, while the null point mutation

[1]*Present address:* Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637.

[2]*Corresponding author:* Department of Genetics, Gardner Hall, North Carolina State University, Raleigh, NC 27695-7614.
E-mail: ggibson@unity.ncsu.edu

Egfr^{f2} specifically failed to quantitatively complement shape specification in the central and anterior wing.

These results encouraged us to initiate a large association study designed first to confirm that variation in *Egfr* modifies wing shape and second to attempt to describe the distribution of allelic effects of single-nucleotide polymorphisms (SNPs). Association mapping generally relies on linkage disequilibrium among SNPs to focus on a region of a locus that may be responsible for an effect. Choice of a subset of SNPs reduces both the cost of genotyping and the statistical challenges due to multiple comparisons (Robin *et al.* 2002; de Luca *et al.* 2003). However, this is a far from comprehensive approach, particularly in Drosophila where linkage disequilibrium typically decays within a kilobase (Aquadro *et al.* 2001) and is heavily biased toward detecting only major-effect common polymorphisms.

Consequently, we chose to sequence almost 11 kb of the *Egfr* in a sample of 207 chromosomes and to test for associations with each of the 238 independently segregating SNPs and 20 indels for which both alleles were at a frequency of at least 5%. Although statistical power can be increased substantially by isolating single chromosomes in a common background, this approach can be biased by the presence of deleterious recessives and by nonadditive interactions with the common background. On the other hand, assessment of associations in outbred individuals is compromised by the large contribution of nongenetic variation, so the intermediate approach adopted here was to measure nearly isogenic lines generated by at least 15 generations of inbreeding. Measurement of multiple individuals of each line minimizes environmental variation, effectively enhancing the genetic effects. The 207 lines were derived from two different populations, from North Carolina and California, affording us the opportunity to examine whether population stratification might affect any observed associations (Pritchard and Rosenberg 1999; Thornsberry *et al.* 2001; Ardlie *et al.* 2002). Although sexual dimorphism for wing shape is limited (Birdsall *et al.* 2000), subtle quantitative contributions of SNPs might be sex biased, so both sexes were also included in the analysis. The results provide evidence for association of a handful of SNPs with different aspects of wing shape, but promote the more cautionary conclusion that resolution of the relative contributions of alleles of large and small effect to complex traits remains beyond the reach of empirical detection.

MATERIALS AND METHODS

Fly culture: All of the wild-type lines of *Drosophila melanogaster* used in this study are near isogenic lines derived by between 15 and 50 generations of sib-pair mating of derivatives of isofemales collected in West End, North Carolina (NC; Palsson *et al.* 2004, accompanying article, this issue), and Winter, California (CA; Yang and Nuzhdin 2003). Homozygosity at *Egfr* was determined by sequencing of a single fly and confirmed for a second individual in a subset, but there may be some genetic variation segregating at *Egfr* in some lines, and in general heterozygosity remains up to 10% throughout the genome of the NC lines in particular, as these were less inbred. All lines were maintained on 12-hr light-dark cycle in vials with 10 ml cornmeal medium supplemented with yeast.

Morphometrics: One wing was dissected from each of 10 flies of each sex from each of three independent vials and mounted carefully under a glass coverslip on a microscope slide. The image was captured using Scion Image Version 4.2 and recorded to a compact disc. The *x-y* coordinates of the nine landmarks were captured and subjected to thin plate spline analysis using the TPS-Relw Version 1.2 package downloaded from http://life.bio.sunysb.edu/morph (Rohlf 1996). The first step of this procedure is a general Procrustes transformation that scales the wings to a unit-squared distance of landmarks from the centroid and rotates pairs of wings iteratively to minimize the sum of the squared distances between equivalent landmarks (Dryden and Mardia 1998). Subsequently, the major axes of variation are captured as relative warps, which are essentially equivalent to principal components (Rohlf 1999), and provide orthogonal descriptors of some fraction of the variation. Two sets of relative warps were obtained, one involving all nine landmarks and the other three sets of four, five, and four landmarks for intervein region B (IVR-B), IVR-C, and IVR-D, respectively (see Figure 1 in Birdsall *et al.* 2000). The shape measures obtained by TPS analysis are unaffected by observer bias (other than landmark selection) and typically capture aspects of variation that can be ascribed *a posteriori* to one or a few regions of the shape, identified in Figure 1. Nine trait measures were extracted from both analyses, W1–W9 and B1–B3, C1–C3, and D1–D3, the key features of which are discussed in the text. For association studies, sex-within-line values for association tests were computed as the arithmetic mean relative warp scores for the 30 individuals, and these values are approximately normally distributed as described in Palsson (2003).

Robustness of the trait measures was tested by monitoring the sensitivity of relative warp scores to modification of the size of the data set. The Pearson correlation between line means by sex in Table 1 contrasts the complete data set involving 126 NC lines and 81 CA lines with scores from just 85 NC lines. The effects of sex, population, and line on each phenotype were assessed by mixed-model analysis of variance using the DIFFS options in PROC MIXED in SAS Version 8.2 (SAS Institute, Cary, NC) with population (P) and sex (S) as fixed effects and line (L) and replicate vial (R) nested random factors in the model

$$\text{Relative warp score} = \mu + P + S + P \times S + L(P) + S \times L(P) + R(LP) + S \times R(LP) + \varepsilon,$$

where μ is the grand mean and ε is the normally distributed error. Results are tabulated in the supplementary information at http://statgen.ncsu.edu/ggibson/SupplInfo/SupplInfo7.htm.

Heritability and genetic correlations between warps were assessed using variance component estimates obtained with PROC VARCOMP in SAS. Lines were assumed to be nearly isogenic, so heritability was estimated for each sex separately as $V_G/2$ divided by $(V_P + V_G/2)$. Genetic correlations between traits were estimated as the covariance of the among-line variance component divided by the square root of the product of variance components (Falconer and Mackay 1996, p. 313).

Association tests: Genotypes were obtained by direct sequencing of PCR products as described in the accompanying article (Palsson *et al.* 2004, this issue) and are also available along with phenotype scores as supplementary information

(http://statgen.ncsu.edu/ggibson/SupplInfo/SupplInfo7.htm). Effects of genotype (G) and interactions with sex and population were assessed with PROC MIXED for each SNP according to the model

$$\text{Relative warp score} = \mu + G + P + S + P \times S + G \times P + G \times S + G \times P \times S + L + \varepsilon,$$

where all terms are fixed except for the line. Note that line is fit to control for the pseudo-replication due to measurement in two sexes, and while it is nested within population and genotype, this nesting does not affect the estimates of interest and so was not included here. Failure to include line results in dramatic inflation of P-values due to pseudo-replication of the sexes, as monitored by permutation of the lines within sex and population (analysis not shown). Inclusion of the population interaction terms controls for possible effect of population structure.

