Evidence That *Egfr* Contributes to Cryptic Genetic Variation for Photoreceptor Determination in Natural Populations of *Drosophila melanogaster*

Ian Dworkin, Arnar Palsson, Kelli Birdsall, and Greg Gibson

Supplemental Experimental Procedures

Genetics

The inbred lines were from two North American populations of *D. melanogaster*: a panel of 90 inbred lines derived from a vineyard near Davis, California, in 1996 (CA lines, courtesy of S. Nuzhdin), and a second set of 120 inbred lines established from a peach orchard near West End, North Carolina, in 2000 (NC lines). Single gravid females were collected from these populations, and their offspring were sib mated for 15–20 (NC) or 50 generations (CA). This inbreeding program was sufficient to generate isogenicity at *Egfr* for greater than 95% of the lines as determined by comprehensive sequence analysis (A.P., A. Rouse, R. Riley-Berger, I.D., and G.G., unpublished data). The remaining individual heterozygous sites were scored as missing data.

Sensitization crosses were performed using a gain-of-function allele of *Egfr* termed *Ellipse*^{1} (*Egfr*^{E1}) that is caused by an alanine to threonine amino acid replacement, A887T [S1]. In order to control for genetic background effects, the experiment was conducted with two strains harboring the *Egfr*^{E1} allele with the following genotypes: BL-1564, *nw*^{D} *Pu*^{2} *Egfr*^{E1} *Pin*^{Yt}/SM1; and BL-5144, P{w$^{+mW.hs}$ = GawB}*elav*^{C155} *w** P{ry$^{+t7.2}$ = neoFRT}19A; *Bc*^{1} *Egfr*^{E1}/CyO. Both strains were obtained from the Bloomington stock center and inbred by three generations of sib mating immediately prior to the experiment. Two virgin females of each of the *Egfr*^{E1} stocks were crossed to two males of each inbred line. Each cross was conducted in replicate and according to a randomized block design, interleaving both *Egfr*^{E1} strains and populations of inbred lines. All flies were grown at standard conditions: 25°C on standard cornmeal yeast medium under constant light/dark cycle.

Variation in eye roughness was scored by simple direct examination of ten individuals of each sex from each of two replicate vials according to a previously described scale [S2]. A single experimenter assigned a numerical score to each specimen along the preset qualitative scale of eye roughness, ranging from complete suppression (0) to extreme enhancement (5).

Polymorphisms in *Egfr* were identified by sequencing 10.9 kb corresponding to the six exons and flanking noncoding regions of each of the inbred lines (A.P., A. Rouse, R. Riley-Berger, I.D., and G.G., unpublished data). A total of 245 SNPs and 30 indels with the more rare allele at a frequency of 0.05 or higher were tested for their effect on eye roughness. In brief, sequencing was performed by PCR amplification of genomic DNA extracted from single flies. Gel-purified PCR products were then used as templates for the sequencing reaction using Big Dye terminator chemistry and ABI 3700 automated sequencers at the NCSU-GRL.

For the case-control and TDT tests, 1000 males from West End, NC, were collected in the summer of 2002 (WE2002). These males were immediately crossed to BL-1564 virgin females bearing the *Egfr*^{E1} allele. Eye phenotypes were determined for the non-*Cy* offspring (namely those that must carry the *Egfr*^{E1} allele). Approximately 15% of crosses produced at least one male with apparently wild-type eyes and 15% produced progeny with blistered eyes due to strong enhancement of the mutation. One male was selected from each of these case and control-producing crosses, along with his father, for subsequent genotypic analysis. These flies were sequenced for ~700 bp using the following primers: sense, AGAT TAACGTGCTCCACAGA or GGCCACTGGAATCTTGACG; antisense, GAACAGGTGTGCTCCAAGTG . Heterozygous individuals could be easily scored from the sequence chromatographs. Any ambiguously called SNPs were resequenced from the opposite direction. Of the approximately 150 father-son pairs genotyped in each class, between one-quarter and one-half were heterozygous in the father as expected and hence contributed to the TDT calculations in Table S2.

Confirmation of the polymorphism for the *pogo* transposable element was made via PCR amplification of flanking regions with the following primers: sense, GGCCAACAACAGAGTGTGTG; antisense, CGATTAGCAACGAGCTTTCC. Presence of the polymorphism could be detected by size alone and was unambiguous.

Statistical Analyses

All statistical analyses were performed using SAS Version 8.2 (Cary, NC) fitting models that included fixed effects of population ([*P*], NC or CA), sex (*S*), cross ([*C*], BL1564 or BL5144), and genotype for each SNP (*G*), and random effects of line (*L*) and replicate vial (*R*). The dependent variable, *y*, was the qualitative eye-roughness score. Line means by sex, denoted as *Y* below, were assessed for each *Egfr*^{E1} strain separately using the least square mean option in PROC GLM with the model $y = L + S + L \times S + R(L \times S) + \epsilon$.

