

# Epistasis dominates the genetic architecture of *Drosophila* quantitative traits

Wen Huang<sup>a</sup>, Stephen Richards<sup>b</sup>, Mary Anna Carbone<sup>a</sup>, Dianhui Zhu<sup>b</sup>, Robert R. H. Anholt<sup>c</sup>, Julien F. Ayroles<sup>a,1</sup>, Laura Duncan<sup>a</sup>, Katherine W. Jordan<sup>a</sup>, Faye Lawrence<sup>a</sup>, Michael M. Magwire<sup>a</sup>, Crystal B. Warner<sup>b,2</sup>, Kerstin Blankenburg<sup>b</sup>, Yi Han<sup>b</sup>, Mehwish Javaid<sup>b</sup>, Joy Jayaseelan<sup>b</sup>, Shalini N. Jhangiani<sup>b</sup>, Donna Muzny<sup>b</sup>, Fiona Ongeri<sup>b</sup>, Lora Perales<sup>b</sup>, Yuan-Qing Wu<sup>b,3</sup>, Yiqing Zhang<sup>b</sup>, Xiaoyan Zou<sup>b</sup>, Eric A. Stone<sup>a</sup>, Richard A. Gibbs<sup>b</sup>, and Trudy F. C. Mackay<sup>a,4</sup>

Departments of <sup>a</sup>Genetics and <sup>b</sup>Biology, North Carolina State University, Raleigh, NC 27695; and <sup>b</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2010.

Contributed by Trudy F. C. Mackay, August 6, 2012 (sent for review May 29, 2012)

**Epistasis—nonlinear genetic interactions between polymorphic loci—is the genetic basis of canalization and speciation, and epistatic interactions can be used to infer genetic networks affecting quantitative traits. However, the role that epistasis plays in the genetic architecture of quantitative traits is controversial. Here, we compared the genetic architecture of three *Drosophila* life history traits in the sequenced inbred lines of the *Drosophila melanogaster* Genetic Reference Panel (DGRP) and a large outbred, advanced intercross population derived from 40 DGRP lines (Flyland). We assessed allele frequency changes between pools of individuals at the extremes of the distribution for each trait in the Flyland population by deep DNA sequencing. The genetic architecture of all traits was highly polygenic in both analyses. Surprisingly, none of the SNPs associated with the traits in Flyland replicated in the DGRP and vice versa. However, the majority of these SNPs participated in at least one epistatic interaction in the DGRP. Despite apparent additive effects at largely distinct loci in the two populations, the epistatic interactions perturbed common, biologically plausible, and highly connected genetic networks. Our analysis underscores the importance of epistasis as a principal factor that determines variation for quantitative traits and provides a means to uncover genetic networks affecting these traits. Knowledge of epistatic networks will contribute to our understanding of the genetic basis of evolutionarily and clinically important traits and enhance predictive ability at an individualized level in medicine and agriculture.**

chill coma recovery | genetic interaction networks | genome-wide association studies | startle response | starvation resistance

Our understanding of genetic architecture of complex traits has been greatly advanced by genome-wide screens for DNA variants associated with phenotypic variation. To date, genome-wide association studies (GWASs) have identified over 6,000 common SNPs robustly associated with human complex traits and common diseases (1). Two general findings have emerged from these studies. First, there are typically a large number of loci associated with each trait, each of which explains a very small fraction of phenotypic variation (2). Second, loci associated with each trait collectively account for only a small proportion of genetic variation, giving rise to the mystery of missing heritability (3). Fitting all SNPs simultaneously and additively in a linear model can substantially increase the fraction of genetic variation explained by DNA variants (4), suggesting the existence of many weak associations. Another potential explanation under active investigation is that rare alleles with large effects, non-SNP variants (e.g., structural variants such as copy number variations and small indels), and nonsequence epigenetic modifications together account for the missing heritability (3). However, heritability in human studies is usually estimated as two times the

difference between the observed phenotypic correlation between monozygotic and dizygotic twins, which estimates the fraction of phenotypic variance caused by additive genetic variation as well as overestimates the variance caused by dominance and epistasis (5). The potential inflation of estimates of narrow sense heritability (i.e., heritability caused by only the additive component of genetic variance) because of genetic interactions in human studies could lead to substantial underestimates of heritability explained by DNA variants (6).

Controversy about the relative importance of epistasis in the genetic architecture of complex traits began with early formulations of quantitative genetic theory (7, 8) and continues today (9, 10). The crux of the controversy stems from the disparate goals of assessing the extent to which interactions affect mean genotypic values vs. estimating the fraction of total genetic variance caused by epistatic interactions in outbred populations. There is extensive evidence for epistatic interactions among quantitative trait loci (QTLs) affecting mean genotypic values in *Drosophila*, mice, *Arabidopsis*, yeast, and chickens (2, 9). Epistatic effects can be as large as main QTL effects, and they can occur in opposite directions between different pairs of interacting loci and between loci without significant main effects on the trait. Knowledge of interactions between loci can be used to infer genetic networks affecting complex traits (11), greatly informing the underlying biology. Epistasis is also the genetic mechanism underlying canalization (genetic homeostasis) (12, 13) and speciation (Dobzhansky-Muller incompatibilities) (14, 15); therefore, identifying interacting loci segregating in natural populations is relevant to understanding both evolutionary stasis and change. Finally, knowledge of interacting loci will improve predictions of long-term response to selection and inbreeding depression (and its converse,

---

Author contributions: S.R., R.R.H.A., R.A.G., and T.F.C.M. designed research; M.A.C., J.F.A., L.D., K.W.J., F.L., M.M.M., C.B.W., K.B., Y.H., M.J., J.J., S.N.J., D.M., F.O., L.P., Y.-Q.W., Y.Z., and X.Z. performed research; E.A.S. contributed new reagents/analytic tools; W.H., D.Z., and J.F.A. analyzed data; and W.H. and T.F.C.M. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The sequences reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) database, [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) (accession nos. SRP013733–SRP013735).

<sup>1</sup>Present address: Faculty of Arts and Sciences Society of Fellows, Harvard University, Cambridge, MA 02138.

<sup>2</sup>Present address: Shell International Exploration and Production, Inc., Westhollow Technology Center, Houston, TX 77082-3101.

<sup>3</sup>Present address: Institute for Applied Cancer Science, MD Anderson Cancer Center, Houston, TX 77054.

<sup>4</sup>To whom correspondence should be addressed. E-mail: [trudy\\_mackay@ncsu.edu](mailto:trudy_mackay@ncsu.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1213423109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1213423109/-DCSupplemental).

heterosis) in agricultural animal and crop species and individual disease risk in humans.

However, nonadditive gene action does not translate to non-additive genetic variance. Pure dominance results in mostly additive variance across the entire range of allele frequencies (5); pure epistasis gives largely additive genetic variance when allele frequencies are low, and most frequencies are low (10). In practice, all estimates of additive genetic variance (and hence, narrow sense heritability) from resemblance among relatives include fractions of the interaction variance (5). Thus, the only approach to discover epistatic interactions is a genome-wide screen in a mapping population, which incurs a severe penalty for multiple testing and hence, requires unreasonably large samples.