Adjustment for multiple comparisons was performed using the standard Bonferroni correction by dividing the testwise α of 0.05 by the number of tests (2142 = 238 SNPs and nine traits per analysis) to give the experimentwise threshold of 0.000023 (the negative logarithm of which is 4.6). However, since linkage disequilibrium is typically significant between up to five adjacent sites, this threshold is almost certainly too conservative, and a fourfold less stringent threshold of 0.0001 (neglogp = 4.0) is also considered. Since it is unlikely that the statistical measures of shape capture axes of variation that are directly affected by *Egfr*, finding associations with correlated warps from two different modes of analysis might be considered as corroborative evidence that a SNP affects some aspect of shape that is only partially captured by the two warps. That is, correlated trait measures may exhibit "morphological disequilibrium" with one another in the same sense that physically adjacent SNPs are in linkage disequilibrium (LD) with one another. However, since the two sets of relative warps combined capture most of the wing-shape variation, they are not independent. This is confirmed by the observation that summation of the logarithms of the P-values of each set of nine warps results in very similar "multitrait" association plots that will be described elsewhere. In our judgment, this nonindependence justifies separate statistical analysis of the two types of shape measure and correction for 9 traits rather than 18.

Multivariate analysis of variance (MANOVA) was also performed so as to allow employment of a single test statistic per site, by fitting a linear model on the six pairs of partial warps and the uniform component of variation derived from the nine *x-y* landmarks (Rohlf *et al.* 1996). Compared with univariate analysis of each relative warp, this strategy requires an order of magnitude of less formal significance, but since no sites were significant even at $P < 0.001$, the results are presented only in supplementary Table 2 at http://statgen.ncsu.edu/ggibson/SupplInfo/SupplInfo7.htm. Various other modes of MANOVA were also considered, but their general failure to identify candidate sites suggests that any effects of *Egfr* are restricted to single components of wing-shape variation.

RESULTS

Relative warp descriptors of wing shape: Since our hypothesis is that *Egfr* affects wing shape through its effect on vein differentiation, our analysis of shape focuses on the nine landmarks defined by the junctions of the internal longitudinal veins 2, 3, 4, and 5 with the anterior and posterior crossveins (acv and pcv) and the wing margin. The *x-y* coordinates of each landmark were captured from single wings of 30 flies of each sex of the 207 nearly isogenic lines and subjected to Procrustes superimposition, which reduces size differences and rotates shapes to optimize their alignment (Rohlf 1996). Shape variance was reduced to its major components by two modes of relative warp analysis, one dividing the wing into three intervein regions (IVRs B, C, and D) corresponding to anterior, central, and posterior regions, respectively, and the other treating the whole wing as a single shape. The advantages of the former are that the warps explain more of the variance and are more robust than the whole-shape warps, are easier to interpret, and are more likely to capture variation due to region-specific gene action. The advantages of the latter are that the warps are by definition orthogonal, and this approach implicitly captures any genetic integration of morphogenesis across the whole wing.

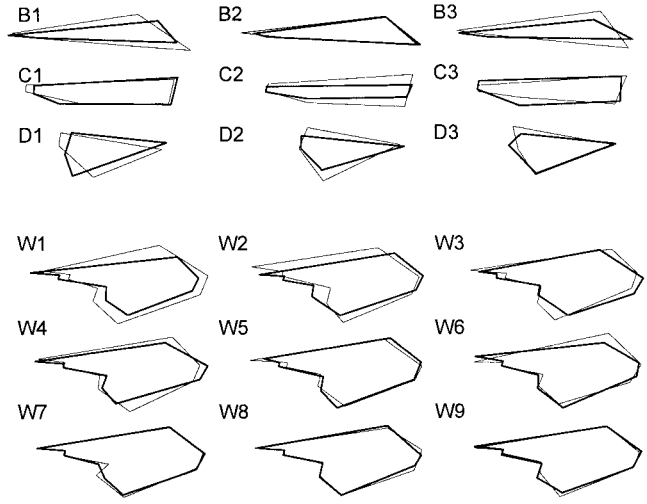

Figure 1.—Aspects of wing-shape variation captured by the 18 relative warp measures. Each outline shows a reconstituted wing outline from the average of the landmarks of the five lines with the most extreme values of the particular relative warp. The nine plots at the top show the IVR traits in the anterior, central, and posterior regions of the wing, while those at the bottom show the whole-wing traits.

Nine warps were captured for each mode. The aspect of morphology captured by each warp is shown in Figure 1, and parameters of variation are indicated in Table 1. Hatcher's rules (Hatcher 1994) suggest retention of all nine IVR warps (three per IVR), since all associated eigenvalues are >1 and each warp explains >10% of the variation for its IVR. In the anterior IVR, B1 captures the width of the region, B2 the placement of the acv, and B3 a composite of all landmarks. In the central IVR, C1 captures the distance between the crossveins, C2 the width of the region, and C3 the taper of the region. The shape of the posterior IVR is more complex, but broadly D1 captures the length of longitudinal vein 5, D2 the width of the region, and D3 the placement of the pcv. For the whole-wing warps, retention is justified

TABLE 1

Relative warp parameters

Warp	Eigenvalue	PVE	F_{pop}	Est	SE	r	h^2
B1	3.30	59.4	0.16	0.0008	0.0019	0.999	0.51
B2	2.10	24.1	8.93**	0.0036	0.0012	0.991	0.46
B3	1.59	13.8	0.05	0.0002	0.0009	0.992	0.46
C1	2.74	58.3	4.07*	0.0036	0.0159	0.999	0.50
C2	1.65	21.1	0.34	0.0005	0.0080	0.982	0.45
C3	1.15	10.3	0.03	0.0000	0.0061	0.967	0.34
D1	4.46	57.5	0.19	0.0011	0.0026	1.000	0.49
D2	2.59	19.3	0.90	0.0015	0.0015	0.991	0.43
D3	2.28	15.0	0.05	0.0003	0.0013	0.986	0.34
W1	3.53	32.7	1.20	0.0023	0.0021	0.986	0.61
W2	2.85	21.3	7.03**	0.0047	0.0018	0.974	0.53
W3	2.33	14.2	0.13	0.0005	0.0013	0.969	0.46
W4	1.87	9.2	0.16	0.0005	0.0011	0.957	0.49
W5	1.65	7.2	7.17**	0.0026	0.0010	0.961	0.46
W6	1.48	5.7	1.86	0.0012	0.0008	0.982	0.40
W7	1.04	2.8	0.00	0.0000	0.0006	0.989	0.28
W8	0.85	1.9	0.00	0.0000	0.0005	0.953	0.39
W9	0.76	1.5	0.03	0.0001	0.0004	0.951	0.37

PVE, percentage of variance explained; F_{pop}, F-ratio for effect of population on warp score with associated significance (*$0.01 < P < 0.05$; **$0.0001 < P < 0.001$); Est and SE, estimate of difference between and standard error of population means; r, correlation between data set for associations and the larger data set; h^2, heritability averages across sexes and populations.

strongly for W1–W3 and more weakly for W4–W7, but we also include W8 and W9 because they both show high heritability, W8 captures some of the variation seen by B3, and W9 shows an unexpected pattern of association with nucleotide polymorphism. The stronger warps W1–W4 capture broad aspects of wing shape relating to length and width, whereas W5, W6, W7, and W9 appear to be most influenced by local aspects of shape relating to placement of the crossveins and L5.

For the purposes of association mapping, it is preferable that the shape measures be both uncorrelated and robust. As expected, none of the whole-wing warps are significantly correlated with one another (Table 2). Some correlation between IVR warps is expected because adjacent IVRs share veins and landmarks, but only three significant similarities were detected at both the phenotypic and genotypic levels: B2 with D1 and D2, and B3 with D2. Since IVR-B and IVR-D do not share landmarks, this observation provides evidence for morphological integration across the wing (Klingenberg and Zaklan 2000), but it should be noted that the effect is weak and most of the variation is IVR specific. Whole-wing and IVR warps clearly capture some of the same underlying genetic factors, since a complex correlation matrix is indicated between the two modes in Table 2. Overall, we regard the IVR warps as better descriptors of shape, but both modes are considered side by side as a check for the internal consistency of the association studies as outlined in materials and methods.

