Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior

John Shorter, Charlene Couch, Wen Huang, Mary Anna Carbone, Jason Peiffer[1], Robert R. H. Anholt, and Trudy F. C. Mackay[2]

Department of Biological Sciences, Program in Genetics and W. M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, NC 27695-7614

Contributed by Trudy F. C. Mackay, May 27, 2015 (sent for review April 21, 2015; reviewed by Jason H. Moore and Daniel R. Weinberger)

Aggression is an evolutionarily conserved complex behavior essential for survival and the organization of social hierarchies. With the exception of genetic variants associated with bioamine signaling, which have been implicated in aggression in many species, the genetic basis of natural variation in aggression is largely unknown. *Drosophila melanogaster* is a favorable model system for exploring the genetic basis of natural variation in aggression. Here, we performed genome-wide association analyses using the inbred, sequenced lines of the *Drosophila melanogaster* Genetic Reference Panel (DGRP) and replicate advanced intercross populations derived from the most and least aggressive DGRP lines. We identified genes that have been previously implicated in aggressive behavior as well as many novel loci, including gustatory receptor 63a (*Gr63a*), which encodes a subunit of the receptor for CO_2, and genes associated with development and function of the nervous system. Although genes from the two association analyses were largely nonoverlapping, they mapped onto a genetic interaction network inferred from an analysis of pairwise epistasis in the DGRP. We used mutations and RNAi knock-down alleles to functionally validate 79% of the candidate genes and 75% of the candidate epistatic interactions tested. Epistasis for aggressive behavior causes cryptic genetic variation in the DGRP that is revealed by changing allele frequencies in the outbred populations derived from extreme DGRP lines. This phenomenon may pertain to other fitness traits and species, with implications for evolution, applied breeding, and human genetics.

Drosophila Genetic Reference Panel | genome-wide association mapping | extreme QTL mapping | advanced intercross population | epistasis

Aggression is a near-universal animal behavior, important for securing food resources, defense against predators, gaining access to mating partners and, among social animals, creating and maintaining dominance hierarchies. Aggressive behavior is a typical quantitative trait, with natural variation attributable to segregating variants at multiple interacting loci, the effects of which are sensitive to the physical and social environments to which the individuals are exposed (1). Sociopathic and violent behaviors place a significant socioeconomic burden on human societies. However, disentangling the relative genetic and environmental contributions to aggressive behavior in natural populations is challenging due to its low heritability, and, in humans, because of the difficulty in accounting for social and other environmental influences, precisely quantifying aggression, and comorbidity with other psychological disorders.

These challenges can be overcome using animal models. Numerous studies have highlighted the evolutionary conservation of neural pathways affecting aggression (2–4). Genes affecting bioamine signaling affect aggressive behavior in humans (4, 5), mice (6–8), and *Drosophila* (9–15). The nuclear receptor subfamily 2, group E, member 1 gene *Nr2e1* affects aggressive behavior in mice (16) and humans (17, 18), and the *Drosophila* ortholog of murine *Nr2e1*, *tailless*, and its transcriptional corepressor, *Atrophin*, both affect fly aggressive behavior (19).

The development of reproducible assays for aggressive behavior in *Drosophila melanogaster* (4, 20, 21) facilitated the resolution of neural circuits affecting aggression. Neurons expressing the male-specific isoform of the sex-determining gene *fruitless* (22–24) are in part responsible for the sexual dimorphism in fighting styles between males and females (25). Olfactory sensory neurons expressing *Or67d* (26) and *Or65a* (27) both detect the male-specific volatile pheromone 11-*cis*-vaccenyl acetate and are, respectively, associated with promoting aggression in isolated males and suppressing aggression in group-housed males (28, 29). The cuticular hydrocarbon (z)-7-tricosene is also required for male–male aggression; this effect is modulated via gustatory receptor neurons expressing *Gr32a* (30). Unbiased mutagenesis screens for genes affecting aggressive behavior (31) and transcriptional profiling of mutations (32, 33), artificial selection lines (34, 35), and wild-derived inbred lines (36) divergent for aggressive behavior have implicated hundreds of genes spanning diverse biological processes.

Although aggressive behavior has a large mutational target size in *Drosophila*, we do not know to what extent genes identified from mutant analysis also harbor alleles affecting naturally occurring variation, nor how allelic variants at multiple segregating loci combine to affect aggressive behavior. The effects of *Drosophila* mutations (33) and natural variants (37) on aggression depend on genetic background, a phenomenon also observed for mutations in neuronal *nitric oxide synthase* (38) and *Nr2e1* (16) in mice. Analysis of naturally segregating variation can leverage polymorphisms at many loci to explore the contribution of epistatic interactions to genetic variation for complex traits (39–41).

Significance

Aggressive behavior is evolutionarily conserved and genetically complex, but the genetic basis of natural variation in aggression is largely unknown. We performed genome-wide association analyses using the inbred, sequenced lines of the *Drosophila* Genetic Reference Panel (DGRP) and an advanced intercross population derived from the most and least aggressive DGRP lines. These analyses identified largely nonoverlapping genes that mapped onto a genetic interaction network inferred from an analysis of pairwise epistasis in the DGRP. We functionally validated candidate genes and genetic interactions. Epistasis for aggressive behavior causes cryptic genetic variation in the DGRP that is revealed by changing allele frequencies. This observation may apply to other fitness traits and species, with implications for evolution, applied breeding, and human genetics.

Author contributions: R.R.H.A. and T.F.C.M. designed research; J.S., C.C., and M.A.C. performed research; J.S., W.H., J.P., and T.F.C.M. analyzed data; and J.S., W.H., R.R.H.A., and T.F.C.M. wrote the paper.

Reviewers: J.H.M., University of Pennsylvania; and D.R.W., Lieber Institute for Brain Development.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]Present Address: Dupont Pioneer, Johnston, IA 50131.

[2]To whom correspondence should be addressed. Email: trudy_mackay@ncsu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1510104112/-/DCSupplemental.

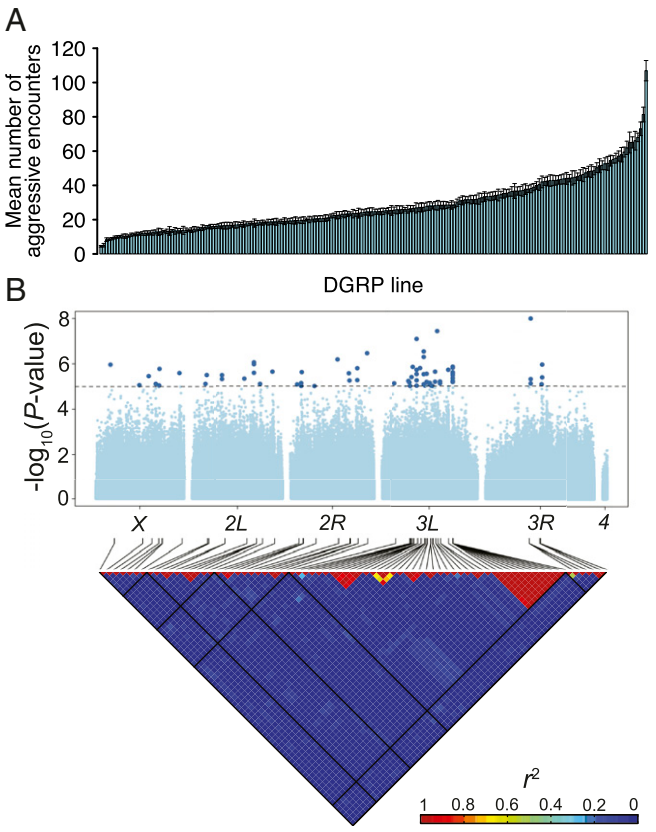

Fig. 1. Aggressive behavior in the DGRP. (*A*) Rank-ordered mean aggression scores in 200 DGRP lines. Error bars are SEM. (*B*) GWA analysis for aggression; $-\log_{10}$(*P* values) are shown for all markers tested. The dashed line indicates the nominal $P < 10^{-5}$ reporting threshold. The heat map depicts LD among the top variants.

Here, we performed a genome-wide association (GWA) analysis of aggressive behavior using the sequenced inbred lines of the *D. melanogaster* Genetic Reference Panel (DGRP) (42, 43) and an extreme quantitative trait locus (QTL) (39, 40, 44, 45) GWA analysis using an advanced intercross population (AIP) derived from aggressive and nonaggressive DGRP lines. We identify novel as well as previously documented genes affecting aggression, show that epistatic interactions affect genetic variation in aggression and can be used to derive a genetic interaction network, and functionally validate candidate genes as well as candidate genetic interactions.