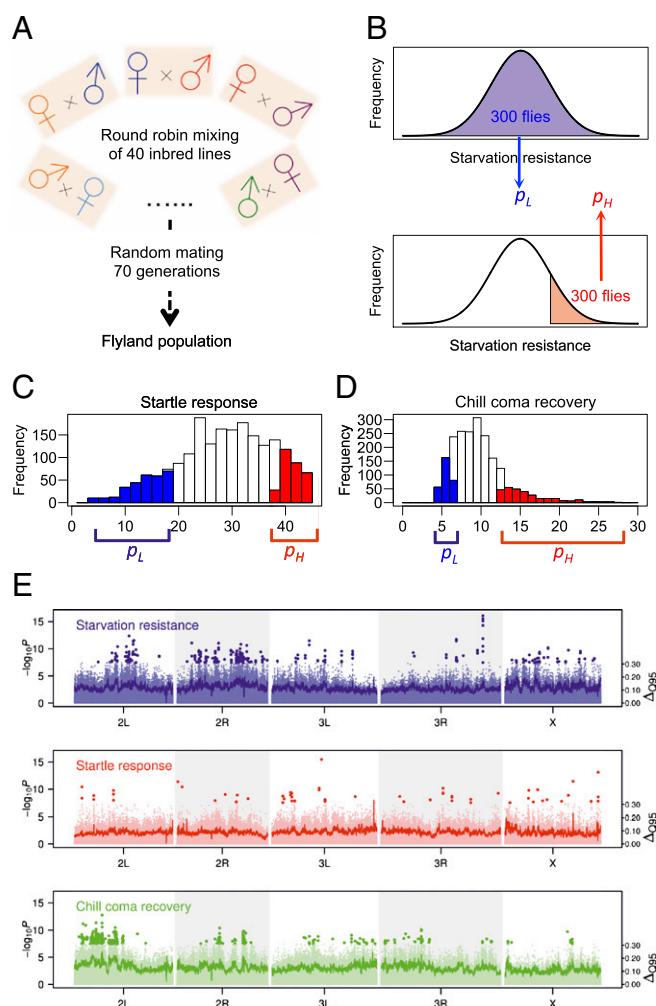
Here, we use the 168 sequenced inbred lines of the *D. melanogaster* Genetic Reference Panel (DGRP) (16) to evaluate the contribution of common and rare variants as well as additive and epistatic gene action to the genetic architecture of three complex life history traits. We constructed an outbred, advanced intercross population from 40 DGRP lines, and we assessed allele frequency changes between extremes of the distribution for each trait by deep DNA sequencing of pools of individuals. We compared the results of GWASs in the DGRP lines to changes of allele frequency between extreme scoring individuals of the outbred population. We expected that common SNPs with additive effects shared between the two populations would be significant in both and that SNPs associated with the traits that were too rare in the DGRP to include in the GWAS might be significant in only the outbred population. Furthermore, we hypothesized that common SNPs shared between the populations with epistatic effects would not replicate across populations, because with epistasis, allelic effects are expected to vary between populations with different background allele frequencies. Surprisingly, we find that all three traits have distinct genetic architectures in the two populations caused by epistasis and that genetic networks inferred from the epistatic interactions are highly interconnected.

## Results

**Extreme QTL Mapping in the Flyland Population.** We previously performed a GWAS in the DGRP to identify common SNPs associated with starvation resistance, startle response, and chill coma recovery time (16). To complement the search for trait-associated SNPs in the DGRP, we generated an advanced intercross outbred population (Flyland) from 40 DGRP lines. We crossed the 40 lines in a round robin design and subsequently, maintained the population by random mating with a large population size for over 70 generations (Fig. 1A). Thus, SNPs that were private to one of the founder DGRP lines are expected to segregate at frequencies of 2.5% in Flyland, enabling estimation of their effects.

We used extreme QTL mapping (17–20), which maps trait-associated loci by contrasting allele frequencies among individuals with extreme phenotypes, to identify SNPs associated with the three traits. Briefly, we phenotyped samples of 2,000 females from the Flyland population for starvation resistance, startle response, and chill coma recovery. For each trait, we created two pools of female flies. For starvation resistance, the 15% longest surviving flies constituted one pool, whereas the other was a random sample of 300 females (Fig. 1B). For startle response and chill coma recovery, the pools consisted of the top and bottom 15% of the phenotypic distribution (Fig. 1C and D).

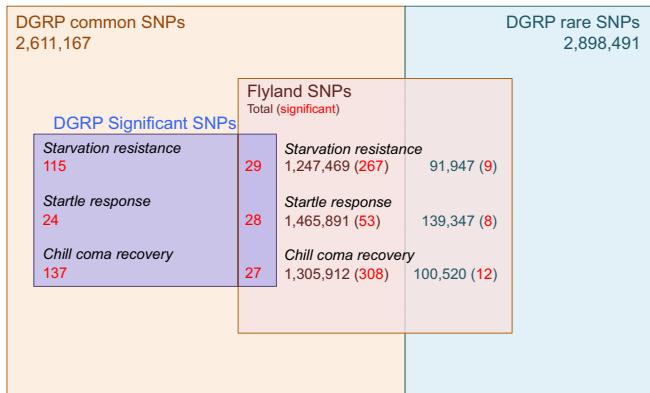
We sequenced DNA from each of these pools to at least ~300× (Dataset S1, Table S1). Among the segregating SNPs in the DGRP, we found 1,339,448, 1,605,264, and 1,406,458 to be segregating in Flyland for the starvation resistance, startle response, and chill coma recovery pools, respectively. We estimated SNP allele frequencies in the sequenced DNA pools using counts of high-quality sequencing reads matching the alleles, and we computed the difference in allele frequency between the two DNA



**Fig. 1.** Extreme QTL mapping in Flyland. (A) The Flyland population was generated by a round robin crossing design. Progeny of these crosses was mixed and outcrossed randomly for 70 generations. The last generation constituted the Flyland population. (B) For starvation resistance, a random sample of 300 and the 300 longest-living female flies constituted the sequenced pools. (C) Distribution of startle response for 2,000 phenotyped female flies in Flyland. (D) Distribution of chill coma recovery for 2,000 female flies in Flyland. The top (red) and bottom (blue) 15% scoring flies were pooled and sequenced. Allele frequencies were estimated in the high and low pools and compared to identify QTLs. (E) Plots of significance ( $-\log_{10} P$ , left axis) and allele frequency differences ( $\Delta Q_{95}$ ; right axis) for extreme QTL mapping. Points of dark color indicate significant SNPs. The dark-colored lines connect consecutive sliding windows of 50 kb, with a step size of 5 kb. In each window, the 95% quantile of the absolute allele frequency difference is plotted.

pools for each trait. We identified 276, 61, and 320 SNPs with significant allele frequency differences between the extreme pools for starvation resistance, startle response, and chill coma recovery, respectively, at a stringent Bonferroni genome-wide threshold of 0.05 (Figs. 1E and 2 and Dataset S1, Tables S2–S4). The majority of significant SNPs had moderate effects, with allele frequency differences of ~0.25 in the extreme pools (SI Appendix, Fig. S1). The significant SNPs were distributed throughout the genome (Fig. 1E), indicating multigenic and complex genetic architecture for the traits in the Flyland population.