The robustness of the relative warp measures was confirmed by examining the effect of altering the size of the data set. The penultimate column in Table 1 indicates that all warps are very highly correlated between the full data set consisting of 12,530 wings and one about half this size consisting of just 70% of the North Carolinian sample. This suggests both that the main axes of variation are shared between the populations and that the warp measures are not overly sensitive to sample size. All warps also have relatively high heritability, as shown in the final column of Table 1. Heritability was consistently several percent lower in the North Carolinian than in the Californian sample, possibly reflecting more complete inbreeding in the latter sample of lines. Furthermore, good congruence with $r > 0.95$ for all warps was observed when the data set was tripled by the addition of three more data sets to be described elsewhere, including a Kenyan sample, a set of crosses among lines, and testcrosses to 15 heterozygous wing mutants. It is noteworthy that two gain- and loss-of-function alleles of *Egfr* had opposite overall effects on several measures of wing shape and that the IVR measures were more sensitive to perturbation by mutants (Palsson 2003). Loss of Drosophila EGF receptor (DER) activity tends to reduce the width of the wing and lessen the spacing between the crossveins.

Association between *Egfr* alleles and wing shape: Association between allelic variation in *Egfr* and each measure of wing shape was assessed by mixed-model analysis of variance with genotype, sex, and population as fixed effects and line as a random effect, for each of 238 SNPs for which both biallelic sites had a frequency of at least 5% in the sample of 207 lines. Table 3 lists the strongest

TABLE 2

Correlation matrix for relative warps

	B1	B2	B3	C1	C2	C3	D1	D2	D3	W1	W2	W3	W4	W5	W6	W7	W8	W9	TA
B1	—				*					***	*	*							*
B2		—				*	*	*			*	**		***					*
B3	*		—			*		*					**		**		*		*
C1	*			—			*	*		**	**			*		*			
C2	*	*			—			*		*	*		*	*	*		**		
C3		*	*			—	*				*	*			**				*
D1		*		*	*	*	—				*	**	**	*					*
D2		*	*	**	*	*		—			**		**			*			
D3									—	*	*		*			***			
W1	***	*		***	*	*			*	—									*
W2	**	*		***	**	*	**	***	*		—								
W3	**	***	*	*	**	**	***	*				—							*
W4	*		***	*	*	*	***	***	*				—						
W5	*	***		*	*		**		*					—					
W6		***	*	*	***										—				*
W7				**				***	***							—			
W8			***		***												—		
W9					*			*	*									—	
TA	*				*				*										—

Phenotypic correlation is above the diagonal and male genetic correlation below for both populations combined. *$0.05 < r^2 <$ 0.25; **$0.25 < r^2 < 0.50$; ***$0.50 < r^2 < 1.00$. Note the absence of phenotypic correlation between relative warps of the same trait, although these sometimes show genetic correlation.

association observed for each of the 18 warps as well as the total area of the wing, and Figure 2 plots significance of the genotype effect against location in the gene for traits C1, D1, W1, and W9. At least one SNP would be expected to associate by chance with each trait with a *P*-value of 0.004 or lower, but even that level of stringency is not met by 5 of the IVR warps and 2 of the whole-wing warps, indicating that *Egfr* does not dramatically affect wing shape in a global manner. Treating each mode of analysis separately, 2142 SNP tests were performed, so a stringent significance threshold of 2.3×10^{-5} would provide experiment-wide confidence in any particular association. For shape, this condition was almost met by just a single site, 30200, that lies in the promoter 5′ to the alternate first exon 2 and affects trait C1. Notably this site was scored in only half of the lines due to difficulties in amplifying DNA from the region, and it disrupts an unusual highly conserved GAGA factor binding motif (Palsson *et al.* 2004). Applying a less conservative correction that accounts for the nonindependence of tests with adjacent sites requires that $P < 0.0001$ (Figure 2, dashed line) and suggests marginally significant associations with two further sites, 39389, a silent substitution in exon 4 that may affect D1, and 30505, which is just 13 bases in front of the translation start site of exon 2 and may affect W7.

Each of these associations would explain in excess of 5% of the genetic variance in the relevant aspects of wing shape, notwithstanding Beavis effects. The study design does not have good power to detect effects of this magnitude, so it is possible that other true associations have not been detected. Two hints that this may be the case are suggested by careful examination of Table 3. First, three sites show the strongest association with traits in both modes of analysis. Site 39389 mentioned above has the strongest association with W4, the whole-wing trait most closely correlated to warp D1. Similarly, site 30505 gives the strongest association with D2, but in this case while W7 is correlated with D2, it is more strongly predictive of D3 yet site 30505 is only very weakly associated with D3 ($P = 0.042$). Site 40110 is not formally significant for any trait, but it provides the strongest test statistic for both W2 and C3, even though these two traits are uncorrelated. Moreover, site 40119, another silent substitution just 9 bases away in exon 6 and in quite strong linkage disequilibrium with site 40110, shows the best association with B2. Second, three of the other traits associate most strongly with relatively rare alleles: D3 with a promoter allele near alternate exon 1 at a frequency of 0.12 in a sample of just 140 chromosomes, W1 with a silent substitution in exon 4 at a frequency of 0.14, and W6 with a site in the conserved portion of intron 2 at a frequency of 0.06 that actually shows the largest standardized effect of any of the SNPs. Further work will be required to confirm the significance of any of these SNPs, but taken as a whole there is good evidence that one or a few sites in the *Egfr* locus impact wing shape. Supplementary Table 3 (http://statgen.ncsu.edu/ggibson/SupplInfo/SupplInfo7.htm) lists the site that has the strongest magnitude of effect for each trait, and all but one of these are relatively rare alleles, many of which lie within intron 2.

TABLE 3

Most significant genotype associations by trait

Trait	Site	Location	*P*-value	Estimate	SDU	Allele 1	Allele 2
B1	38581	Exon 4	0.01121	0.00626	0.46	58 A	64 C
B2	40119	Exon 6	0.00766	0.00354	0.41	111 T	80 G
B3	37498	Intron 2	0.01275	0.00300	0.47	130 A	42 T
C1	30200	5′ to exon 2	2.7E-05**	0.01037	0.82	95 T	33 C
C2	39160	Exon 4	0.00117	0.00303	0.49	137 T	67 C
C3	40110	Exon 6	0.00900	0.00166	0.38	74 C	122 T
D1	39389	Exon 5	6.1E-05*	0.01298	0.69	54 C	126 T
D2	6412	3′ to exon 1	0.00589	0.00496	0.45	59 T	134 C
	30505	5′ to exon 2	0.00603	0.00751	0.68	78 C	24 A
D3	5917	5′ to exon 1	0.00405	0.00839	0.89	123 C	17 A
W1	38914	Exon 4	0.00044	0.01066	0.72	181 C	29 T
W2	40110	Exon 6	0.00332	0.00559	0.43	122 T	74 C
W3	39160	Exon 4	0.00155	0.00431	0.48	67 C	137 T
W4	39389	Exon 5	0.00567	0.00404	0.50	54 C	126 T
W5	41925	Exon 6	0.02437	0.00259	0.36	49 G	158 C
W6	36248	Intron 2	0.00221	0.00560	0.92	12 G	173 A
W7	30505	5′ to exon 2	0.00016	0.00360	0.90	24 A	78 C
W8	42043	3′ UTR	0.00076	0.00275	0.81	23 T	179 A
W9	38022	Intron 3	0.00047	0.00159	0.52	105 T	89 C
TA	41712	Exon 6	0.00999	0.42850	0.14	120 C	90 A

P-value is for the genotype term after fitting the full model with sex and population. Estimate and SDU are estimate in relative warp units and as a fraction of the sex-averaged standard deviation units of the trait line means. Alleles 1 and 2 are arbitrarily defined to give a positive difference in warp least-squares means. TA is total area of the wing, a representative size measure.