Results

Quantitative Genetics of Aggression in the DGRP. To characterize natural variation in aggression, we quantified male–male aggressive behavior for 200 DGRP lines. We found substantial genetic variation in aggressive behavior, with a greater than 20-fold difference between the most and least aggressive lines (Fig. 1*A* and Dataset S1). This range of variation greatly exceeds the less than threefold difference achieved by 28 generations of divergent artificial selection from an outbred population derived from the Raleigh population (35). The broad sense heritability (±SE) for aggressive behavior is $H^2 = 0.69 \pm 0.07$ (Table S1). Surprisingly, the narrow sense heritability estimated from the pairwise genomic relationship matrix was $h^2 = 0.00$. We infer from the discrepancy between the broad and narrow sense heritabilities that the genetic architecture of aggressive behavior does not satisfy the assumptions of the additive, infinitesimal model used to determine heritability from the realized genomic relationship. Alleles affecting natural variation in aggressive behavior in the DGRP may have large effects and/or interact epistatically.

GWA Mapping in the DGRP. To identify genes and variants associated with aggressive behavior, we performed two GWA analyses in the DGRP. First, we assessed the effects of 1,914,528 common [minor allele frequencies (MAF) ≥ 0.05] variants using a mixed model analysis that accounts for the effects of *Wolbachia* infection, inversion karyotypes, and polygenic relatedness (43). A quantile–quantile (Q–Q) plot of the resulting *P* values (Fig. S1) shows enrichment above random expectation for *P* values below 10^{-5}. Therefore, we chose this threshold for reporting candidate associations of 74 variants in 39 genes (Fig. 1*B* and Dataset S2, Tab A). One SNP in the intron of *CG14869* (*AdamTS-A*) met the Bonferroni-adjusted 5% significance threshold (2.61×10^{-8}). Only one of the top candidate genes—a serotonin receptor (*5-HT1A*)—has been previously associated with aggressive behavior (10, 13–15); others are plausible candidates based on their expression in the brain (46).

To evaluate the effects of rare alleles on aggression, we performed the sequence kernel association test (SKAT) (47), which tests the aggregate effect of rare variants (MAF < 0.05) for each gene. We identified 22 genes significant at $P < 10^{-4}$ (an approximate false discovery rate of FDR ≈ 0.07), of which 10 were significant following a Bonferroni correction (Dataset S2, Tab B). None of these genes have been previously associated with aggression. These associations implicate genes affecting diverse biological processes in aggressive behavior, including chitin binding (*Cht7*); motor activity (*CG10859*); reproduction (*Sfp87B*); and chemosensation via the gustatory receptor 63a (*Gr63a*), which is part of the receptor complex for CO_2 (48), which flies emit when stressed (49). Only one gene (*CG7110*) overlapped between the single-variant and gene-based GWA analyses. This is because the single-variant GWA evaluates association between individual common variants and the trait variation, whereas the gene-based SKAT test assesses the aggregate effects of all rare variants within a gene without the necessity for any to have individually strong effects.

In summary, our GWA analyses in the DGRP have implicated both common and rare alleles at largely novel loci affecting aggressive behavior. Although only some tests for individual variants and genes reach Bonferroni-adjusted significance thresholds, departures from a uniform distribution of *P* values above the reporting thresholds after properly accounting for residual population structure indicate significant enrichment of the top associations for true positives.

Extreme QTL GWA Mapping. We can test the inferences about genetic architecture derived from GWA analyses in the DGRP by creating large outbred populations that segregate for the variants associated with the trait in the DGRP (39, 40). We created two independent AIPs using three DGRP lines with extremely high and three lines with extremely low aggression scores. We crossed these lines using a full diallel design (excluding parental genotypes) to ensure equal initial contribution of all genotypes, and maintained them by random mating in large populations over 17 generations ($n = 300$ per AIP). We selected the 10% most and 10% least aggressive individuals from 1,500 scored in each replicate population (Fig. 2 *A* and *B*), pooled the individuals from the top and bottom extremes of the phenotypic distribution, and sequenced the pools. We evaluated the difference in allele frequency between extreme pools in both replicate populations for the 914,798 SNPs that segregate in the six parental lines, and combined the evidence of association from both populations. We observed 746 SNPs in or near 355 genes below $P < 10^{-5}$, of which 26 were significant after applying a Bonferroni correction for multiple testing ($P < 5.47 \times 10^{-8}$) (Dataset S2, Tab C, and Fig. 2*C*). The top genes from this analysis included *5-HT2*, *neur*, *Nmdar1*, *pxb*, *siz*, *Tk*, and *TkR86C*, which have previously been associated with aggressive behavior (24, 31–33), as well as octopamine (*Oct-TyrR*, *Octβ2R*, *Octβ3R*), serotonin (*5-HT7*), and dopamine (*DopR*) receptors, consistent

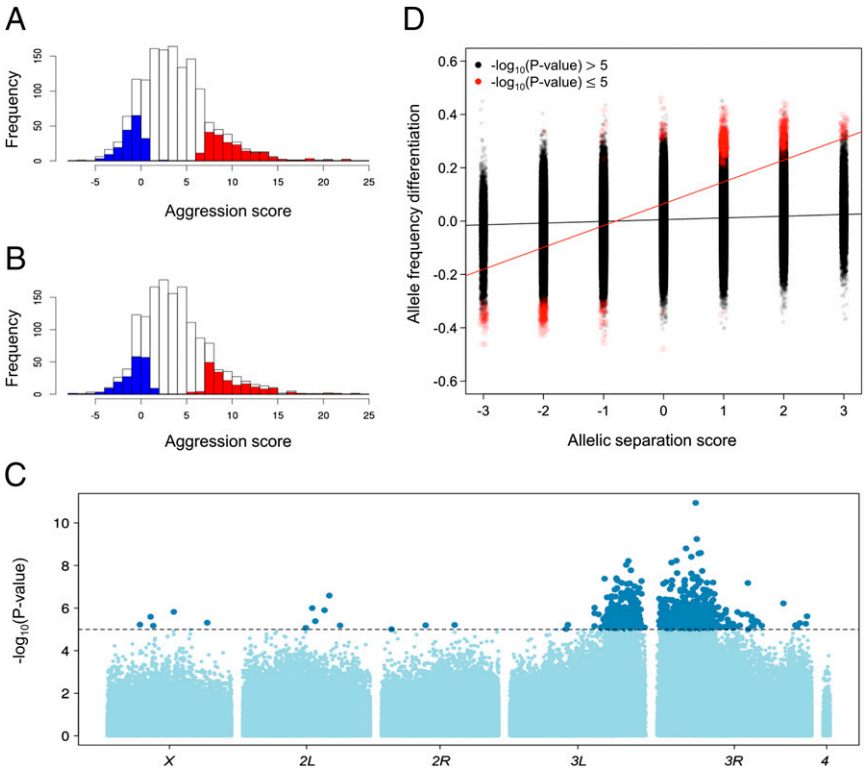

Fig. 2. Aggressive behavior in the AIP. (*A* and *B*) Histograms of aggression scores in the AIPs. Individuals with extreme high and low scores selected for pooled sequencing are shown in red and blue, respectively. (*A*) Replicate 1. (*B*) Replicate 2. (*C*) GWA analysis for aggression using extreme QTL mapping. The dashed line indicates the nominal $P < 10^{-5}$ reporting threshold. (*D*) Association of allelic separation scores for all loci segregating in the pooled AIPs (*x* axis) and the allele frequency difference between high and low parental lines (*y* axis). The red circles denote the top variants associated with aggressive behavior ($P < 10^{-5}$), and the black circles denote the remaining variants. The trend lines depict the regression of allele frequency difference in the high and low pools in the AIP on the allelic separation score for the significant variants (red line) and the remaining variants (black line).

with the well-known role of bioamine signaling in aggressive behavior (9–15).

We constructed an allelic separation score for each variant that summarizes the direction of the allelic effect on aggression in the parental lines (Fig. 2*D*). If the extreme QTL GWA analysis is enriched for variants that explain the difference in aggression between the six parental lines, we expect a positive correlation between the allelic separation scores and the difference between allele frequencies in the high and low pools for variants associated with aggression in the AIP, but not for variants that do not reach the significance threshold. Indeed, there is a strong correlation between the allelic separation score and allele frequency difference for the top variants ($r = 0.92$), in contrast to very weak correlation ($r = 0.17$) for the remaining variants (Fig. 2*D*). Thus, alleles that explain the difference in aggression in the parental lines were more likely to be associated with aggression in the AIPs.

Comparison Between DGRP and Extreme QTL GWA Analyses. We identified variants with additive effects on aggressive behavior in both the DGRP and the AIP populations derived from six DGRP lines. Only one gene (*Rbp6*), but no variant, was in common between the two experiments. One explanation could be that all of the variants identified in both experiments were false positives given a lenient reporting threshold of $P < 10^{-5}$. However, this is unlikely given that there was enrichment of low *P* values in the DGRP beyond expectation based on a uniform distribution, and many AIP variants exceeded stringent Bonferroni-adjusted significance thresholds. Further, the high correlation between the allelic separation score and allele frequency difference for the top variants in the extreme QTL GWA analysis indicates the majority of variants are true positives.

A second explanation for the lack of concordance between the two experiments is that rare variants in the DGRP that were not tested for individual association may be frequent in the AIP, and, conversely, the six founder lines of the AIP may not harbor all of the variants that were tested in the DGRP. Indeed, of the 746 SNPs that were significant in the AIP, 103 were rare in the DGRP and not tested in the GWA based on individual variants (Dataset S2, Tab C). Similarly, among the 74 variants that were significant in the DGRP GWA analysis, 31 were either not polymorphic in the six parental lines or did not pass the filters for inclusion in the AIP GWA analysis (Dataset S3). However, these only represented a minority of all variants tested.