**Distinct Genetic Architecture of Complex Traits in the DGRP and Flyland.** We next asked if the significant SNPs identified in Flyland recapitulated the genetic architecture determined by



**Fig. 2.** No overlap between significant SNPs in Flyland and the DGRP. Numbers of SNPs overlapping between different categories of SNPs are shown in boxes. Numbers within each box are for SNPs in the corresponding category. SNPs belonging to multiple categories are indicated by overlaps between boxes.

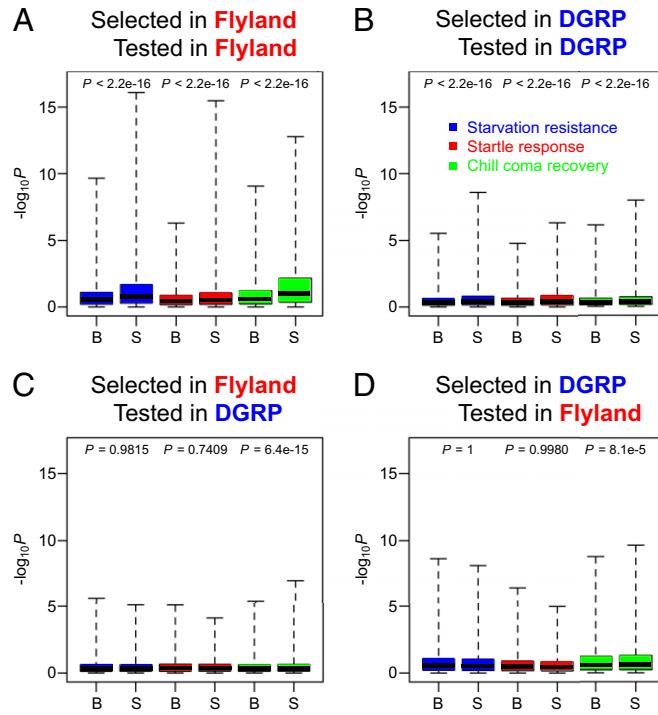
GWAS in the DGRP and vice versa. Among the significant Flyland SNPs, 267 (starvation resistance), 53 (startle response), and 308 (chill coma recovery) were also tested in the DGRP GWAS (16), but none were even nominally ( $P < 10^{-5}$ ) significant in this analysis (Fig. 2). The remaining 9 (starvation resistance), 8 (startle response), and 12 (chill coma recovery) SNPs were too rare (with fewer than four lines with the minor alleles) to be included in the DGRP GWAS. However, 144 (starvation resistance), 52 (startle response), and 164 (chill coma recovery) SNPs had  $P$  values less than  $10^{-5}$  in the DGRP female GWAS (Fig. 2) (16). Of these SNPs, 29 (starvation resistance), 28 (startle response), and 27 (chill coma recovery) were tested in Flyland, and none reached statistical significance. The remaining SNPs were of relatively low frequency in the DGRP and either became fixed for one allele in Flyland or were too rare to be distinguished from sequencing error.

The complete lack of overlap between significant SNPs in Flyland and the DGRP cannot be explained by thresholding  $P$  values, because relaxing thresholds in either analysis gives no more overlap than expected by chance (SI Appendix, Fig. S2). In addition, we estimated the statistical power to detect the DGRP SNPs in Flyland and vice versa by simulation. Although Flyland did not always possess sufficient power to detect the DGRP SNPs, particularly for those SNPs with small effect sizes and low allele frequency, the DGRP was adequately powered (>50%) to detect the majority of Flyland SNPs (SI Appendix, Fig. S3). Moreover, there was no correlation between the ranks of  $P$  values for SNPs tested in the two populations [Spearman's correlation  $p = -0.003$  (starvation resistance),  $p = 0.001$  (startle response), and  $p = 0.002$  (chill coma recovery)]. Additionally, signs and magnitudes of effects of alleles were not qualitatively similar for SNPs that were significant in one population but not the other. We compared the estimated effects for minor alleles of SNPs in the DGRP and the change of frequency for the same allele in the high pool relative to the low pool for each trait in Flyland. We found no correlation between the effects of the same alleles in the Flyland and DGRP populations for all three traits (SI Appendix, Fig. S4).

We next investigated the possibility that different SNPs were associated with the traits in Flyland and the DGRP, because clusters of closely linked SNPs were associated with the traits, and different representatives of each cluster were identified in the different analyses. We selected SNPs that were within 10 kb of each significant SNP in Flyland and tested them for association with the same trait in the Flyland population. As expected with local linkage disequilibrium (LD) that still remains after

crossing inbred lines for 70 generations (SI Appendix, Fig. S5), SNPs near significant signals were generally more strongly associated with traits than randomly selected SNPs (Fig. 3A and SI Appendix, Fig. S6). The same was also true for SNPs in close physical proximity to SNPs with significant signals in the DGRP population (Fig. 3B and SI Appendix, Fig. S6). Conversely, when we tested SNPs close to significant Flyland signals for starvation resistance and startle response in the DGRP population, they did not differ significantly from randomly selected SNPs (Fig. 3C and SI Appendix, Fig. S6). For chill coma recovery, there was a small but statistically significant enrichment of SNPs with stronger association in the DGRP around Flyland signals (Fig. 3C). The same relationships were also observed for SNPs selected from the DGRP analyses when tested in Flyland (Fig. 3D and SI Appendix, Fig. S6). However, analyses of chill coma recovery in both the Flyland and DGRP populations revealed a disproportionate number of significant SNPs on chromosome 2L, and therefore, it is difficult to disentangle effects caused by LD from overrepresentation of significant SNPs on this chromosome arm. We tested whether the enrichment was solely induced by LD by narrowing the window size for selecting SNPs, which should strengthen the enrichment. This was not observed (SI Appendix, Fig. S7), suggesting that the enrichment was likely caused by the disproportionate numbers of SNPs on chromosome 2L that have a large effect on chill coma recovery in both populations.