Among the replacement substitutions segregating in the population, only site 6065 (Ser17Ile in exon 1) shows any putative significance. It is associated with D1 at $P = 0.0022$ and the correlated trait W3 at $P = 0.0102$, as well as with W9 at $P = 0.0070$. Intriguingly, this site also showed a strong interaction with the ability to enhance the dominant gain-of-function $Egfr^{Ellipse}$ phenotype in photoreceptor determination. Most of the other replacement polymorphisms are rare, and molecular population genetic analysis suggests that the protein, with the exception of exon 1, experiences strong purifying selection (Palsson *et al.* 2004, this issue). As a class, these sites are marginally significant for association with traits B2/W3 and W9 ($P \sim 0.03$) and hence with the location of the anterior crossvein.

Sex and population dependence of associations: Aside from the main effect of genotype, another way to evaluate the significance of associations is to assess their robustness across sexes and populations. Although no significant association was observed between *Egfr* and wing size, the plots in Figure 3 suggest that there may be a significant interaction between genotype and sex or population in relation to the total area of the wing. These effects are the most formally significant in the entire analysis, but the effects of sex, population, and sex-by-population interaction regardless of genotype were much larger for size than for shape, and permutations of the data result in associations at this level more often than would be expected by chance (data not shown). Since size is much more susceptible to environmental variation than is shape, this association is much less likely than the shape associations to be robust and indeed fails to replicate in preliminary follow-up crosses (our unpublished data; Palsson 2003). For the shape measures, similarity between the sexes occurs despite the fact that several traits show overall sex effects, but also reflects the contribution of shared genetic backgrounds within lines to the correlation between sexes. Population effects are likely to be due to a combination of sampling biases, overall differences between populations due to genetic factors other than those at *Egfr*, and population-by-genotype interactions at each SNP.

Due to the enormous number of contrasts involved in the full models for 238 SNPs and 18 traits, it is almost impossible to disentangle these factors across the experiment. The ANOVAs in Table 4 indicate the absence of interaction effects for the sites that show the strongest overall genotype effects. The strongest population-by-genotype interactions were actually observed for sites 40344 and 42788 with D2 ($P = 0.00067$ and 0.00078) and 42383 with C1 ($P = 0.00087$), but this level of effect is expected by chance. No associations at the $P < 0.001$ level were seen for the sex-by-genotype interaction, while two closely linked sites in exon 3 showed suggestive associations with B2 (site 37878, $P = 0.00064$) and C2 (site 37959, $P = 0.00085$) for the three-way genotype-by-sex-by-population interaction.

Locus-wide and haplotype associations: An alternative

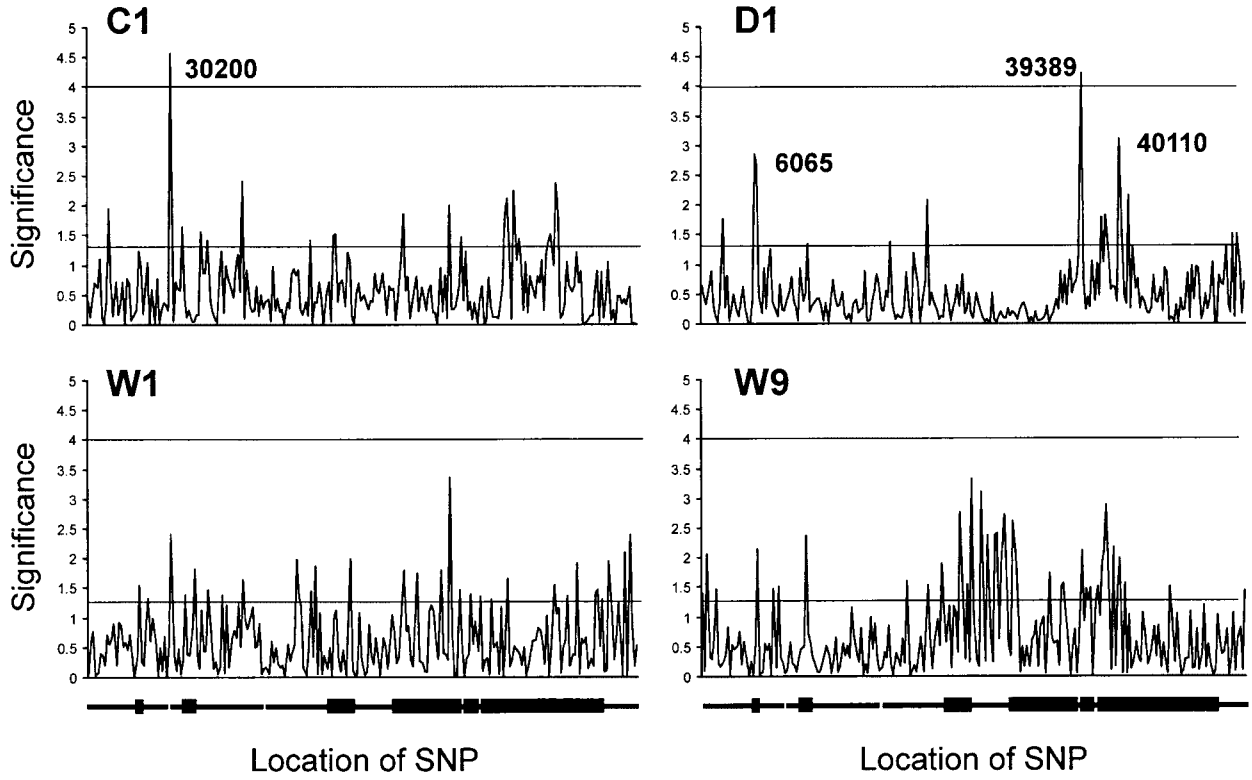

FIGURE 2.—Significance of relative warps along the *Egfr* locus. Each plot shows the negative logarithm of the *P*-value associated with the genotype term from the mixed-model ANOVA for the indicated trait for the 238 common SNPs aligned from 5′ to 3′. Boxes below the plots indicate the extent of exons in three contiguous blocks of sequence (see PALSSON *et al.* 2004, this issue, for details of SNP location). Bottom line indicates testwise $\alpha = 0.05$ cutoff, whereas the line at neglogp = 4.0 indicates an approximate Bonferroni threshold adjusted for nonindependence of adjacent sites. Plots show association with traits C1, D1, W1, and W9.

approach to association mapping that does not attempt to identify the discrete sites that may cause a phenotype difference is to ask whether more sites than expected by chance are significant at a prechosen threshold (LAI *et al.* 1994). Two whole-wing warps have a great excess of sites significant at the 0.05 level as shown in Figure 2: W1 and W9. However, the variance of this statistic is greatly inflated by even small levels of linkage disequilibrium such as those detected in *Egfr*. Whereas if 238 genotypes are randomly generated independent of one another, the number of sites significant at the 0.05 level almost always lies between 8 and 16, with the *Egfr* data this number ranges up to 21 for more than a quarter of the traits, while W1 and W9 show 30 and 39 significant sites, respectively. However, values up to 30 are seen in 20% of permutations of the *Egfr* genotype and phenotype matrices, holding sexes within lines and lines within populations constant. The reason is that if a chance strong association happens to fall in a haplotype block then several SNPs will exceed the low threshold, as appears to be the case for W9 where most of the weak associations fall in a large block of LD near exon 3. Consequently, even these extreme test values provide only weak evidence for a contribution of *Egfr* to wing shape.