Associations between sequence variants in *Egfr* and eye roughness were tested using PROC MIXED for each SNP separately on least square mean phenotype estimates with the model $Y = P + S + C + G + G \times C + G \times P + L(G \times P) + \epsilon$. The random effect L(G × P) controls for the correlation between sexes and crosses due to line effects that are nested within genotype and population. A number of submodels were also considered, including cross separately and sex and cross separately using PROC GLM. Where possible, results were confirmed using a nonparametric permutation procedure in Tassel (www.maizegenetics.net), which allows for the control of population admixture [S3, S4]. In general, to correct for multiple comparisons we used a sequential Bonferroni procedure [S5].

To determine whether sites of significant interest were interacting in some fashion, an a posteriori approach was employed. First, we assessed the significance of haplotypes based on the results from the initial association test using the same ANOVA approach as discussed above but with multiple-state haplotype codes replacing biallelic SNPs. Second, epistatic contributions in the BL-1564 cross were assesed in PROC MIXED using a multisite association test and a model of the form: $Y = P + S + G1 + G2 + G1 \times G2 + G1 \times G2 \times P + L(G1 \times G2 \times P) + \epsilon$, where *G1* and *G2* are the two SNP genotypes.

For the case control test, approximately 150 individuals with the most extreme class of eye-roughness score were treated as "cases," and the individuals without any visible eye roughening (i.e., the same phenotype as wild-type flies) were labeled "controls." The case control test essentially tests for differences between observed and expected allele frequencies between the two groups using a G test statistic and the traditional "χ^2," which provided similar results [S6, S7]. This test was only performed for four candidate SNPs, three of which are in strong linkage disequilibrium, and thus are nonindependent. To test for haplotype effects, we used haplotypes that showed significant differences in eye roughness as our candidate haplotypes. As well, a log-linear analysis was performed to test for interactions between candidate SNPs and the phenotypic classes. Analysis was performed in SAS using PROC FREQ for the likelihood-ratio tests and PROC CATMOD for the log-linear analysis.

The transmission-disequilibrium test (TDT) was performed according to Spielman et al. [S8] using the same SNP variants that were employed in the case-control tests. The TDT evaluates the significance of unequal SNP allele transmission from heterozygous fathers to their affected offspring. Given the sample size, we used both conventional χ^2 and an exact binomial test for the equality of two proportions, which provided comparable results. PROC FREQ was used for both of these tests.

Table S1. Components of Variation Affecting the Eye Roughness Phenotype

Line	Sex	Mean	Standard Deviation	MS_{sex}	MS_{pop}	$MS_{line/(pop)}$	$MS_{rep(P \times L)}$	MS_{error}
BL-1564	male	2.60	0.80					
				435.6	233.5	8.57	3.87	0.27
BL-1564	female	2.09	0.81					
BL-5144	male	1.85	0.68					
				341.5	1.1	4.82	3.14	0.20
BL-5144	female	1.41	0.61					

Table S2. Case Control Statistics

	Site 8196		Site 8535		Site 8541		Site 8697	
	G	T	C	T	C	T	C	T
Phenotype			Counts for Case Control					
High	34	97	86	68	70	84	110	42
Low	30	101	89	54	55	88	123	27
p (χ^2)	0.56		0.26		0.22		0.046	
p(g)	0.56		0.26		0.22		0.045	

Table S3. Significant Sites for the Main Association

Site	GenBank	−log(p)	Common/Rare Allele	Frequency of Rare Allele
00654	06065	4.3	G/T	0.14
01085	06410	4.1	C/T	0.08
03418	35391	4.0	G/A	0.31
06889	38816	3.8	C/T	0.12
08196	40119	4.2	T/G	0.42
08535	40458	7.4	C/T	0.49
08541	40464	7.6	T/C	0.48
08697	40620	6.6	C/T	0.32
09633	41556	4.8	C/T	0.19
10322	42241	4.5	C/A	0.35

Information is provided for each of the sites that were significant for the cross × genotype interaction term of the primary association test. Site refers to the site number used for the contig. GenBank refers to the site number for the polymorphism in Genbank record 17571116 (Flybase FBgn000373) for *Egfr*.

Supplemental References

S1. Dominguez, M., Wassarman, J.D., and Freeman, M. (1998). Multiple functions of the *EGF* receptor in *Drosophila* eye development. Curr. Biol. *8*, 1039–1048.

S2. Lesokhin, A.M., Yu, S.-Y., Katz, J., and Baker, N.E. (1999). Several levels of *EGF* receptor signaling during photoreceptor specification in wild-type, *Ellipse*, and null mutant *Drosophila*. Dev. Biol. *205*, 129–144.

S3. Pritchard, J., Stephens, M., Rosenberg, N., and Donnelly, P. (2002). Association mapping in structured populations. Am. J. Hum. Genet. *67*, 170–181.

S4. Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D., and Buckler, E., IV. (2001). *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. *28*, 286–289.

S5. Sokal, R.R., and Rohlf, F.J. (1998). Biometry: The Principles and Practice of Statistics in Biological Research, Fourth Edition. (San Francisco: W.H. Freeman and Company).

S6. Sasieni, P.D. (1997). From genotypes to genes: doubling the sample size. Biometrics *53*, 1253–1261.

S7. Weir, B.S. (1996). Genetic Data Analysis II (Sunderland, MA: Sinauer Associates).

S8. Spielman, R.S., McGinnis, R.E., and Ewens, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. *52*, 506–516.