Although there was no overlap of SNPs associated with the three complex traits in the Flyland and DGRP populations, we considered the possibility that the SNPs, nevertheless, were associated with the same genes and pathways. We compiled lists of



**Fig. 3.** Significance of SNPs around association signals selected in Flyland and the DGRP. (A) SNPs within 10 kb of significant Flyland SNPs were tested for associations with each trait in Flyland. The same number of background SNPs (B) was randomly drawn from the genome and tested for associations with the traits.  $P$  values are from Wilcoxon tests of the difference between the level of significance in selected (S) and background (B) sets. Similar analyses were also performed for (B) SNPs selected around DGRP signals and tested in the DGRP, (C) SNPs selected around Flyland signals and tested in the DGRP, and (D) SNPs selected around DGRP signals and tested in Flyland.

genes from FlyBase Release 5.44 (21) that contained at least 1 of the 1,000 most significant SNPs in the Flyland and DGRP populations that were located within 50 bp of exons and 1,000 bp of transcription start sites. *P* value cutoffs for the top 1,000 SNPs were  $6.5 \times 10^{-7}$  (starvation resistance),  $1.0 \times 10^{-5}$  (startle response), and  $2.8 \times 10^{-7}$  (chill coma recovery) in Flyland and  $1.4 \times 10^{-4}$  (starvation resistance),  $2.6 \times 10^{-4}$  (startle response), and  $1.1 \times 10^{-4}$  (chill coma recovery) in the DGRP. This result led to the identification of 282, 237, and 292 genes in Flyland and 230, 220, and 247 genes in the DGRP for starvation resistance, startle response, and chill coma recovery, respectively. A total of 2 (starvation resistance), 7 (startle response), and 10 (chill coma recovery) genes were in common between Flyland and the DGRP. The overlap was not significant for starvation resistance (*permutation P* = 0.950) and startle response (*P* = 0.059) but was nominally significant for chill coma recovery (*P* = 0.028).

We next assessed whether there was overrepresentation of Gene Ontology (GO) terms among the genes associated with each trait in Flyland and the DGRP individually as well as in the combined gene set. At a False Discovery Rate (FDR) < 0.10, we found no overrepresentation of biological process and molecular function terms for starvation resistance. We did find small enrichment for several GO terms for startle response and chill coma recovery (Dataset S1, Table S5), including asymmetric neuroblast division for startle response (*P* =  $2.6 \times 10^{-4}$ ). Among the 29 annotated genes for this GO term, 4 genes were associated with startle response in Flyland. One of these four genes and an additional two genes were associated with startle response in the DGRP, leading to a significant enrichment of this term in the combined set of significant genes in Flyland and the DGRP. This finding suggests that different candidate genes within a common biological process affect startle response in the two populations. Nonetheless, such enrichment was rare, and the magnitude was very small.

Taken together, our results clearly show that association analyses in the Flyland and DGRP populations detected largely distinct QTLs for the three traits, indicating that genetic architecture at the level of individual QTLs is highly background-specific in the two populations.

**Widespread Epistasis for SNPs Associated with Complex Traits in Flyland and the DGRP.** Both the single-marker regression association analyses in the DGRP and the extreme QTL mapping analyses in Flyland detect additive effects of individual SNPs, but they do not distinguish between SNPs with strict additive gene action independent of genotypes at other loci and apparent additive effects induced by epistasis. Purely epistatic models can result in covariance between relatives (5) and a substantial amount of apparent additive variance (10). Therefore, we tested whether SNPs associated with complex traits in Flyland and the DGRP had epistatic effects.

We performed GWASs for pairwise interactions for each trait in the DGRP population in which the focal SNP was one of the significant SNPs in the Flyland and DGRP populations, and the GWAS tested for interactions with all other SNPs in the DGRP. We restricted these analyses to pairs of SNPs for which at least two lines were represented in the minor haplotype class. Because the sample size is relatively small, we chose a liberal significance threshold of *P* <  $1 \times 10^{-5}$  for the interaction term to summarize the results. Rather than making inferences on individual significant interactions, we investigate the overall architecture of epistasis, tolerating occasional false positives. Remarkably, the majority of significant SNPs in either Flyland or the DGRP interacted with at least one other SNP in the DGRP for all three traits (Table 1 and Dataset S1, Table S6–S8). Most of the significant SNPs for which we could not detect epistasis were relatively rare variants, and thus, all four haplotypes were not represented in the DGRP by at least two lines.

Our observation of extensive epistasis shows that estimates of additive effects of individual SNPs are highly context-dependent. In the simplest form of epistasis, illustrated in Fig. 4, a change of allele frequency of the context SNP can significantly affect the estimated main effect of the SNP being tested. Therefore, we asked whether SNPs that interacted with SNPs significantly associated with each trait differed in allele frequency between Flyland and the DGRP. Among the 56,653 interacting SNPs, 8,918 had significant (FDR < 0.05) and large (FDR > 0.25) changes in population allele frequencies. Furthermore, SNPs significantly associated with traits in single-marker analyses generally interacted with multiple additional SNPs, illustrating higher-order epistatic interactions. Thus, it is unlikely that the main effect of an SNP is determined by the allele frequency of a single interacting SNP. Nevertheless, widespread epistasis and changes in population allele frequencies suggest that genetic architecture is highly dynamic and sensitive to the exact allelic composition in the population.

**Epistatic Interactions Reveal Genetic Networks.** We hypothesized that epistatic interactions between SNPs reflect underlying genetic networks. To test this hypothesis, we mapped all interacting SNPs to annotated genes. First, we asked whether genes interacting with genes harboring SNPs associated with the traits were common between Flyland and the DGRP. We found substantial overlap (*P* < 0.001) for starvation resistance and chill coma recovery and little but nominally significant overlap for startle response (SI Appendix, Fig. S8). The lack of overlap for startle response is likely attributable to the relatively small number of genes associated with this trait. The overlapping interacting genes formed highly interconnected networks (Fig. 5), suggesting that common biological networks were perturbed through epistatic interactions in both Flyland and the DGRP, despite the lack of overlap of individually associated genes. Second, we assessed whether epistatic genes were enriched for GO terms and found several biologically plausible GO terms significant at FDR < 0.10 for all three traits (Dataset S1, Table S9). Third, we asked if the epistatic genes were overrepresented in the signaling and metabolic pathways from the Reactome and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (22). We found significant enrichment of subnetworks for epistatic genes associated with starvation resistance (*P* < 0.005) (SI Appendix, Fig. S9A) and chill coma recovery (*P* < 0.005) (SI Appendix, Fig. S9B). Collectively, these results show that epistatic interactions recapitulate known as well as candidate genetic networks affecting complex traits.

## Discussion

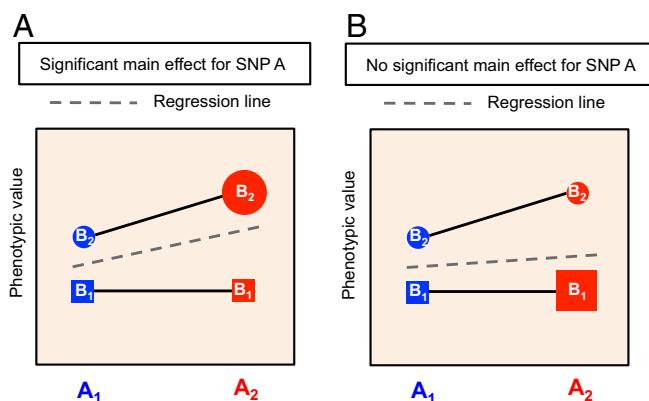
Using the rapid and powerful extreme QTL mapping approach, we identified a large number of SNPs associated with three complex traits in an outbred population (Flyland) derived from a subset of 40 DGRP lines. Surprisingly, none of the SNPs identified in the Flyland population replicated associations found for these traits in the DGRP (16). Although a few of the SNPs associated with the traits in the Flyland population were too rare in the DGRP to be included in the association analyses, most of the SNPs were common in both populations. Our systematic evaluation of potential sources of background-specific effects provided strong evidence that the genetic architectures in the two populations were distinct at the QTL, gene, and biological pathway levels when associations were considered individually. Such a high degree of context dependence of genetic architecture for complex traits has important implications. First, it suggests that genetic associations need to be interpreted within the context or background in which they are identified. Second, it immediately leads to questions about the nature of the contexts and how they modify genetic architecture.