Suggestive SNP associations with disease susceptibility are now routinely followed by haplotype tests (*e.g.*, MOHAMED *et al.* 2003; NORTH *et al.* 2003; ZHANG *et al.* 2003). The rationale for this is that where genotyping is incomplete, a cluster of linked sites are more likely to capture the variation associated with the true but untyped causal SNP. In some cases, the haplotype might also sum weak effects of two or more sites to give a larger overall effect, particularly if epistatic interactions are present (DE LUCA *et al.* 2003). There are too many sites and traits to test systematically for haplotype effects here, but no significant epistatic interactions were detected between the three most significant sites and the next two most significant sites for their respective traits.

However, one intriguing result is that linkage disequilibrium exists between sites that have the strongest association with D1, namely 39389, 40110, and 6065 (all of which were mentioned in relation to other traits as well). Although the former two sites are within 1 kb of one another, they show considerably less LD with 12 common polymorphisms between them. Site 6065 is >35 kb distant, and the LD is only weakly significant since the more rare T allele has a frequency of only 0.1. Nevertheless, in each case the pairwise linkage disequilibria result in an excess of alleles with extreme values

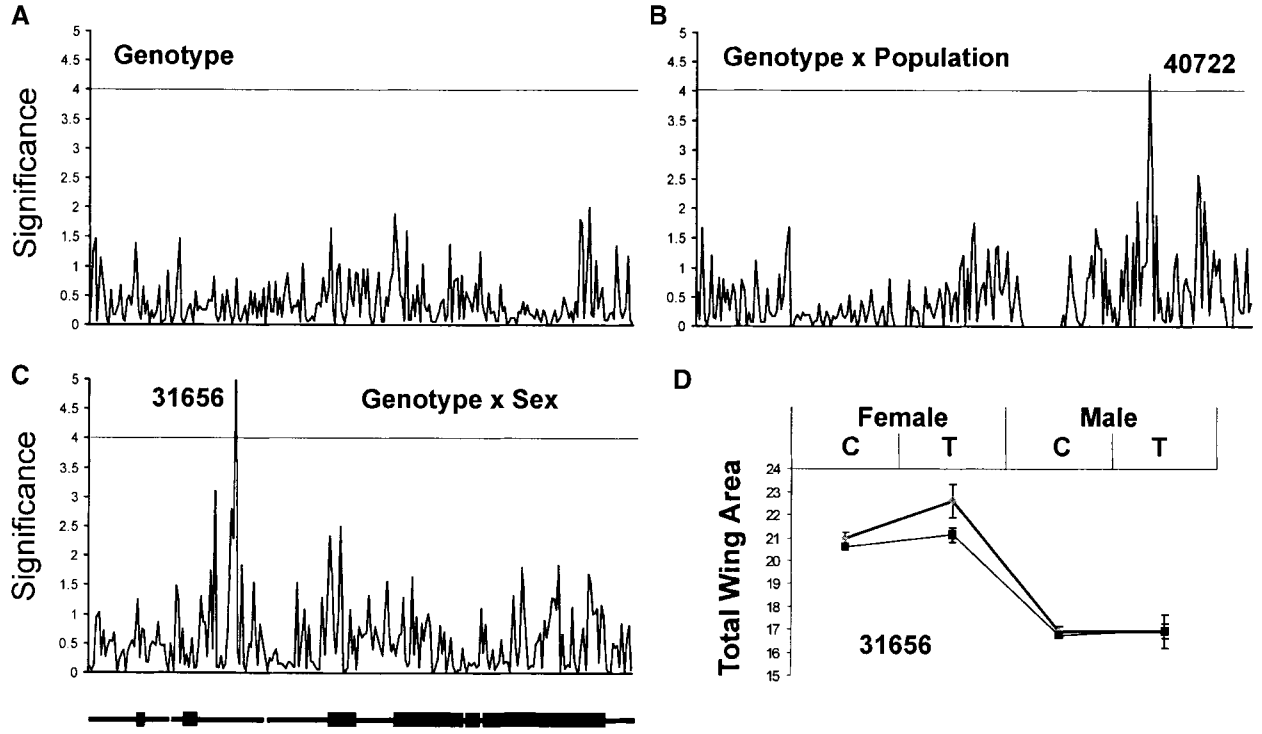

FIGURE 3.—Genotype interaction effects with sex and population for total area of the wing. (A–C) Similar plots as in Figure 2, but showing significance of the association with wing size for the indicated interaction effects. (D) Plot of genotype and sex means for total area (in arbitrary units) plus or minus one standard error unit for site T31656C (thick line, CA; thin line, NC). Total area is highly correlated with other measures of size such as the area of each IVR, or wing length.

for warp D1. Differences in D1 score are highly significant for both the two-site and three-site haplotypes ($P < 0.000002$ and $P = 0.00002$, respectively), and the effects of each SNP are nearly additive such that the two most extreme haplotypes differ by more than a full standard deviation unit of the line means for the trait.

DISCUSSION

Does polymorphism in the *Egfr* locus affect wing shape? This study represents the largest investigation of genotype-phenotype association yet reported in a model organism, in terms of the combination of depth of sequence coverage, sample size of chromosomes, and inclusion of two populations. Developmental genetic arguments (BIEHS *et al.* 1998), documentation of the effect of loss- and gain-of-function *Egfr* alleles on wing patterning (CLIFFORD and SCHÜPBACH 1994; MARTIN-BLANCO *et al.* 1999; WANG *et al.* 2000; CROZATIER *et al.* 2002; ZECCA and STRUHL 2002), QTL mapping, and deficiency complementation tests all implicate *Egfr* as an excellent candidate locus for modification of wing shape in Drosophila (PALSSON and GIBSON 2000; ZIMMERMAN *et al.* 2000). Molecular population genetic analysis demonstrates that purifying selection is quite strong throughout the locus, but variable haplotype structure in particular short regions of the gene is consistent with the operation of weak positive selection as well (PALSSON *et al.* 2004, this issue). Nevertheless, no blazing signal of comprehensive association with multiple aspects of wing shape was detected, and the conclusion must be that any contributions of the sequenced portion of *Egfr* to the trait are subtle and likely to be due to multiple polymorphisms of weak effect within the gene.