Third, we induced linkage disequilibrium (LD) in the initial generation to construct the AIPs. Individuals from the AIPs were collected following 10–17 generations of random mating, which is not sufficient time for LD to decay between closely linked variants (1). Thus, it is possible that variants in close proximity to top aggression-associated variants in the DGRP were identified in the AIP GWA analysis and vice versa. To test this, we mapped all variants 500 bp upstream and downstream of each significant DGRP variant in the AIP, and reciprocally mapped all variants 500 bp upstream and downstream of each significant AIP variant in the DGRP, and computed the *P* values for association for all variants in these windows for both populations (Fig. S2). For the vast majority of significant variants in the DGRP and the AIP, no variants within 500 bp showed even weak association ($P < 0.001$) in the other population (0.8% of top AIP variants and 18% of top DGRP variants). Therefore, LD is unlikely to be a major factor contributing to the lack of concordance between the two analyses.

Fourth, the DGRP lines are inbred and largely homozygous, whereas the AIP individuals are outbred, with homozygous and heterozygous genotypes at each segregating locus. Nonadditive gene action from dominance and overdominance could partially account for the lack of replication of top DGRP variants in the AIP. With complete dominance, heterozygotes and one of the two homozygotes will be at one end of the distribution while the other homozygote is at the other extreme. With overdominant gene action, heterozygotes will be at one extreme while one of the homozygote genotypes will be at the other extreme. In both cases, there is reduced power to detect allele frequency differences in the AIP using DNA pools of extreme individuals, because the pools with extreme individuals contain both alleles. On the other hand, dominance or overdominance cannot account for the failure of variants with strong additive effects detected by extreme QTL mapping in the AIP to replicate in the DGRP.

The only remaining plausible explanation for the lack of correspondence between variants detected in the two populations is pervasive epistasis for aggressive behavior. With epistasis, the additive (marginal) effect of a focal locus depends on the frequency of the alleles at interacting loci (39, 41). Because only six lines from the DGRP were used to construct the AIP, the allele frequencies of the variants and their combinations in the AIP are different from the DGRP, leading to distinct genetic backgrounds in the two populations. Consistent with this explanation is the observation that the effects of top variants in the two GWA analyses are poorly correlated (Fig. S3).

Epistasis for Aggressive Behavior. To test whether epistasis could explain the lack of replication between the DGRP and AIP GWA analyses, we performed a genome-wide screen for pairwise epistatic interactions in the DGRP. After filtering variants for local LD and minor allele frequency, we tested for interactions between 682,515 variants, for 2.33×10^{11} pairwise tests for association. We observed 985 interactions below a threshold of 5.0×10^{-12}, for an approximate FDR = 0.0012, and the top 34 interactions were significant at a Bonferroni-corrected threshold of 2.15×10^{-13} (Dataset S4). These interactions involved 1,299 variants in or near 948 genes. The nature of the interactions was such that the genotypes homozygous for both minor alleles were

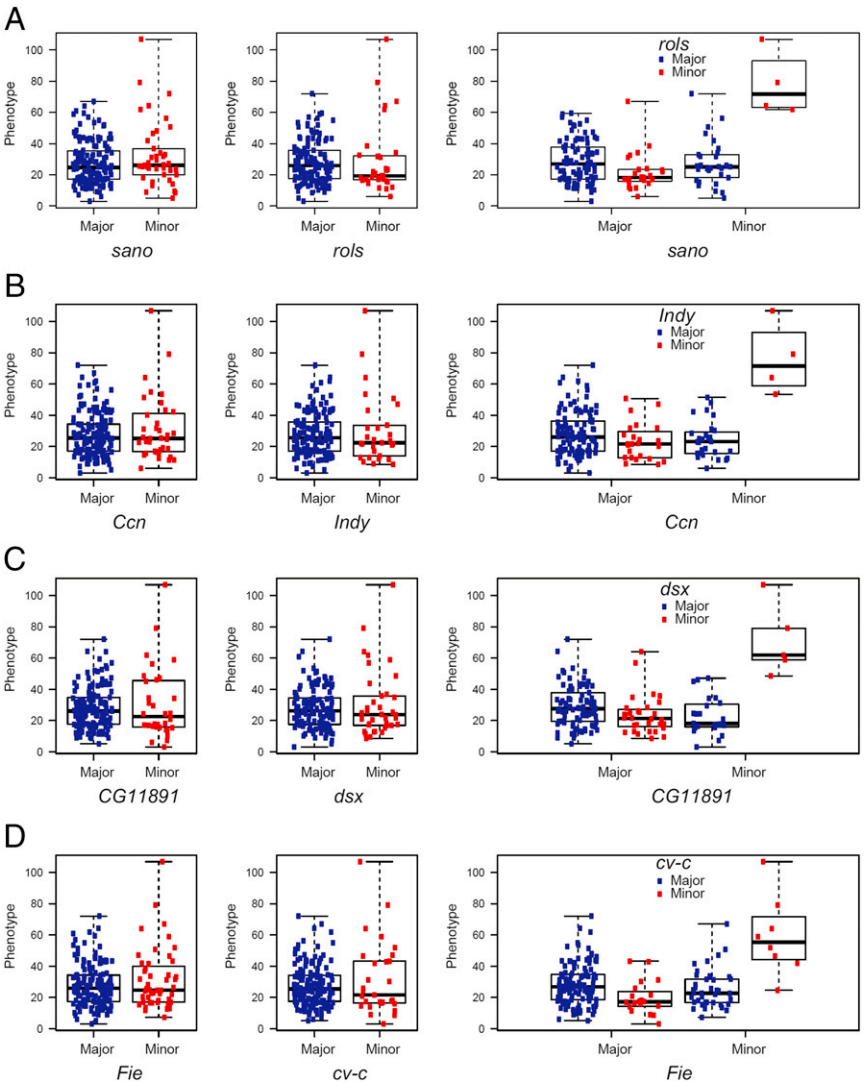

Fig. 3. Epistatic interactions in the DGRP. Box plots illustrating nonsignificant additive effects of variants with significant ($P < 5.0 \times 10^{-12}$) additive by additive epistatic interactions. "Major" and "minor" denote major and minor alleles at each locus, respectively. (*A*) *2R*_14845120_SNP (*sano*) and *3L*_12009722_SNP (*rols*), (*B*) *3L*_17718187_SNP (*Ccn*) and *3L*_18823109_DEL (*Indy*), (*C*) *3R*_21095783_SNP (*CG11891*) and *3R*_3774964_INS (*dsx*), and (*D*) *3L*_3891736_SNP (*Fie*) and *3R*_10263276_SNP (*cv-c*).

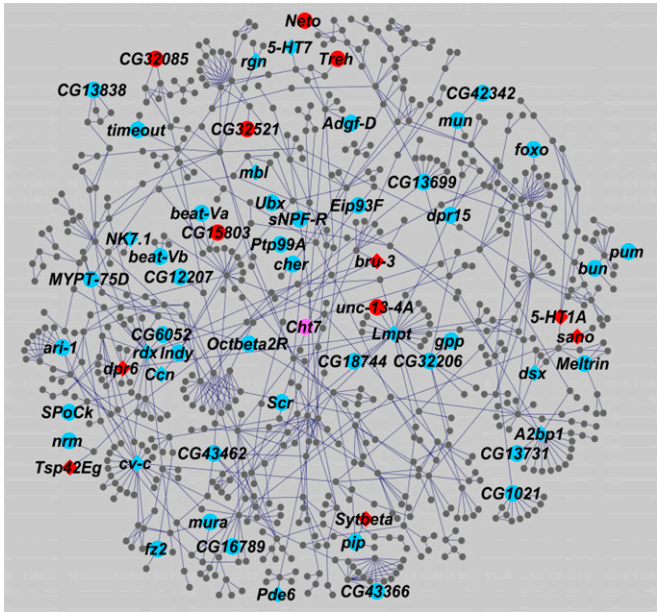

Fig. 4. Genetic interaction network for aggressive behavior. A network of 741 genes derived from significant ($P < 5.0 \times 10^{-12}$) pairwise interactions in the DGRP. Red, pink, and blue symbols show top candidate genes with additive effects from the common and rare (SKAT) DGRP and the AIP GWA analyses, respectively.

highly aggressive, with no difference in aggressive behavior for the other three genotypes (Fig. 3 *A*–*D*).