**Table 1.** Epistasis between trait-associated SNPs and other SNPs in the genome

Trait	Trait-associated common SNPs (DGRP + Flyland)	Significant epistatic pairs	Trait-associated SNPs with at least one significant interacting SNP (Flyland + DGRP)
Starvation resistance	411 (144 + 267)	19,629	288 (57 + 231)
Startle response	105 (52 + 53)	1,252	48 (11 + 37)
Chill coma recovery	472 (164 + 308)	57,863	314 (79 + 235)

There are many characteristics of a population that can define context and for which the Flyland and DGRP populations differ. First, the Flyland population is outbred, and thus, it has heterozygous as well as homozygous individuals at each locus and an additional source of genetic variability because of dominance and/or overdominance. Dominance and overdominance may explain the inability to replicate the DGRP signals in Flyland, because extreme QTL mapping is intrinsically less powerful for these modes of gene action. However, they do not explain the failure of associations in the DGRP to replicate Flyland signals, because SNPs detected by extreme QTL mapping have strong additive effects, which should also be detectable in the DGRP (*SI Appendix*, Fig. S3). Second, the pattern of local LD may differ between Flyland and the DGRP, which would lead to different patterns of associations, particularly at noncausative loci in LD with causal variants. There is little local LD in the DGRP (16) (*SI Appendix*, Fig. S5), and after 70 generations, the LD in Flyland would increase slightly but remain very small (*SI Appendix*, Fig. S5). Furthermore, our association analyses were nearly exhaustive for all variants identifiable by sequencing, including causal variants. However, even when LD was taken into account, distinct genetic architectures were found in the two populations (Fig. 3 and *SI Appendix*, Fig. S6). Third, the composition and site frequency distribution of alleles differ between Flyland and the DGRP. The DGRP has more SNPs than Flyland, whereas Flyland has a larger population size and hence, more potential multilocus genotypes. A large number of SNPs has substantial differences in allele frequency between Flyland and the DGRP (*SI Appendix*, Fig. S10). Some of these differences are caused by the selection of 40 lines from the DGRP, whereas others are caused by genetic drift, and potentially natural selection.

Changes in allele frequencies at a locus can alter the estimated effect of a second locus if the two loci interact epistatically (Fig. 4).

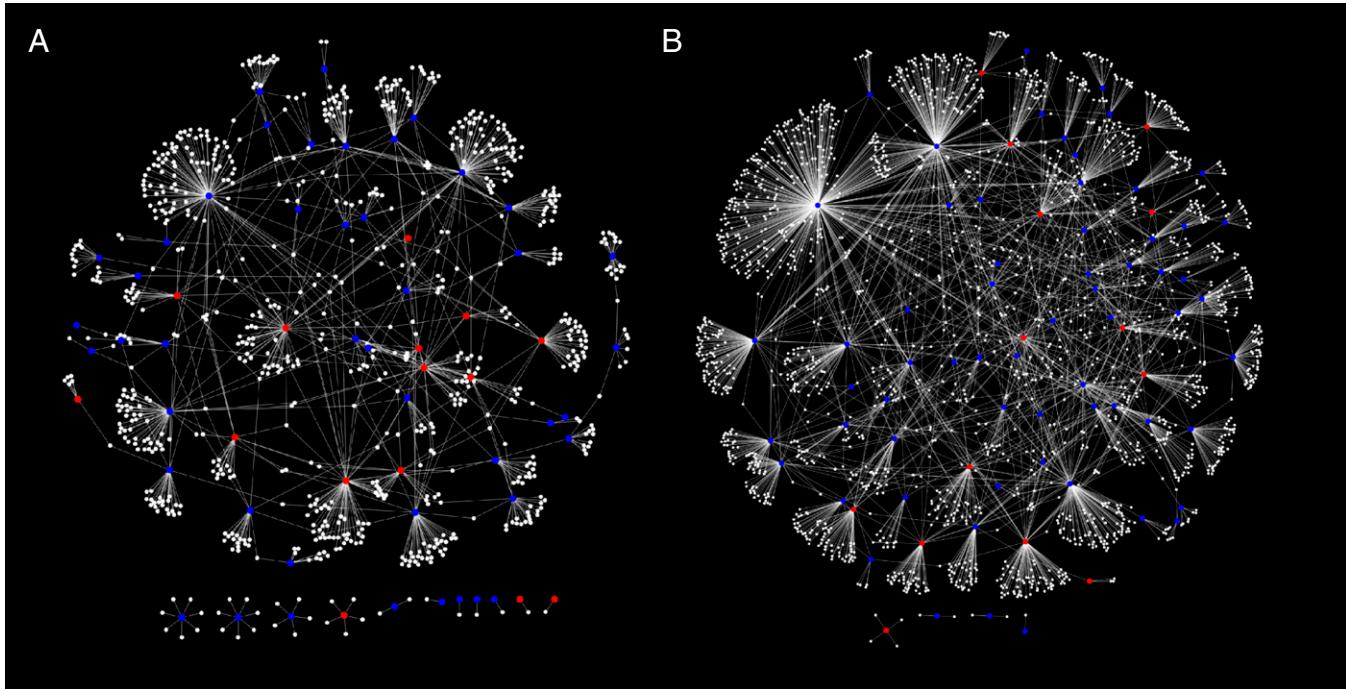


**Fig. 4.** Context-dependent additive effects. A significant epistatic interaction is shown between two SNPs, A and B, each with two alleles (indicated by subscripts 1 and 2). There is no difference in the effect of the two alleles of SNP A in the B<sub>1</sub> genetic background but a large difference in the B<sub>2</sub> background. The sizes of the symbols are proportional to the genotype frequency. (A) SNP A has a large main effect (shown by the regression line) when the A<sub>2</sub>B<sub>2</sub> genotype has a high frequency. (B) The main effect of SNP A is diminished when the A<sub>2</sub>B<sub>1</sub> genotype has a high frequency.

We showed that the majority of SNPs associated with each trait was involved in at least one pairwise interaction with other SNPs. Although we focused our analysis of epistasis on only the SNPs with marginal additive effects in either of the two populations, the large number of tests performed gave the search essentially no statistical power to declare significance of any particular interaction. However, the goal of this analysis was not de novo discovery of epistasis but rather, determination of whether the epistatic models could better explain the phenotypic variation given that the identity of one of the candidate SNPs was known. Therefore, we used the same nominal significance threshold as the single SNP GWAS in the DGRP (16) to enrich for even weak epistatic interactions and analyzed global patterns of epistasis rather than individual significance. Because our search for epistasis was restricted to the DGRP, we could only evaluate the effects of additive by additive interactions.