The strongest evidence for association comes from a T-to-C substitution at a frequency of 0.26 in a putative regulatory element 341 bp upstream of the start codon of the second alternate first exon. As shown in Figure 4, this element consists of two runs of >12 CN repeats in a 100-bp sequence fragment that is highly conserved in *D. pseudoobscura*, which diverged from *D. melanogaster* 45 million years ago (POWELL 1996), and a similar motif in the bithoraxoid region is known to be required for regulation of transcription by a GAGA factor (HODGSON *et al.* 2001). The association is with the shape of the central intervein region, and the less common C allele decreases the spacing between the anterior and posterior crossveins, consistent with slight loss of function of EGFR signaling. It was detected predominantly in the North Carolina sample, owing to incomplete sequencing of the fragment in the California sample, but the polymorphism is at a similar frequency in a set of 30 Kenyan alleles and so has been maintained in the species for some time. Three other putative associations were detected with polymorphisms flanking the two first exons, all with aspects of shape in the posterior intervein

TABLE 4

Analysis of variance of representative associations

Trait	Site	Source	*F*	*P*	Trait	Site	Source	*F*	*P*
C1	C30200T	Genotype	19.06	0.00003	D1	T39389C	Genotype	16.93	0.00006
		Population	5.58	0.020			Population	0.12	0.72
		Sex	17.38	0.00006			Sex	343.56	0.00000
		$G \times P$	0.28	0.60			$G \times P$	3.31	0.071
		$G \times S$	0.22	0.64			$G \times S$	0.06	0.80
		$P \times S$	0.98	0.32			$P \times S$	2.00	0.16
		$G \times P \times S$	0.16	0.69			$G \times P \times S$	0.00	0.97
Area	T31656C	Genotype	2.06	0.15	Area	T40722C	Genotype	0.15	0.70
		Population	1.45	0.23			Population	14.65	0.00018
		Sex	2044.02	0.00000			Sex	6287.36	0.00000
		$G \times P$	0.31	0.58			$G \times P$	16.97	0.00006
		$G \times S$	25.59	0.000002			$G \times S$	0.12	0.73
		$P \times S$	18.72	0.00004			$P \times S$	33.23	0.00000
		$G \times P \times S$	11.28	0.0011			$G \times P \times S$	4.49	0.035
C2	C31634T	Genotype	0.03	0.86					
		Population	0.24	0.62					
		Sex	317.09	0.00000					
		$G \times P$	3.31	0.071					
		$G \times S$	8.94	0.0037					
		$P \times S$	10.20	0.0020					
		$G \times P \times S$	18.35	0.00005					

region. All other trait measures were most strongly associated with synonymous substitutions in the long exons that encode the bulk of the protein, including one site that is marginally significant experiment-wide, 39389, potentially with the length of longitudinal vein 5 in the posterior of the wing.

There are a variety of mechanisms by which wing shape could affect fitness and by which synonymous polymorphisms could affect wing shape. Placement of the wing veins is thought to impart rigidity to the wing, which is important for aerodynamic performance (Dickenson *et al.* 1999; Dudley 2000) and might also affect other wing functions such as the mating dance. Wing size and shape clines are well documented in Drosophila (Imasheva *et al.* 1995; Gilchrist and Partridge 1999; Huey *et al.* 2000) and appear to be under strong selection, and while shape is quite stable to size variation, subtle variation in venation might interact with size to affect performance. Intriguingly, the first two relative warps for IVR-C are significantly correlated with two aspects of take-off velocity in a panel of recombinant inbred lines (K. Montooth and A. G. Clark, personal communication). Presumably, venation can be quantitatively affected by the amount of DER (EGFR) protein in the vein primordium, as a result of either differential transcription or translation. Codon bias can affect protein levels (Carlini and Stephan 2003), but it might also affect the fidelity of amino acid incorporation. In a parallel study of association between *Egfr* polymorphism and eye development, we found strong evidence for an effect of a cluster of synonymous sites in exon 6 on photoreceptor determination (Dworkin *et al.* 2003). These sites are not implicated in wing shape here (and nor do the putative wing sites affect the eye), which may indicate some tissue specificity to codon usage effects or could be attributed to low statistical power.

Experimental design issues: Our experimental design differs from similar published work with Drosophila in three major features: the type of inbreeding, inclusion of two populations, and comprehensive nature of the

```
                                                                  *
GenBank TCCTACACTCACGCACACACTCACTCACT - 27bp LINKER - CACACACTCACTCTCACGCGCATACACA
NC001   ............................          .      ............................ 70%
CA120   ............................          .      .......................C..... 26%
NC004   ............................          .      ....A....................... 3%
CA034   ......................TC....          .      ............................ 1%
Sim     A..CT...A.G.A..... - 27bp linker - .......A.T.A...G......C.....
```

Figure 4.—Conservation and polymorphism in the CN repeat. The sequences of the 86-bp bipartite CN motif upstream of exon 2 in five representative alleles at the indicated frequencies in *D. melanogaster* and one from *D. simulans* are shown. Dots represent conserved sites. Underlined sites are conserved in *D. pseudoobscura* as well: note that almost all of the C residues are conserved, but the N residues vary. The linker is just 7 bp long in *D. pseudoobscura*. The asterisk identifies the C/T polymorphism (site 30200) in *D. melanogaster* that is associated with wing-shape variation.

genotyping. The major attraction of Drosophila for these experiments is the ability to manipulate the genetics (Mackay 2001), principally through inbreeding that allows multiple measurements to be made per genotype and thereby reduces the environmental error. Mackay and co-workers have demonstrated the gains to be had by taking this a step further, by substituting whole chromosomes into a common background, so that the contribution of the candidate gene need be a significant proportion of only one-third of the genetic variation (Long *et al.* 1998; Robin *et al.* 2002). Unfortunately, in a pilot experiment we encountered several drawbacks to this approach. Sequencing detected an appreciable level of probable gene conversion between the wild-type alleles at *Egfr* and the balancer chromosome, approaching 10%. Since most natural chromosomes harbor lethals, an unbiased approach is to retain them over the balancer and take measurements in a heterozygous cross—but this strategy surrenders any gain in power if site effects are recessive. Chromosome extraction is also laborious, expensive, and error prone, and while it would be useful for the community to develop a large panel of such lines (Gibson and Mackay 2002), in the mean time simple inbreeding to homogenize whole genomes is a viable alternative that also ensures that site effects are averaged over the effects of other modifiers.

Population stratification is likely to be a concern only where two populations differ for the trait or for the frequency of a particular allele (Pritchard and Rosenberg 1999). The NC and CA populations do show different phenotype distributions (Palsson 2003), presumably due to undocumented genetic differences. Since more marginally significant genotype-by-population interactions than genotype-by-sex interactions were detected in our study, despite the fact that sex has at least a 10 times larger fixed effect than population, interactions between *Egfr* polymorphisms and the genetic background may affect the association studies. However, only a handful of sites differ in frequency between NC and CA (Palsson *et al.* 2004, this issue), and none of these showed significant contributions to any trait. A more pressing concern is whether each of the two main populations are themselves produced by admixture between local populations, perhaps including inversion polymorphism types. Even if only a few percent of all SNPs differ in frequency between populations, the total amount of genetic differentiation could be of the same order of magnitude as that observed in humans, for example. It thus seems prudent to include population as a factor in Drosophila quantitative genetics, at least in the initial phase of a study restricting the survey to two or three localities.