A total of 741 of the genes with top pairwise interactions formed a large genetic interaction network (Fig. 4). We next asked how many of the genes with significant additive effects from the single-marker DRGP and AIP GWA analyses also participated in epistatic interactions. We found 16 (41%) genes significant in the GWA analysis of the DGRP, and 68 (19%) genes significant in the extreme QTL analysis in AIP represented among the top pairwise epistatic interactions. Among these genes, 61 were in the largest interaction network formed by the top pairwise epistatic interactions (Fig. 4). Thus, although genes identified in the GWA analysis of the DGRP and extreme QTL analysis of the AIPs were largely inconsistent, they are connected through direct epistatic interactions or via interactions with another gene in the epistatic network. Indeed, there is significant enrichment of candidate genes from the AIP in the epistatic network constructed from DGRP marker interactions ($P = 0.001$ by 1,000 permutations), indicating that it captures the emergent additive effects in the AIP. Finally, biological process gene ontology terms involved in aspects of development and morphogenesis, including neuron development and differentiation, are highly overrepresented among the genes participating in epistatic interactions (Dataset S5), suggesting that the interaction network is not merely a statistical construct but indeed reflects biologically relevant cellular processes.

Functional Validation of Candidate Genes and Gene–Gene Interactions. The advantage of the *Drosophila* model is that we can use publicly available mutant and RNAi knock-down alleles that have been constructed in a common isogenic background to test whether different alleles in genes implicated by the GWA analyses indeed affect aggressive behavior. We tested 27 *Mi{ET1}* element insertional mutations (50) in top candidate genes in the DGRP and AIP GWA analyses based on availability of mutations in a common isogenic background; of these, 25 were present in the genetic interaction network (Fig. 4). We also tested two RNAi knock-down alleles (51). A total of 23 (79%) of these functional tests were significant (Fig. 5*A*). In all cases, the mutant or knock-down alleles were associated with decreased aggression. This did not reflect a general decrease in locomotor ability, because only three (10%) of the mutant or knock-down alleles showed a significant decrease in negative geotaxis behavior (Fig. S4). We functionally validated genes affecting bioamine signaling (*5-HT1A*, *5-HT7*) and the function and development of the nervous system (*A2bp1*, *Ccn*, *cv-c*, *dsx*, *fz*, *mbl*, *Sytβ*), as well as genes affecting chemosensation (*dpr6*, *Gr63a*), transcription (*NK7.1*), translation (*bru-3*), proteolysis (*rdx*), cGMP metabolism (*Pde6*), eye and tracheal system development (*sano*), chitin-based cuticle development (*TwdlC*), wing development (*Glut4EF*), defense response to fungus (*Lmpt*), tissue regeneration (*rgn*), and stem cell development (*Rbp6*), and genes of previously unknown function that we now annotate to affect aggression (*CG14459*, *Tsp42Eg*).

Finally, we tested whether mutations in pairs of candidate genes from the genetic interaction network (Fig. 4) that were validated from single-mutant analyses also exhibited epistasis. We compared the effects on aggression of the wild-type control, single-mutant heterozygotes, and the double-mutant heterozygote for pairs of *Mi{ET1}* element insertional mutations. Two pairs were directly connected in the network (*dpr6-Lmpt*, *Ccn-rdx*), and two were connected through an intermediate gene (*sano-5-HT1A*, *A2bp1-dsx*). Three of the gene pairs tested (*dpr6-Lmpt*, *Ccn-rdx*, *A2bp1-dsx*) exhibited epistasis (Fig. 5 *B*–*E*). Our ability to demonstrate functional causality of computationally predicted epistasis is remarkable because the genetic background of the *Mi{ET1}* insertion lines is unrelated to the DGRP lines, and, further, the DGRP-derived epistatic interaction network was derived from homozygous (additive by additive) interactions whereas the functional tests were based on heterozygous (dominance by dominance) interactions between the candidate genes.

Discussion

Low narrow-sense heritabilities, lack of control of the external and social environment, and difficulty in quantifying aggressive behavior all pose a considerable challenge to understanding the genetic basis of variation in aggression in natural populations. We were able to overcome these limitations in *D. melanogaster* by leveraging standardized phenotyping assays, strict control of the physical and social environment, and experimental designs for gene mapping only possible in a model organism that can be inbred to homozygosity and crossed to produce maximally genetically diverse outbred populations with respect to alleles affecting aggressive behavior.

The realized narrow-sense heritability of aggressive behavior in the Raleigh, NC, population of *D. melanogaster*, estimated by replicated responses to divergent artificial selection (35), is $h^2 = 0.094$. Thus, most of the variation in aggressive behavior in a natural population is attributable to environmental variation. Consequently, we would need a huge sample of unrelated individuals from an outbred population to identify contributing genetic variants by association mapping (1). However, we can use three strategies in *Drosophila* to increase the contribution of genetic variation to phenotypic variation and, hence, increase the power of mapping in smaller mapping populations, as well as to determine the relative contributions of additive and epistatic genetic variance to aggressive behavior (1). First, the heritability of inbred line means increases as a function of the number of individuals measured per line. Second, the genetic variation among inbred lines is expected to be twice that of the outbred population from which they were derived, assuming strictly additive gene action. In the presence of pairwise epistasis, the genetic variance among the inbred lines is further augmented by 4 times the additive by additive epistatic variance (52). Third, outbred populations derived from a small number of inbred lines have intermediate allele frequencies at all segregating loci, thus

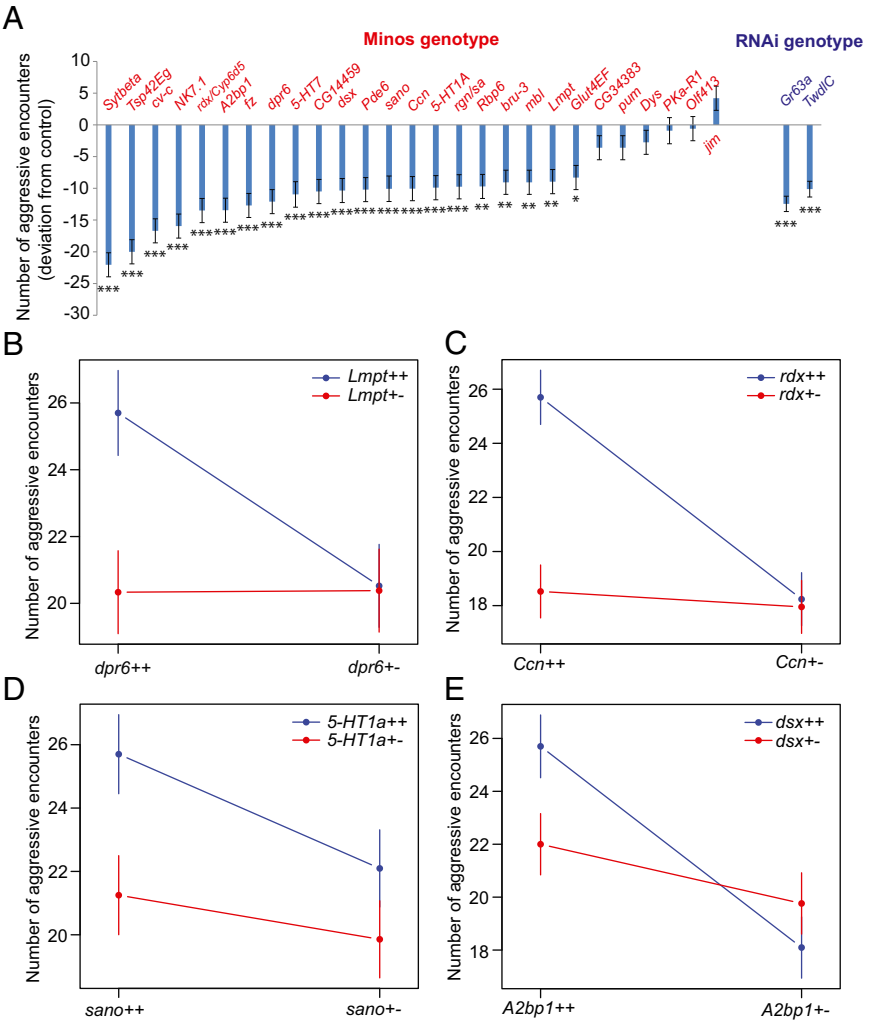

Fig. 5. Functional validation. (*A*) Effects on aggression of homozygous *Mi{ET1}*-element insertional mutations and RNAi knock-down alleles, expressed as deviations from respective coisogenic controls. Error bars are SEM. ***$P < 0.001$; **$P < 0.01$; *$P < 0.05$. (*B*–*E*) Dominance by dominance epistatic interactions. Each panel depicts means (± SEM) of four genotypes for two potentially interacting genes, *G1* and *G2*: *G1*++ *G2*++ (wild type), *G1*++ *G2*+– (*G2* mutant heterozygote), *G1*+– *G2*++ (*G1* mutant heterozygote), *G1*+– *G2*+– (*G1 G2* double-mutant heterozygote). Plus (+) denotes the wild-type allele, and minus (–) denotes the *Mi{ET1}*-element mutation. (*B*) The *P* value for the *dpr6* – *Lmpt* interaction (P_I) is $P_I = 0.040$, (*C*) *Ccn* – *rdx* $P_I = 0.0078$, (*D*) *sano* – 5-*HT1A* $P_I = 0.37$, and (*E*) *A2bp1* – *dsx* $P_I = 0.024$.

increasing heritability relative to a natural population, in which the majority of alleles are rare (53).