The genetic networks perturbed by epistatic interactions with SNPs with significant main effects in Flyland or the DGRP were highly connected. Thus, a major result that emerges from our analysis is that, despite the apparent additive effects at individual loci pointing to distinct genes in the two populations, they actually perturb common genetic networks through epistatic interactions. Genes in these genetic networks were enriched for biologically plausible processes and functions as well as signaling and metabolic pathways. None of these observations would be expected if the interactions were largely false positives. Because we only investigated a small fraction of possible pairwise associations and the high network connectivity implies even higher-order interactions, we conclude that the genetic architecture of these quantitative traits is dominated by extensive epistasis.

There has been controversy about the contribution of epistasis to the genetic architecture of quantitative traits and response to directional selection. Numerous studies have reported evidence for epistatic gene action affecting quantitative traits in humans, plants, animals, and flies (2). However, epistatic gene action causes largely additive genetic variance in natural populations in which the allele frequency spectrum is U-shaped (10), and hence, it is thought to be of little importance in predicting the response to directional selection (23). Although the statistical property of additive variance is useful for genetic improvement of agricultural species, it remains important to distinguish between genetic variation that is caused by epistasis and the fraction of genetic variation that can be statistically explained by an additive model. In fact, variation induced by all of the epistatic interactions identified in the present study could be largely explained by the marginal additive effects at the trait-associated loci. Indeed, statistical models of the genetic architecture of starvation resistance and startle response showed no evidence for departure from a purely additive model (24). However, the predictive ability of genome selection models for these traits (24) was not high (0.23–0.24), which is expected with extensive epistasis, because allele frequencies and hence, allelic effects will be different between the populations used to derive the model and the population in which the model is tested. Furthermore, for selection to work primarily on additive variance, it has to act on the population as a whole. Consider, for example, a trait for which loci A and B are interacting antagonistically. Selecting for this trait in extreme subpopulations in which the alternative alleles of locus B are



**Fig. 5.** Networks of epistatic interactions. Interaction networks are depicted for (A) starvation resistance and (B) chill coma recovery. Nodes depict genes, and edges significant interactions. Red nodes are genes containing significant SNPs from the Flyland analysis. Blue nodes are genes containing significant SNPs from DGRP analysis.

fixed will result in changes of allele frequency at locus A in opposite directions.

Although our analysis did not explore the entire spectrum of possible epistatic interactions, we have shown that the majority of background-specific additive effects can be equally well or better explained by epistatic interactions between genes that are organized into highly connected genetic networks. An appreciation for and knowledge of the contribution of epistatic gene action to quantitative trait phenotypes is required for understanding the molecular mechanisms of variation for quantitative traits and developing predictive models at the individualized level. We speculate that epistatic gene action is also an important feature of the genetic architecture of quantitative traits in other organisms, including humans. Our analysis paradigm (first identifying loci associated with a quantitative trait in two populations with different allele frequencies and then using these loci as foci for a genome-wide screen for pairwise epistatic interactions) can be applied to any organism for which such populations exist. For example, human GWASs have been plagued by a lack of replicated associations across populations in even large studies (25, 26). We argue that this finding is expected under epistatic gene action and variable allele frequencies. This hypothesis is testable by analysis of these data using our framework.

In the future, understanding the full extent of genetic interaction networks for quantitative traits will enrich our understanding of their underlying biology.

## Methods

**Creation of Flyland and Phenotyping.** We crossed 40 DGRP lines (RAL\_208, RAL\_301, RAL\_303, RAL\_304, RAL\_306, RAL\_307, RAL\_313, RAL\_315, RAL\_324, RAL\_335, RAL\_357, RAL\_358, RAL\_360, RAL\_362, RAL\_365, RAL\_375, RAL\_379, RAL\_380, RAL\_391, RAL\_399, RAL\_427, RAL\_437, RAL\_486, RAL\_514, RAL\_517, RAL\_555, RAL\_639, RAL\_705, RAL\_707, RAL\_712, RAL\_714, RAL\_730, RAL\_732, RAL\_765, RAL\_774, RAL\_786, RAL\_799, RAL\_820, RAL\_852, and RAL\_859) in a round robin design, giving 40 F1 genotypes. We crossed females of RAL\_208 to males of RAL\_301, females of RAL\_301 to males of RAL\_303, and so on until females of RAL\_859 were crossed to males of RAL\_208. We placed one male

and one female of each genotype into each of 10 bottles, removing the parents after 2 d. For each subsequent generation, we collected 40 males and 40 females from each bottle and randomly allocated them to 10 bottles to minimize genetic drift. Thus, the census size of the Flyland population is  $n = 800$  per generation. Beginning at generation 70, we collected 6,000 Flyland females and phenotyped 2,000 flies for startle response, chill coma recovery, and starvation resistance exactly as described previously (16). We collected the 300 top- and 300 bottom-scoring individuals for startle response and chill coma recovery time and 300 unstarved plus 300 longest lived individuals from the starvation treatment, and we froze them for subsequent DNA extraction.

**Genomic DNA Isolation.** Genomic DNA was extracted from 100 mg flies using Genomic-tip 100/G columns (Qiagen) according to the manufacturer's instructions. The flies were ground in liquid nitrogen using a mortar and pestle to a fine white powder. Digestion Buffer (19 mL Buffer G2; Qiagen), 75  $\mu$ L RNase A (20 mg/mL; 5 Prime), and 600  $\mu$ L Proteinase K (20 mg/mL; 5 Prime) were added to the homogenized flies and incubated at 55 °C for 2 h. Following centrifugation, the lysate was loaded onto the 100/G columns that had been equilibrated with Buffer QBT (Qiagen). The column was washed two times with Buffer QC (Qiagen), and the DNA was eluted with prewarmed (55 °C) elution buffer (Buffer QF; Qiagen). The DNA was precipitated with 100% isopropanol, and the pellet was washed with 70% ethanol. The DNA pellet was rehydrated in Tris-EDTA buffer (pH 8.0; Ambion) and quantified using picogreen on the Qubit fluorometer (Invitrogen).

**Illumina Library Construction.** We sequenced each of the six extreme genomic DNA samples to a total of 300 $\times$  coverage with six technical replicates of 50 $\times$  coverage each to generate a total of 36 libraries. Libraries were barcoded using Illumina Multiplexing oligos (Illumina), enabling two to be sequenced per HiSeq. 2000 lane.