The most important difference in our study is the decision to completely sequence all of the alleles, rather than to genotype just the most common SNPs identified in 12 sequences. This decision was made because haplotype blocks are so short in Drosophila, rarely more than five sites long, and polymorphism is so high, that linkage disequilibrium mapping is generally unlikely to detect weak associations. The disadvantage is that the multiple-comparison problem becomes severe, but the complexity of wing shape and length of the locus already assures statistical uncertainty. Our approach is validated by two considerations. Most importantly, neither of the two formally significant sites would have been scored if we had chosen sites to genotype on the basis of their frequency in a small sample, and similarly several of the other weaker, not formally significant, associations were detected with relatively rare alleles. Moreover, if the hypothetical experiment of assessing only the 33 most common polymorphisms spaced by at least 100-bp intervals is performed, no meaningful associations are detected: only the expected 15 of 297 contrasts are significant at $P = 0.05$ for the IVR warps, and the strongest association is 0.007, well short of the 0.0001 cutoff for nine warps or even for just the three strongest warps. We conclude that even though linkage-disequilibrium mapping has produced significant results in other studies in flies (Lyman *et al.* 1999; Long *et al.* 2000), success is far from guaranteed.

The genetic architecture of QTL: A more critical interpretation of this study would call into question the reality of the identified quantitative trait nucleotide (QTN) effects and, even if they are real, their relevance to evolutionary biology. The experimentwise adjustment for multiple comparisons is appropriate for testing the hypothesis that variation in the candidate gene affects the trait, but actually builds on the prior assumption that the gene is more likely than other genes to make a quantitative contribution. There are almost certainly >2 million common SNPs in the fly genome on the basis of observed nucleotide diversity, so 200 associations with each trait are expected genome-wide at the 0.0001 level by chance (and 20 at the 0.00001 level, even though none were observed in *Egfr*). In a sample of 200 chromosomes, each of these 200 *prima facie* significant associations would appear to explain >5% of the genetic variation. So, in a complete genome scan the ratio of false positives to true positives will be at least 10 to 1 and false positives are expected one in every 75 genes per trait (or one in every 10 genes for any of the wing-shape traits). Had we sampled one of the other loci in *Df(2R)Pu-D17* in the vicinity of the *Egfr* QTL peak and observed a similar association, we would be less likely to believe that the result represented a true association. Even though our two most significant associations are among the strongest yet reported in Drosophila, the results must be regarded as suggestive only. Experimental verification is unlikely to have sufficient resolution to detect such a small effect, so further replication is required to confirm the result.

The even more pertinent question is whether detection of one or two major-effect QTNs addresses the fundamental issue of whether genetic variation is pre-

dominantly contributed by a few sites of large effect or many with small effects. Although the reality of major-effect QTL has been emphasized recently, QTL are slowly being eroded by fine-structure mapping that can resolve single into multiple peaks (Pasyukova *et al.* 2000; Steinmetz *et al.* 2002) and by the realization that Beavis effects lead to overestimation of contributions in initial studies (Ioannidis *et al.* 2001). It would take sample sizes of 10,000 fully sequenced alleles to confidently identify QTNs that account for as little as 0.5% of the variation for a trait, but 100 such sites spread over multiple loci could explain half of the genetic variation for any trait. Given the high level of recombination in Drosophila, loci such as *Egfr* are made up of numerous small haplotype blocks in linkage equilibrium. If several of these harbor small-effect QTN, it is inevitable that some alleles in the population would differ by several percent in their contribution to a trait. This presents a conundrum in the resolution of QTL to individual genes and nucleotides, because there is simply no power to test the alternate hypothesis that numerous sites of small effect add up to produce QTL haplotypes.

We thank Ian Dworkin, Naruo Nikoh, Lisa Goering, and Ed Buckler for discussions. A.P. was funded in part by awards from the American Scandinavian Foundation and NATO, and the project was funded by grants to G.G. from the David and Lucille Packard Foundation and the National Institutes of Health (R01 GM61600).

LITERATURE CITED

Aquadro, C. F., D. Bauer and F. Reed, 2001 Genome-wide variation in the human and fruitfly: a comparison. Curr. Opin. Genet. Dev. **11:** 627–634.

Ardlie, K. G., K. Lunetta and M. Seielstad, 2002 Testing for population subdivision and association in four case-control studies. Am. J. Hum. Genet. **71:** 304–311.

Atchley, W. R., and B. K. Hall, 1991 A model for development and evolution of complex morphological structures. Biol. Rev. **66:** 101–157.

Biehs, B., M. Sturtevant and E. Bier, 1998 Boundaries in the Drosophila wing imaginal disc organize vein-specific genetic programs. Development **125:** 4245–4257.

Birdsall, K., E. Zimmerman, K. Teeter and G. Gibson, 2000 Genetic variation for the positioning of wing veins in *Drosophila melanogaster*. Evol. Dev. **2:** 16–24.

Bookstein, F. L., 1991 *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge, UK.

Carlini, D. B., and W. Stephan, 2003 *In vivo* introduction of unpreferred synonymous codons into the Drosophila *Adh* gene results in reduced levels of ADH protein. Genetics **163:** 239–243.

Clifford, R., and T. Schüpbach, 1994 Molecular analysis of the Drosophila EGF receptor homolog reveals that several genetically defined classes of alleles cluster in subdomains of the receptor protein. Genetics **137:** 531–550.

Crozatier, M., B. Glise and A. Vincent, 2002 Connecting Hh, Dpp and EGF signalling in patterning of the Drosophila wing: the pivotal role of collier/knot in the AP organiser. Development **129:** 4261–4269.

De Luca, M., N. Roshina, G. Geiger-Thornsberry, R. Lyman, E. Pasyukova *et al.*, 2003 Dopa decarboxylase (Ddc) affects variation in Drosophila longevity. Nat. Genet. **34:** 429–433.

Dickenson, M. H., F. Lehmann and S. Sane, 1999 Wing rotation and the aerodynamic basis of insect flight. Science **284:** 1954–1960.

Dryden, I. L., and K. V. Mardia, 1998 *Statistical Shape Analysis*. John Wiley & Sons, Chichester, UK/New York.

Dudley, R., 2000 *The Biomechanics of Insect Flight*. Princeton University Press, Princeton, NJ.

Dworkin, I., A. Palsson, K. Birdsall and G. Gibson, 2003 Evidence that *Egfr* contributes to cryptic genetic variation for photoreceptor determination in natural populations of *Drosophila melanogaster*. Curr. Biol. **13:** 1888–1893.

Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to Quantitative Genetics*, Ed. 4. Longman Group, Essex, UK.

Gibson, G., and T. F. C. Mackay, 2002 Enabling population genomics. Genet. Res. **80:** 1–6.

Gilchrist, A. S., and L. Partridge, 1999 A comparison of the genetic basis of wing size divergence in three parallel body size clines of *Drosophila melanogaster*. Genetics **153:** 1775–1787.

Gilchrist, A. S., R. B. Azevedo, L. Partridge and P. O'Higgins, 2000 Adaptation and constraint in the evolution of Drosophila melanogaster wing shape. Evol. Dev. **2:** 114–124.

Hatcher, P. J., 1994 *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling*. SAS Institute, Cary, NC.

Hodgson, J. W., B. Argiropoulos and H. W. Brock, 2001 Site-specific recognition of a 70-base-pair element containing d(GA)n repeats mediates *bithoraxoid* Polycomb group response element-dependent silencing. Mol. Cell Biol. **21:** 4528–4543.