We estimated the broad-sense heritability of aggressive behavior in 200 inbred DGRP lines derived from the Raleigh, NC, population (42, 43) as $H^2 = {\sigma_L}^2/({\sigma_L}^2 + {\sigma_\varepsilon}^2) = 0.69$ (${\sigma_L}^2$ and ${\sigma_\varepsilon}^2$ are, respectively, the among- and within-line variance components). In contrast to the high broad-sense heritability, the narrow-sense heritability estimated from average pairwise genomic relationships of the sequenced DGRP lines was, surprisingly, $h^2 = 0$, suggesting that additive by additive epistatic interactions are prevalent. Previous studies of *Drosophila* (33) and mouse (16, 38) mutations and *Drosophila* inbred lines (37) have demonstrated that allelic effects of genes affecting aggression depend on genetic background. A hallmark of epistasis is that the additive effect of one of the loci participating in an interaction depends on the frequency of the interacting locus (39, 41), and thus, epistatic variance can be converted to additive variance (and vice versa) by changing allele frequencies of the interacting loci (41, 54–56). Therefore, in an outbred population founded from three high and three low DGRP lines, we expect to be able to detect additive alleles whose effects were masked by epistatic interactions in the DGRP.

Our GWA analyses in the DGRP and AIP derived from extreme DGRP lines reveal a highly polygenic genetic architecture of aggressive behavior that is dominated by epistatic gene action. We identified several genes (*5-HT1A*, *5-HT2*, *neur*, *Nmdar1*, *pxb*, *siz*, *Tk*, and *TkR86C*) previously implicated in aggression as well as additional genes involved in bioamine signaling (*5-HT7, DopR*, *Oct-TyrR*, *Octβ2R*, *Octβ3R*) at which segregating variation is associated with natural variation in aggressive behavior. However, the vast majority of the candidate genes are novel with respect to aggression, and for many, this study represents their first biological function annotation. These candidate genes represent a wide diversity of molecular functions and biological processes consistent with the multiple developmental and sensory modalities required for the manifestation of aggressive behavior, including the receptor for CO_2, which flies emit when stressed (49).

We found only one gene (*Rbp6*) in common between the single-variant DGRP and AIP GWA analyses. Given the quantitative genetic evidence for the prevalence of epistatic variance for aggressive behavior in the DGRP, we postulated that lack of

correspondence between the GWA analyses in the DGRP and the AIP was due to the difference in frequency of alleles affecting aggression in the two populations. Therefore, we performed a GWA screen for pairwise epistatic interactions in the DGRP. We identified a large interaction network of 741 genes from interactions that are significant at FDR = 0.0012 in which genes from the DGRP and AIP GWA analyses are either directly connected or connected through intermediate genes. In all cases, the nature of the epistatic interactions was such that significant variants in the AIP were not detectable in single-variant tests in the DGRP, because the genotypes with minor alleles at the interacting loci were associated with increased aggression but were not at high enough frequency to cause a difference in mean between the major and minor genotypes at one of the loci. This pattern of epistasis could, however, be detectable in a single-variant GWA for differences in variance among lines (40, 54).

We functionally validated 79% of all individual candidate genes tested and 75% of predicted epistatic interactions tested using mutations generated in an unrelated genetic background, suggesting that the genetic interaction network may reflect underlying biological processes common to all genotypes. However, the nodes of this network that are visible in any one analysis will vary according to the particular constellation of alleles present. This network provides context to the heterogeneous list of candidate genes and a blueprint for future studies assessing how genetic interactions translate to the underlying neurobiology and neural circuitry underpinning variation in aggressive behavior.

Our genome-wide approaches based on analyses of natural variation uncovered several genes associated with aggression that were previously identified in studies on single genes, thereby placing independent observations based on single-gene analyses into a common framework. Our approach also identifies several genes with no previous association with aggression, shifting the study of aggression beyond bioamine signaling. Furthermore, we were able to not only derive epistatic interactions computationally but also validate several pairwise gene–gene interactions experimentally, thereby corroborating functional causality of computationally predicted epistasis.

Our observation of pervasive epistasis is not likely to be unique to aggressive behavior or to *Drosophila*, and could be a feature of the genetic architecture of most traits related to fitness, which are expected to have little additive genetic variance, but can have substantial nonadditive genetic variation from dominance and epistasis (1). Epistasis gives rise to cryptic genetic variation, potentially contributing to the "missing heritability" in human GWA studies (57). In the presence of epistasis, the relative contribution of additive and epistatic variance can vary markedly between populations with different allele frequencies, with important implications for applied plant and animal breeding, evolution (55, 56), and interpretations of genetic studies in human populations aimed at identifying disease risk factors or allelic variants that contribute to variation in behavior.

Methods

***Drosophila* Stocks.** The 200 DGRP lines (42, 43) and the isogenic strain w^{1118} Canton S B (CSB) (58) are maintained in our laboratory. We obtained *Mi{ET1}* insertional mutations (*5-HT1A*^{MB09812}, *5-HT7*^{MB04445}, *CG31008*^{MB04445}, *A2bp1*^{MB03305}, *bru-3*^{MB04010}, *Ccn*^{MB12229}, *CG14459*^{MB02564}, *CG34383*^{MB05180}, *cv-c*^{MB03489}, *dpr6*^{MB04096}, *dsx*^{MB08902}, *Dys*^{MB03186}, *fz*^{MB07478}, *Glut4EF*^{MB03996}, *jim*^{MB04583}, *Lmpt*^{MB09664}, *mbl*^{MB06084}, *NK7.1*^{MB02679}, *olf413*^{MB02820}, *Pde6*^{MB06146}, *Pka-R1*^{MB04145}, *pum*^{MB06187}, *Rbp6*^{MB06777}, *rdx*^{MB09381} *Cyp6d5*^{MB09381}, *rgn*^{MB02115} *sa*^{MB02115}, *sano*^{MB03560}, *Sytbeta*^{MB11464}, and *Tsp42Eg*^{MB08050}) and their corresponding w^{1118} isogenic control line (50) from the Bloomington *Drosophila* Stock Center (flystocks.bio.indiana.edu/). *UAS*-RNAi knock down alleles of *Gr63a*^{KK106612} and *TwdlC*^{KK110000} were obtained from the Vienna *Drosophila* Stock Center (stockcenter.vdrc.at/control/main) (51). We obtained a *tubulin-GAL4* driver line (y^1 *w**; *P{tubP-GAL4} LL7/TM3*, *Sb*^{1}) from the Bloomington *Drosophila* Stock Center. All flies were reared and behavioral measurements assessed under standard culture conditions (cornmeal–molasses–agar medium, 25 °C, 60–75% relative humidity, 12-h light:dark cycle).

We created two replicate AIPs from three DGRP lines with extremely high (DGRP_57, DGRP_324, DGRP_852) and three with extremely low (DGRP_45, DGRP_228, DGRP_235) aggression scores by crossing them in a full diallel cross design, excluding homozygous parental genotypes, to create starting populations of 30 reciprocal F1 genotypes. In the F2 (G1) generation, we seeded each of five bottles per replicate population with one male and one virgin female from each of the 30 genotypes, for a population size per replicate population of $n = 300$. The flies were allowed to mate randomly and lay eggs for 2 days. In subsequent generations, six males and six females from each of the five bottles of each replicate were placed into five new bottles and discarded after 2 days. The two replicate populations were raised contemporaneously but independently for 17 generations.

For functional validation experiments using RNAi alleles, F1s were generated by crossing *UAS*-RNAi virgin females to *tub-GAL4* males. To generate F1 individuals for double-heterozygote *Mi{ET1}* mutant assays, the w^{1118} isogenic control line was crossed to each homozygous mutant line to generate single heterozygous genotypes, and the homozygous mutant lines were crossed to create a double heterozygote.

Behavior Assays. We quantified aggressive behavior for groups of eight flies of the same genotype ("eight fly" assay) or of a single focal fly and seven flies of a different genotype ("focal fly" assay), as previously described (35). All males tested were 3–7 d old, and were maintained with other males and females before the assay. Assays were performed in a behavioral chamber (25 °C, 70% humidity) between 0800 hours and 1100 hours. At least 24 h before the assay, flies were anesthetized using CO_2 and placed into vials containing medium in groups of eight appropriate for the assay. Before the assays, flies were placed in a vial without food for 90 min, after which they were transferred (without anesthesia) to a vial containing a droplet of food and allowed to acclimate for 2 min. After the acclimation period, the number of aggressive encounters (wing threats, charges, head butts, chases, kicks, and boxing) was scored for 2 min. In the eight fly assays, all aggressive encounters from all flies in the group were summed to give a single aggression score per replicate vial. In the focal fly assays, only aggressive encounters in which the focal male participated were scored.