High-molecular weight double-strand genomic DNA samples were used to construct Illumina paired-end libraries according to the manufacturer's protocol (Illumina). Briefly, 1  $\mu$ g genomic DNA in 100  $\mu$ L volume was sheared into fragments of ~300 bp with the Covaris S2 or E210 system (settings: 10% duty cycle, intensity of 4,200 cycles per burst, 90 s; Covaris). Fragments were processed through DNA End-Repair in 100  $\mu$ L containing sheared DNA, 10  $\mu$ L 10x buffer, 5  $\mu$ L End Repair Enzyme Mix, and H<sub>2</sub>O (NEBNext End-Repair Module) at 20 °C for 30 min; A-tailing was performed in 50  $\mu$ L containing End-Repaired DNA, 5  $\mu$ L 10x buffer, and 3  $\mu$ L Klenow Fragment (NEBNext dA-Tailing Module) at 37 °C for 30 min, and each step was followed by

purification using a QIAquick PCR purification kit (Qiagen). Resulting fragments were ligated with Illumina Index PE adaptor oligo mix and the NEB-Next Quick Ligation Module. After ligation, size selection was carried out by using 3% Ready agarose gels running in 1x TBE (BioRad). Gel slices were excised from ~370 to 420 bp, and the size-selected DNA was purified using a Qiagen MinElute gel extraction kit and eluted in 30 µL EB buffer (Qiagen). PCR with Illumina PE Index primers was performed in 50-µL reactions containing 25 µL 2x Phusion High-Fidelity PCR master mix, 10–20 ng size-selected fragment DNA, 1.0 µL each primer, and H<sub>2</sub>O. The standard thermocycling for PCR was 30 s at 98 °C for the initial denaturation followed by 12–18 cycles of 10 s at 98 °C, 30 s at 65 °C, and 30 s at 72 °C and a final extension of 5 min at 72 °C. Agencourt XP Beads (Beckman Coulter Genomics) were used to purify the PCR products. After bead purification, PCR products were quantified using PicoGreen, and their size distribution was analyzed using the Agilent Bioanalyzer 2100 DNA Chip 7500; 15-µL aliquots at 10 nM of each pool of two libraries were passed for Illumina sequencing.

**Sequencing.** Shotgun DNA libraries were sequenced on Illumina's HiSeq2000 with V3 chemistry. Briefly, sequencing libraries were quantified with an Agilent 2100 Bioanalyzer. Cluster generations were performed on an Illumina C-Bot with Illumina's pair end flow cell; 2 × 100 cycles of sequencing with 7 cycles index sequencing were carried out according to the manufacturer's standard protocol. Two barcoded libraries are run together in one lane on the HiSeq flow cell. Imaging analysis and base calling were done with RTA software on HiSeq2000. CASAVA 1.7 was then used to demultiplex the sequences into a set of fastq files that were used in the mapping analysis.

**Extreme QTL Mapping.** Illumina sequence reads were aligned to the Dmel 5.13 reference genome with the Burrows-Wheeler Aligner (BWA) (27) using default parameters. GATK (28) software was used to locally realign regions around indels, remove duplicate sequence reads, and recalibrate base quality scores. Local realignment was performed on the combined BAM files of the two pools for each trait. Alignments were piled up at each base position in the genome by SAMTools (29). We considered SNPs segregating in DNA pools for each trait by the following criteria: alleles were present in the founding strains; coverage of Q13 bases was between 20 and 1,500; at least 80% of the coverage was at least Q13; the two most frequent alleles constituted at least 95% of all observed alleles; minor alleles were present by at least 2.5% in one of the pools; the Chernoff bound of the *P* value for

the null hypothesis that the observed minor alleles were caused by sequencing error (30) was smaller than 10<sup>-5</sup>; and strand bias was not significant (*P* > 10<sup>-5</sup>) in both pools. Allele frequencies were estimated by calculating the proportion of reads supporting the alleles. We tested for a difference in allele frequencies between the two pools by computing

$$Z = \frac{(p_1 - p_2)}{\sqrt{p_0(1-p_0)\left(\frac{1}{n} + \frac{1}{d_1} + \frac{1}{d_2}\right)}}. P \text{ values were obtained by comparing the } Z \text{ statistics to the standard normal distribution. } p_1 \text{ and } p_2 \text{ are the estimated allele frequencies in the high and low pools, respectively; } p_0 \text{ is the allele frequency under the null hypothesis } H_0: p_1 = p_2 \text{ estimated from the average of } p_1 \text{ and } p_2; n \text{ is the number of flies } (n = 300) \text{ in the pools; and } d_1 \text{ and } d_2 \text{ are the sequencing depths for the high and low pools, respectively.}$$

**Identification of Epistatic Interactions.** We used FastEpistasis (31) to test for pairwise epistasis between significant SNPs in Flyland and the DGRP with other SNPs in the DGRP genome. Results from FastEpistasis were refined to have *P* values from an *F* distribution as opposed to the default asymptotic  $\chi^2$  test, and they were furthered filtered for epistatic interactions for which the two locus genotype was represented by at least two lines.

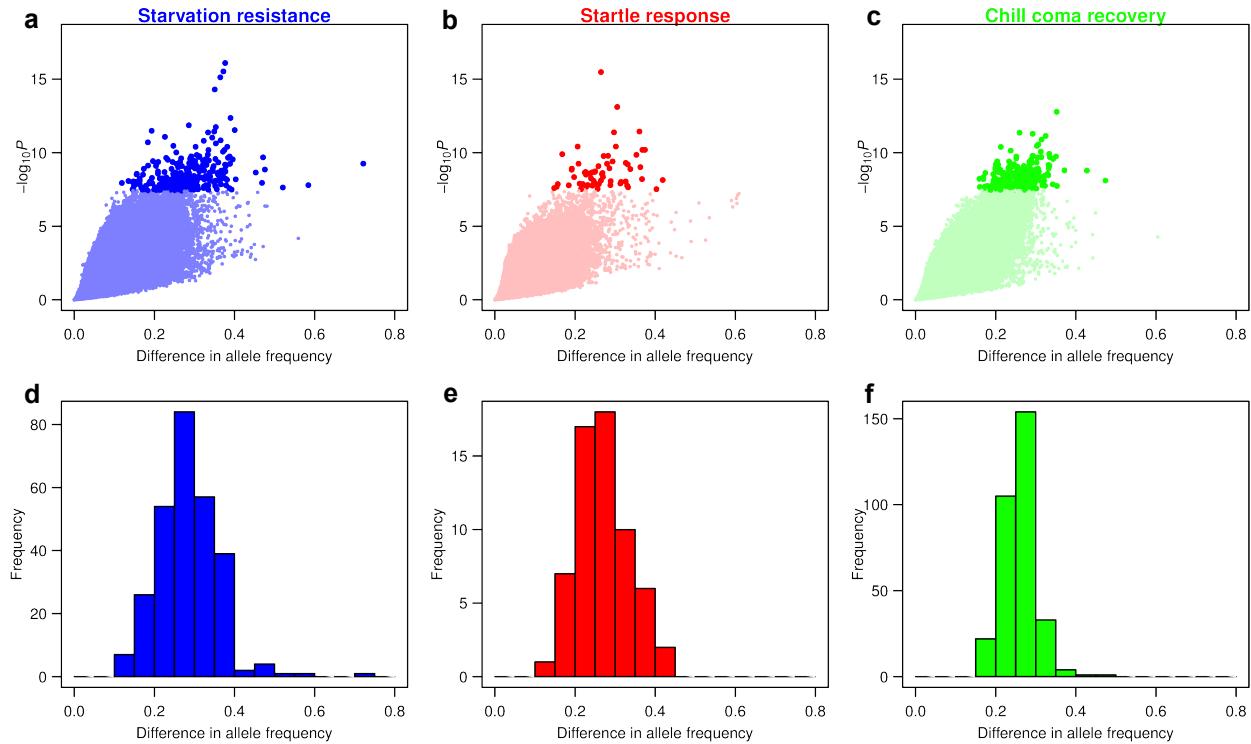
**Annotation of SNPs and Analysis of Trait-Associated and Epistatic Genes.** We annotated SNPs using the FlyBase annotation (release 5.44) (21). We considered SNPs within 1,000 bp of the transcription start site and within 50 bp of exons to be in a gene. When SNPs fell within more than one gene, we annotated the SNPs with multiple genes. For GO analyses, we obtained GO annotations of genes from FlyBase and performed hypergeometric tests for enrichment of genes in all GO terms that contained at least 20 genes in the background gene list, which consisted of genes that contained at least one segregating SNP in the DGRP. We identified subnetworks enriched for epistatic genes using the R-spider web server (22). R-spider compiles the global network from KEGG and Reactome and tests for significance of the subnetwork with the maximal number of input genes through a Monte Carlo simulation-based inference (22). We considered only the model where epistatic genes were directly connected without any missing nodes.