Huey, R. B., G. W. Gilchrist, M. Carlson, D. Berrigan and L. Serra, 2000 Rapid evolution of a geographic cline in size in an introduced fly. Science **287:** 308–309.

Imasheva, A. G., O. Bubli, O. Lazenby and L. A. Zhivotovsky, 1995 Geographic differentiation in wing shape in *Drosophila melanogaster*. Genetica **96:** 303–306.

Ioannidis, J. P. A., E. Ntzani, T. Trikalinos and D. Contopoulos-Ioannidis, 2001 Replication validity of genetic association studies. Nat. Genet. **29:** 306–309.

Klingenberg, C. P., and S. D. Zaklan, 2000 Morphological integration between developmental compartments and the Drosophila wing. Evolution **54:** 1273–1285.

Lai, C., R. Lyman, A. D. Long, C. H. Langley and T. F. C. Mackay, 1994 Naturally occurring variation in bristle number and DNA polymorphisms at the *scabrous* locus of *Drosophila melanogaster*. Science **266:** 1697–1702.

Laurie, C. C., J. Bridgeham and M. Choudhary, 1991 Association between DNA sequence variation and variation in expression of the *Adh* gene in natural populations of *Drosophila melanogaster*. Genetics **129:** 489–499.

Long, A. D., R. F. Lyman, C. H. Langley and T. F. C. Mackay, 1998 Two sites in the *Delta* gene region conotribute to naturally occurring variation in bristle number in *Drosophila melanogaster*. Genetics **149:** 999–1017.

Long, A. D., R. F. Lyman, A. Morgan, C. H. Langley and T. F. C. Mackay, 2000 Both naturally occurring insertions of transposable elements and intermediate frequency polymorphisms in the *achaete-scute* complex are associated with variation in bristle number in *Drosophila melanogaster*. Genetics **154:** 1255–1269.

Lyman, R. F., and T. F. C. Mackay, 1998 Candidate quantitative trait loci and naturally occurring phenotypic variation for bristle number in *Drosophila melanogaster*: the *Delta-Hairless* gene region. Genetics **149:** 983–998.

Lyman, R. F., C. Lai and T. F. C. Mackay, 1999 Linkage disequilibrium mapping of molecular polymorphisms at the scabrous locus associated with naturally occurring variation in bristle number in *Drosophila melanogaster*. Genet. Res. **74:** 303–311.

Mackay, T. F. C., 2001 The genetic architecture of quantitative traits. Annu. Rev. Genet. **35:** 303–339.

Martin-Blanco, E., F. Roch, E. Noll, A. Baonza, J. Duffy *et al.*, 1999 A temporal switch in DER signalling controls the specification and differentiation of veins and interveins in the Drosophila wing. Development **126:** 5739–5747.

Mohamed, H. S., M. Ibrahim, E. Miller, C. Peacock, E. Khalil *et al.*, 2003 Genetic susceptibility to visceral leishmaniasis in The Sudan: linkage and association with IL4 and IFNGR1. Genes Immun. **4:** 351–355.

North, B. V., D. Curtis, P. Cassell, G. Hitman and P. C. Sham, 2003 Assessing optimal neural network architecture for identifying disease-associated multi-marker genotypes using a permutation test, and application to calpain 10 polymorphisms associated with diabetes. Ann. Hum. Genet. **67:** 348–356.

Palsson, A., 2003 Molecular quantitative genetics of wing shape

in Drosophila melanogaster. Ph.D. Thesis, North Carolina State University, Raleigh, NC.

Palsson, A., and G. Gibson, 2000 Quantitative developmental genetic analysis reveals that the ancestral dipteran wing vein prepattern is conserved in *Drosophila melanogaster*. Dev. Genes Evol. **210:** 617–622.

Palsson, A., A. Rouse, R. Riley-Berger, I. Dworkin and G. Gibson, 2004 Nucleotide variation in the *Egfr* locus of *Drosophila melanogaster*. Genetics **167:** 1199–1212.

Pasyukova, E. G., C. Vieira and T. F. C. Mackay, 2000 Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. Genetics **156:** 1129–1146.

Peralta, V., M. J. Cuesta and C. Farre, 1997 Factor structure of symptoms in functional psychoses. Biol. Psychiatry **42:** 806–815.

Powell, J. R., 1996 *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York.

Pritchard, J. K., and N. A. Rosenberg, 1999 Use of unlinked genetic markers to detect population stratification in association studies. Am. J. Hum. Genet. **65:** 220–228.

Robin, C., R. F. Lyman, A. D. Long, C. H. Langley and T. F. C. Mackay, 2002 *hairy*: a quantitative trait locus for Drosophila sensory bristle number. Genetics **162:** 155–164.

Rohlf, J. F., 1993 Relative warp analysis and an example of its application to mosquito wings, pp. 131–159 in *Contributions to Morphometrics*, Vol. 8, edited by L. F. Marcus, L. Bello and A. Garcia-Valdecasas. Monografias del Museo de Sciencias Naturales de Madrid, Madrid.

Rohlf, J. F., 1996 Morphometric spaces, shape components, and the effects of linear transformations, pp. 117–129 in *NATO Advances in Morphometrics*, edited by L. F. Marcus. Plenum, New York.

Rohlf, J. F., 1999 Shape statistics: Procrustes superimposition and tangent spaces. J. Classification **16:** 197–223.

Rohlf, J. F., A. Loy and M. Corti, 1996 Morphometric analysis of old world Talpidae (Mammalia, Insectivora) using partial warp scores. Syst. Biol. **45:** 344–362.

Shastry, B. S., 1999 Recent developments in the genetics of schizophrenia. Neurogenetics **2:** 149–154.

Steinmetz, L., H. Sinha, D. Richards, J. Spiegelman, P. Oefner *et al.*, 2002 Dissecting the architecture of a quantitative trait locus in yeast. Nature **416:** 326–330.

Thornsberry, J., M. Goodman, J. Doebley, S. Kresovich, D. Nielsen *et al.*, 2001 *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. **28:** 286–289.

Wang, S. H., A. Simcox and G. Campbell, 2000 Dual role for Drosophila epidermal growth factor receptor signaling in early wing disc development. Genes Dev. **14:** 2271–2276.

Weber, K. E., 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. E., R. Eisman, S. Higgins, L. Morey, A. Patty *et al.*, 2001 An analysis of polygenes affecting wing shape on chromosome 2 in *Drosophila melanogaster*. Genetics **159:** 1045–1057.

Whitlock, M. C., and K. Fowler, 1999 The changes in genetic and environmental variance with inbreeding in *Drosophila melanogaster*. Genetics **152:** 345–353.

Yang, H. P., and S. V. Nuzhdin, 2003 Fitness costs of Doc expression are insufficient to stabilize its copy number in *Drosophila melanogaster*. Mol. Biol. Evol. **20:** 800–804.

Zecca, M., and G. Struhl, 2002 Subdivision of the Drosophila wing imaginal disc by EGFR-mediated signaling. Development **129:** 1357–1368.

Zhang, S., Q. Sha, H. Chen, J. Dong and R. Jiang, 2003 Transmission/disequilibrium test based on haplotype sharing for tightly linked markers. Am. J. Hum. Genet. **73:** 566–579.

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