We used the eight fly assay to quantify aggressive behavior of the DGRP lines, with $n = 20$ replicate vials per genotype. We used a block design, scoring all 20 replicates for 10 DGRP lines plus 20 replicates of CSB flies each week. We used the focal fly assay to quantify aggression in the AIP, beginning at generation 10. The focal fly was a single AIP male, and the competitors were white-eyed CSB males. We scored 20 males from each population for 10 consecutive days from generations 10–17 inclusive, for a total of ∼3,000 individuals (∼1,500 per population). An adjusted aggression score was calculated by taking the number of aggressive encounters initiated by the focal fly and subtracting the number of aggressive encounters displayed to the focal fly by the other males. All scored focal flies were individually frozen at –80 °C for subsequent DNA extraction. We used the eight fly assay to quantify aggressive behavior of *Mi{ET1}* mutant lines and their corresponding w^{1118} isogenic control line. These lines were partitioned into two blocks, with replicates tested over several days with at least 17 replicates per line, similar to the assay design used for the DGRP. Quantifying RNAi F1s was done in a similar manner, with the two knock-downs and a control measured over 3 d, with a total of 21 replicates per line.

To measure negative geotaxis, flies were placed in 25 mm × 150 mm glass vials (Pyrex–Corning flat bottom) with a ruler marking 5-mm increments from 0, indicating the lowest position, to 29, indicating the highest possible position. Each fly was tapped to the bottom of the vial, and the distance traveled upward was scored based on the highest point reached in 5 s. Twenty individual flies per genotype were assayed on each of 3 d, for a total sample size of 60 males per genotype. We used Student's *t* tests to assess the significance of differences between mutant/RNAi alleles and appropriate controls.

Quantitative Genetics of Aggressive Behavior. We adjusted the aggression scores of the DGRP lines as the difference from the appropriate CSB control line to account for weekly uncontrollable fluctuations in environmental conditions affecting behavior. We added the average of all CSB replicates over the course of the experiment to each adjusted score so the raw and adjusted scores are on the same scale. We performed a random effects analysis of variance on the adjusted data using the model $Y = \mu + L + \varepsilon$, where Y is the adjusted aggression score, μ is the overall mean, L is the DGRP line, and ε is the residual. We estimated broad sense heritability (H^2) from the variance components (σ^2) as $H^2 = \sigma_L{}^2/(\sigma_L{}^2 + \sigma_E{}^2)$ where $\sigma_L{}^2$ and $\sigma_E{}^2$ are, respectively, the among line and error variance components. We estimated

the additive genetic variance (σ_A^2) by a linear mixed model with covariance between lines determined by the genomic relationship matrix and calculated narrow sense heritability (h^2) as $h^2 = \sigma_A^2/(\sigma_A^2 + \sigma_E^2)$ (59).

GWA Analyses in the DGRP. We performed two GWA analyses. First, we performed single-variant tests of association for additive (marginal) effects of 1,914,528 variants that were present at MAFs of at least 0.05 in the 200 DGRP lines. We adjusted the aggression data for the effects of *Wolbachia* infection and major polymorphic inversions [*In(2L)t*, *In(2R)NS*, *In(3R)P*, *In(3R)K*, and *In(3R)Mo*]. To test for significance of each sequence variant, we fitted a linear mixed model $Y = \mu + m + u + \varepsilon$, where Y is the adjusted aggression score, m is the effect of the variant, u is the polygenic effect whose covariance matrix is determined by the genomic relationship matrix, and ε is the residual (43). These analyses were implemented using the DGRP website (dgrp2.gnets.ncsu.edu/).

Second, we performed gene-based GWA analyses, using SKAT (47) to assess the cumulative effect of rarer (MAF < 0.05) variants within 1 kb of each annotated gene (43). We performed these tests using the SKAT package (47) in R v.3.0.1 (60). *Wolbachia* infection status, major inversions, and the first 11 principal components of the genetic variation in the DGRP were used to adjust the aggression scores before the analyses.

Genomic DNA Isolation and Pooling from the AIPs. We selected the 10% (n = 150) most aggressive and the 10% (n = 150) least aggressive individuals from each of the two AIP populations from flies pooled across all eight generations. Genomic DNA was extracted from the 600 flies individually using the DNeasy blood and tissue kit 96-well protocol (Qiagen) according to the manufacturer's instructions. The flies were ground using sterile 1.44-mm beads and a shaker with a digestion buffer (20 μL Proteinase K, 180 μL Buffer ATL; Qiagen), followed by 55 °C incubation for 1–3 h and removal of lysate. Protein precipitation solution (30 μL; Qiagen) was added, and the samples were placed on ice for 30 min. Supernatant without precipitated protein was transferred to new plates. For DNA precipitation, 3 μL of glycogen was added to each sample, followed by 100 μL of 100% Isopropanol precipitation at −20 °C and centrifuged at 9,279 × *g* for 30 min in an Eppendorf 5415R microcentrifuge. The supernatant was removed, and the DNA pellet was washed with 80% (vol/vol) ethanol. DNA was rehydrated using 12 μL 1× Tris-EDTA buffer and quantitated using PicoGreen. All DNA samples were diluted to a common concentration, and a 10 ng/μL aliquot was taken from each individual. We created four pools of DNA from the 150 high-aggression and 150 low-aggression flies from each of the replicate populations.

Extreme QTL GWA Analyses. The DNA pools from flies with extreme aggression scores were sequenced using the Illumina HiSeq2000 platform. Libraries were barcoded, multiplexed across two lanes, and sequenced using 100-bp paired-end reads. Sequence reads were aligned to the *D. melanogaster* reference genome (FB5.49) with the Burrows–Wheeler Aligner (61). After postalignment processing, we tested for differences in allele frequencies between the high and low pools for variants that were polymorphic among the six founding parental lines, separately for the two replicate AIPs. Significance of the association (two-sided) was assessed by a Z test with the test statistic computed as $Z = (p_1 - p_2)/\sqrt{p_0(1-p_0)(1/n + 1/c_1 + 1/c_2)}$, where p_1 and p_2 are the allele frequencies in the two pools respectively, p_0 is the average allele frequency of p_1 and p_2, n is the number of flies pooled in each pool (n = 150), and c_1 and c_2 are the sequence coverages in the two pools. Under the null hypothesis of no difference between p_1 and p_2, Z is distributed as standard normal. Evidence for joint association from the two replicate populations was obtained by calculating a combined χ^2 statistic, weighted by sequence coverage, and obtaining P values from the χ^2 distribution. We also calculated an allelic separation score for variants polymorphic among the six parental lines as: $ASS = \sum N_H - \sum N_L$, where N_H is the number of major (minor) alleles in the three high-aggression lines and N_L is the sum of the number of minor (major) alleles in the three low-aggression lines.

Analysis of Epistasis. We performed a genome-wide screen for pairwise epistatic interactions, fitting models of form $Y = \mu + M_A + M_B + M_A \times M_B + \varepsilon$, where M_A and M_B denote the effects of the two interacting polymorphic markers, and $M_A \times M_B$ is the effect of the interaction using the same adjusted line means described for the gene-based tests of association. We first pruned variants in the DGRP for LD using PLINK (62), such that for any 100 consecutive variants with an MAF > 0.15, no pair had an $r^2 > 0.8$. This resulted in 682,515 variants and 2.33×10^{11} possible pairwise tests for association, which we assessed using FastEpistasis (63). FastEpistasis reports asymptotic P values from the χ^2 distribution, which we corrected with P values from the F distribution with appropriate degrees of freedom. We required that at least two lines were present for each of the four genotypic combinations. The epistatic network for the top interactions was visualized using Cytoscape (64).

To test the enrichment of AIP genes in the epistatic network, we used a permutation procedure in which we randomly sampled the same number of significant genes from the genome and counted the number of genes that were present in the network. This was repeated 1,000 times to obtain the null distribution of the number of genes overlapping the extreme QTL analysis in the AIP and the epistatic network. The P value was calculated as the number of permutations where the number of overlapping genes exceeded the observed number divided by the total number of permutations.

1. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics* (Addison Wesley Longman, Harlow, UK).
2. Kravitz EA, Huber R (2003) Aggression in invertebrates. *Curr Opin Neurobiol* 13(6): 736–743.
3. Nelson RJ, Trainor BC (2007) Neural mechanisms of aggression. *Nat Rev Neurosci* 8(7): 536–546.
4. Anholt RRH, Mackay TFC (2012) Genetics of aggression. *Annu Rev Genet* 46:145–164.
5. Brunner HG, Nelen M, Breakefield XO, Ropers HH, van Oost BA (1993) Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. *Science* 262(5133):578–580.
6. Saudou F, et al. (1994) Enhanced aggressive behavior in mice lacking *5-HT1B* receptor. *Science* 265(5180):1875–1878.
7. Popova NK, et al. (2001) Behavioral characteristics of mice with genetic knockout of *monoamine oxidase type A*. *Neurosci Behav Physiol* 31(6):597–602.
8. Holmes A, Murphy DL, Crawley JN (2002) Reduced aggression in mice lacking the serotonin transporter. *Psychopharmacology (Berl)* 161(2):160–167.
9. Baier A, Wittek B, Brembs B (2002) *Drosophila* as a new model organism for the neurobiology of aggression? *J Exp Biol* 205(Pt 9):1233–1240.
10. Dierick HA, Greenspan RJ (2007) Serotonin and neuropeptide F have opposite modulatory effects on fly aggression. *Nat Genet* 39(5):678–682.
11. Hoyer SC, et al. (2008) Octopamine in male aggression of *Drosophila*. *Curr Biol* 18(3): 159–167.
12. Zhou C, Rao Y, Rao Y (2008) A subset of octopaminergic neurons are important for *Drosophila* aggression. *Nat Neurosci* 11(9):1059–1067.
13. Johnson O, Becnel J, Nichols CD (2009) Serotonin 5-HT(2) and 5-HT(1A)-like receptors differentially modulate aggressive behaviors in *Drosophila melanogaster*. *Neuroscience* 158(4):1292–1300.
14. Alekseyenko OV, Lee C, Kravitz EA (2010) Targeted manipulation of serotonergic neurotransmission affects the escalation of aggression in adult male *Drosophila melanogaster*. *PLoS ONE* 5(5):e10806.
15. Alekseyenko OV, Chan YB, Li R, Kravitz EA (2013) Single dopaminergic neurons that modulate aggression in *Drosophila*. *Proc Natl Acad Sci USA* 110(15):6151–6156.
16. Young KA, et al. (2002) Fierce: A new mouse deletion of Nr2e1; violent behaviour and ocular abnormalities are background-dependent. *Behav Brain Res* 132(2):145–158.
17. Kumar RA, et al. (2008) Initial association of *NR2E1* with bipolar disorder and identification of candidate mutations in bipolar disorder, schizophrenia, and aggression through resequencing. *Am J Med Genet B Neuropsychiatr Genet* 147B(6):880–889.
18. Reif A, et al. (2009) Influence of functional variant of neuronal nitric oxide synthase on impulsive behaviors in humans. *Arch Gen Psychiatry* 66(1):41–50.
19. Davis SM, Thomas AL, Nomie KJ, Huang L, Dierick HA (2014) Tailless and Atrophin control *Drosophila* aggression by regulating neuropeptide signalling in the *pars intercerebralis*. *Nat Commun* 5:3177.
20. Chen S, Lee AY, Bowens NM, Huber R, Kravitz EA (2002) Fighting fruit flies: A model system for the study of aggression. *Proc Natl Acad Sci USA* 99(8):5664–5668.
21. Zwarts L, Versteven M, Callaerts P (2012) Genetics and neurobiology of aggression in *Drosophila*. *Fly (Austin)* 6(1):35–48.
22. Vrontou E, Nilsen SP, Demir E, Kravitz EA, Dickson BJ (2006) *fruitless* regulates aggression and dominance in *Drosophila*. *Nat Neurosci* 9(12):1469–1471.
23. Chan YB, Kravitz EA (2007) Specific subgroups of *FruM* neurons control sexually dimorphic patterns of aggression in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 104(49):19577–19582.
24. Asahina K, et al. (2014) Tachykinin-expressing neurons control male-specific aggressive arousal in *Drosophila*. *Cell* 156(1-2):221–235.
25. Nilsen SP, Chan YB, Huber R, Kravitz EA (2004) Gender-selective patterns of aggressive behavior in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 101(33): 12342–12347.
26. Wang L, Anderson DJ (2010) Identification of an aggression-promoting pheromone and its receptor neurons in *Drosophila*. *Nature* 463(7278):227–231.
27. Liu W, et al. (2011) Social regulation of aggression by pheromonal activation of *Or65a* olfactory neurons in *Drosophila*. *Nat Neurosci* 14(7):896–902.
28. Hoffmann AA (1990) The influence of age and experience with conspecifics on territorial behavior in *Drosophila melanogaster*. *J Insect Behav* 3(1):1–12.
29. Svetec N, Ferveur JF (2005) Social experience and pheromonal perception can change male-male interactions in *Drosophila melanogaster*. *J Exp Biol* 208(Pt 5):891–898.

30. Wang L, et al. (2011) Hierarchical chemosensory regulation of male-male social interactions in *Drosophila*. *Nat Neurosci* 14(6):757–762.
31. Edwards AC, Zwarts L, Yamamoto A, Callaerts P, Mackay TFC (2009) Mutations in many genes affect aggressive behavior in *Drosophila melanogaster*. *BMC Biol* 7:29.
32. Rollmann SM, et al. (2008) Pleiotropic effects of *Drosophila neuralized* on complex behaviors and brain structure. *Genetics* 179(3):1327–1336.
33. Zwarts L, et al. (2011) Complex genetic architecture of *Drosophila* aggressive behavior. *Proc Natl Acad Sci USA* 108(41):17070–17075.
34. Dierick HA, Greenspan RJ (2006) Molecular analysis of flies selected for aggressive behavior. *Nat Genet* 38(9):1023–1031.
35. Edwards AC, Rollmann SM, Morgan TJ, Mackay TFC (2006) Quantitative genomics of aggressive behavior in *Drosophila melanogaster*. *PLoS Genet* 2(9):e154.
36. Edwards AC, et al. (2009) A transcriptional network associated with natural variation in *Drosophila* aggressive behavior. *Genome Biol* 10(7):R76.
37. Edwards AC, Mackay TFC (2009) Quantitative trait loci for aggressive behavior in *Drosophila melanogaster*. *Genetics* 182(3):889–897.
38. Le Roy I, et al. (2000) Loss of aggression, after transfer onto a *C57BL/6J* background, in mice carrying a targeted disruption of the neuronal *nitric oxide synthase* gene. *Behav Genet* 30(5):367–373.
39. Huang W, et al. (2012) Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci USA* 109(39):15553–15559.
40. Swarup S, Huang W, Mackay TFC, Anholt RRH (2013) Analysis of natural variation reveals neurogenetic networks for *Drosophila* olfactory behavior. *Proc Natl Acad Sci USA* 110(3):1017–1022.
41. Mackay TFC (2014) Epistasis and quantitative traits: Using model organisms to study gene-gene interactions. *Nat Rev Genet* 15(1):22–33.
42. Mackay TFC, et al. (2012) The *Drosophila melanogaster* Genetic Reference Panel. *Nature* 482(7384):173–178.
43. Huang W, et al. (2014) Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res* 24(7):1193–1208.
44. Ehrenreich IM, et al. (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464(7291):1039–1042.
45. Ehrenreich IM, et al. (2012) Genetic architecture of highly complex chemical resistance traits across four yeast strains. *PLoS Genet* 8(3):e1002570.
46. St Pierre SE, Ponting L, Stefancsik R, McQuilton P; FlyBase Consortium (2014) FlyBase 102—Advanced approaches to interrogating FlyBase. *Nucleic Acids Res* 42(Database issue):D780–D788.
47. Wu MC, et al. (2011) Rare-variant association testing for sequencing data with the Sequence Kernel Association Test. *Am J Hum Genet* 89(1):82–93.
48. Kwon JY, Dahanukar A, Weiss LA, Carlson JR (2007) The molecular basis of CO_2 reception in *Drosophila*. *Proc Natl Acad Sci USA* 104(9):3574–3578.
49. Suh GSB, et al. (2004) A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila*. *Nature* 431(7010):854–859.
50. Bellen HJ, et al. (2011) The *Drosophila* gene disruption project: Progress using transposons with distinctive site specificities. *Genetics* 188(3):731–743.
51. Dietzl G, et al. (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* 448(7150):151–156.
52. Bulmer MG (1985) *The Mathematical Theory of Quantitative Genetics* (Oxford Univ Press, Oxford).
53. Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4(2):e1000008.
54. Rönnegård L, Valdar W (2011) Detecting major genetic loci controlling phenotypic variability in experimental crosses. *Genetics* 188(2):435–447.
55. Goodnight CJ (1988) Epistasis and the effect of founder events on the additive genetic variance. *Evolution* 42(3):441–454.
56. Cheverud JM, Routman EJ (1996) Epistasis as a source of increased additive genetic variance at population bottlenecks. *Evolution* 50(3):1042–1051.
57. Manolio TA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461(7265):747–753.
58. Norga KK, et al. (2003) Quantitative analysis of bristle number in *Drosophila* mutants identifies genes involved in neural development. *Curr Biol* 13(16):1388–1396.
59. Ober U, et al. (2012) Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genet* 8(5):e1002685.
60. R Core Team (2013) *R: A Language and Environment for Statistical Computing* (R Found Stat Comput, Vienna).
61. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595.
62. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
63. Schüpbach T, Xenarios I, Bergmann S, Kapur K (2010) FastEpistasis: A high performance computing solution for quantitative trait epistasis. *Bioinformatics* 26(11):1468–1469.
64. Cline MS, et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2(10):2366–2382.

Supporting Information

Shorter et al. 10.1073/pnas.1510104112

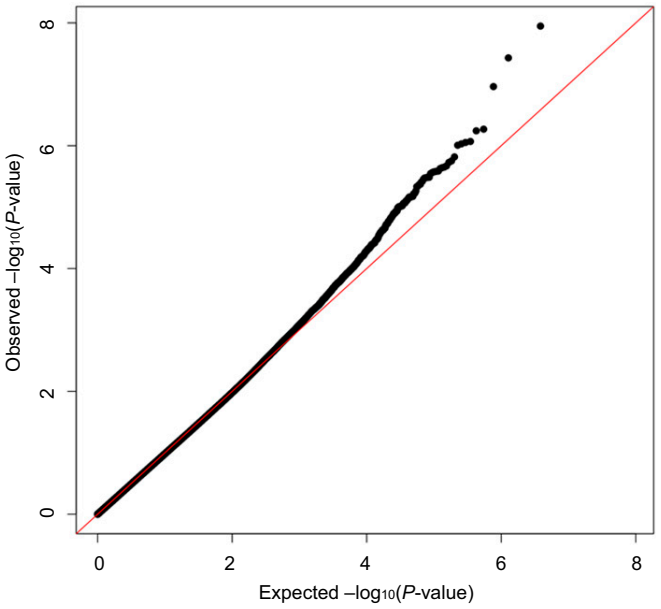

Fig. S1. Q-Q plot of *P* values from the DGRP single variant GWA analysis.

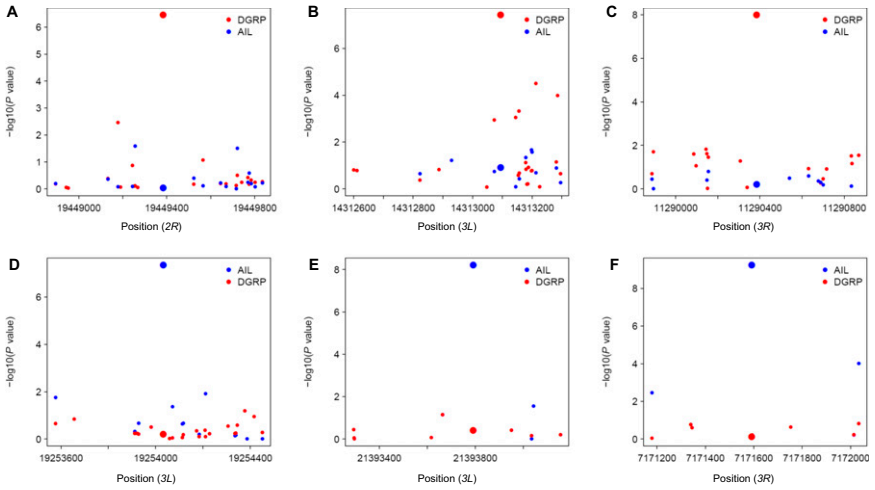

Fig. S2. LD within 1 kb of top ($P < 10^{-5}$) variants in the DGRP and AIP. Variants in the DGRP and AIP are indicated by red and blue circles, respectively. The focal variant in each window is depicted in larger size. (*A*–*C*) Top DGRP variants: (*A*) *2R*_19449384_SNP, (*B*) *3L*_14313095_SNP, and (*C*) *3R*_11290385_SNP. (*D*–*F*) Top AIP variants: (*D*) *3L*_19254033_SNP, (*E*) *3L*_21393792_SNP, and (*F*) *3R*_7171592_SNP.

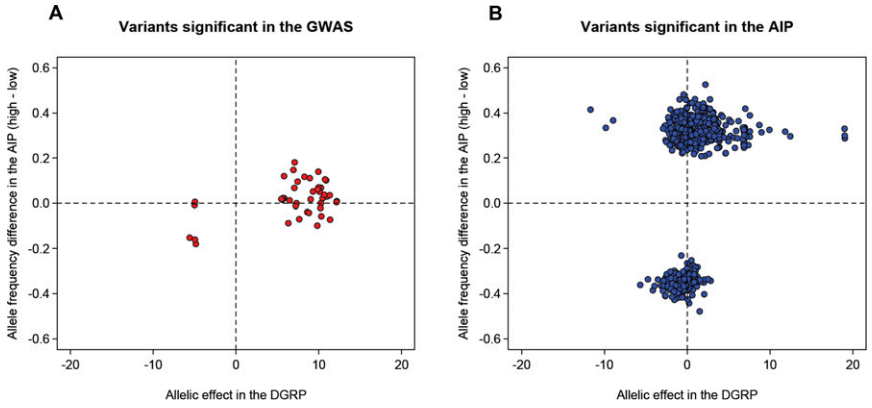

Fig. S3. Comparison of allelic effects in the DGRP (the difference between the mean aggressive behavior of lines homozygous for the minor and major alleles) and allele frequency differences in the AIP. (*A*) Top ($P < 10^{-5}$) variants from DGRP GWA analysis. (*B*) Top ($P < 10^{-5}$) variants from the AIP GWA analysis.

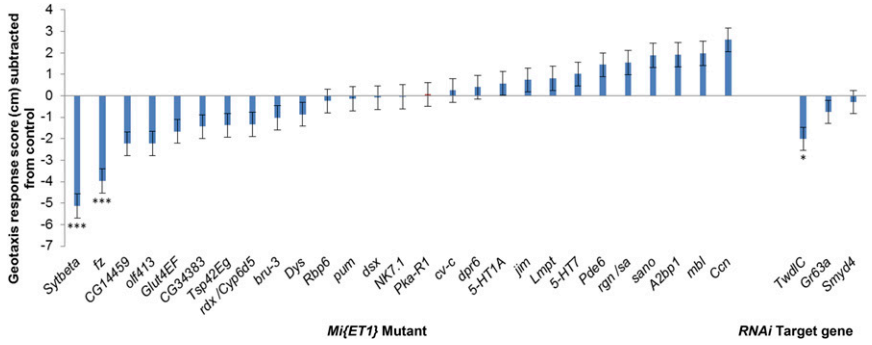

Fig. S4. Negative geotaxis. Effects on negative geotaxis of homozygous *Mi{ET1}*-element insertional mutations and RNAi knock-down alleles, expressed as deviations from respective coisogenic controls. Error bars are SEM. ***$P < 0.001$; *$P < 0.05$.

Table S1. Analysis of variance of male aggressive behavior in the DGRP

Source	df	MS	F	*P*	σ^2	H^2
Line	199	4,829.568	45.75	<0.0001	236.20	0.691
Error	3,800	105.570			105.57	

df, degrees of freedom; MS, mean squares; F, *F* statistic; σ^2, variance component; and H^2, broad-sense heritability.

Dataset S1. Mean adjusted aggression scores of DGRP lines. Raw data were adjusted for weekly environmental fluctuations using the deviations from contemporaneous CSB control line means. The overall CSB mean over all weeks was added to all adjusted scores

Dataset S1

Dataset S2. Genome-wide association analysis for aggressive behavior. (Tab A) DGRP single marker GWA analysis. Top ($P < 10^{-5}$) polymorphisms associated with aggression from single-marker analysis. The chromosome and genomic location is given for each polymorphism, as well as the gene and FlyBase ID with which it is associated. In cases where a single polymorphism could affect two genes, both are given. When polymorphisms are located within a gene, their position in the coding sequence, intron and 3′ and 5′ UTR are given. For intergenic polymorphisms, the locations in bp upstream or downstream from the nearest gene are given. The *P* values are from the mixed model accounting for *Wolbachia* infection status, karyotype of polymorphic inversions, and polygenic relatedness. Effect sizes are one half the difference in means between lines in the major and minor allele class. Information on the base call for major and minor alleles and the reference allele, and minor allele frequency and total minor allele count, is also given. (Tab B) DGRP gene-based GWA analyses. Results are from SKAT analyses of rare alleles in a gene region. *P* values $< 10^{-4}$ are reported; cells highlighted in pink indicate tests significant at a Bonferroni-corrected significance level. Genes are listed in order of their location on each chromosome arm. Vertical black bars denote nested or contiguous genes. (Tab C) Extreme QTL GWA in the AIP. Top ($P < 10^{-5}$) polymorphisms associated with aggression from single-marker analysis. The number of alleles is given for each of the six parental lines used to found the AIP populations, as well as the minor allele frequencies in the low and high pools of each replicate and the combined *P* values. Allele frequencies and *P* values are given for the same variants in the DGRP for comparison. The chromosome and genomic location is given for each polymorphism, as well as the gene and FlyBase ID with which it is associated. In cases where a single polymorphism could affect two genes, both are given. When polymorphisms are located within a gene, their position in the coding sequence, intron and 3′ and 5′ UTR are given. For intergenic polymorphisms, the locations in bp upstream or downstream from the nearest gene are given

Dataset S2

Dataset S3. Extreme QTL GWA results for top variants in the DGRP GWA analysis. Top ($P < 10^{-5}$) polymorphisms associated with aggression from single-marker analysis in the DGRP are shown with the corresponding statistics from the AIP GWA analysis. Column headings are as described in Dataset S2

Dataset S3

Dataset S4. Pairwise epistasis GWA analysis in the DGRP. Only common (minor allele frequency ≥ 0.15), LD-pruned variants were tested. An additional filter was that the rare–rare allele class was represented by a minimum of two lines. Top ($P < 5 \times 10^{-12}$) interactions are listed. The approximate FDR = 0.0012

Dataset S4

Dataset S5. Biological Process Gene Ontology enrichment analysis from pairwise epistasis GWA analysis in the DGRP

Dataset S5