**ACKNOWLEDGMENTS.** This work is a publication of the W. M. Keck Center for Behavioral Biology. This work was funded by National Institutes of Health Grants R01 GM 45146 and R01 GM 59469 (to R.R.H.A. and T.F.C.M.) and National Human Genome Research Institute Grant U54 HG003273 (to R.A.G.).

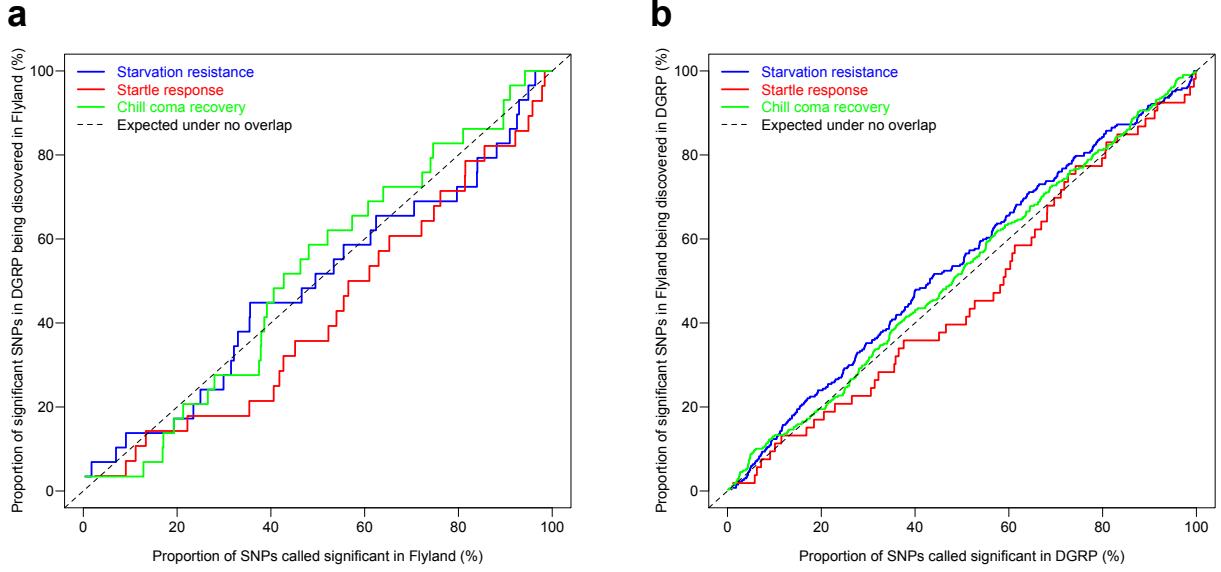
1. Hindorff LA, et al. A Catalog of Published Genome-Wide Association Studies. Available at [www.genome.gov/gwastudies](http://www.genome.gov/gwastudies). Accessed May 8, 2012.
2. Flint J, Mackay TF (2009) Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res* 19:723–733.
3. Manolio TA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753.
4. Yang J, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42:565–569.
5. Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics* (Addison Wesley Longman, Essex, England), 4th Ed.
6. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci USA* 109: 1193–1198.
7. Fisher RA (1930) *The Genetical Theory of Natural Selection* (Clarendon, Oxford).
8. Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159.
9. Carlborg O, Haley CS (2004) Epistasis: Too often neglected in complex trait studies? *Nat Rev Genet* 5:618–625.
10. Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4:e1000008.
11. Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9:855–867.
12. Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150:563–565.
13. Waddington CH (1957) *The Strategy of Genes* (George Allen & Unwin, London).
14. Dobzhansky T (1937) *Genetics and the Origin of Species* (Columbia Univ Press, New York).
15. Muller HJ (1940) *Bearing of the Drosophila Work on Systematics. The New Systematics*, ed Huxley JS (Clarendon, Oxford).
16. Mackay TF, et al. (2012) The *Drosophila melanogaster* genetic reference panel. *Nature* 482:173–178.
17. Werner JD, et al. (2005) FRIGIDA-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. *Genetics* 170:1197–1207.
18. Wolyn DJ, et al. (2004) Light-response quantitative trait loci identified with composite interval and eXtreme array mapping in *Arabidopsis thaliana*. *Genetics* 167:907–917.
19. Lai CQ, et al. (2007) Speed-mapping quantitative trait loci using microarrays. *Nat Methods* 4:839–841.
20. Ehrenreich IM, et al. (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464:1039–1042.
21. McQuilton P, St Pierre SE, Thurmond J; FlyBase Consortium (2012) FlyBase 101—the basics of navigating FlyBase. *Nucleic Acids Res* 40:D706–D714.
22. Antonov AV, Schmidt EE, Dietmann S, Krestyaninova M, Hermjakob H (2010) R spider: A network-based analysis of gene lists by combining signaling and metabolic pathways from Reactome and KEGG databases. *Nucleic Acids Res* 38:W78–W83.
23. Crow JF (2010) On epistasis: Why it is unimportant in polygenic directional selection. *Philos Trans R Soc Lond B Biol Sci* 365:1241–1244.
24. Ober U, et al. (2012) Using whole-genome sequence data to predict quantitative trait phenotypes in *Drosophila melanogaster*. *PLoS Genet* 8:e1002685.
25. Ioannidis JP (2009) Population-wide generalizability of genome-wide discovered associations. *J Natl Cancer Inst* 101:1297–1299.
26. Visscher PM (2008) Sizing up human height variation. *Nat Genet* 40:489–490.
27. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
28. McKenna N, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303.
29. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
30. Bansal V (2010) A statistical method for the detection of variants from next-generation resequencing of DNA pools. *Bioinformatics* 26:i318–i324.
31. Schüpbach T, Xenarios I, Bergmann S, Kapur K (2010) FastEpistasis: A high performance computing solution for quantitative trait epistasis. *Bioinformatics* 26: 1468–1469.

# **Epistasis Dominates the Genetic Architecture of *Drosophila* Quantitative Traits**

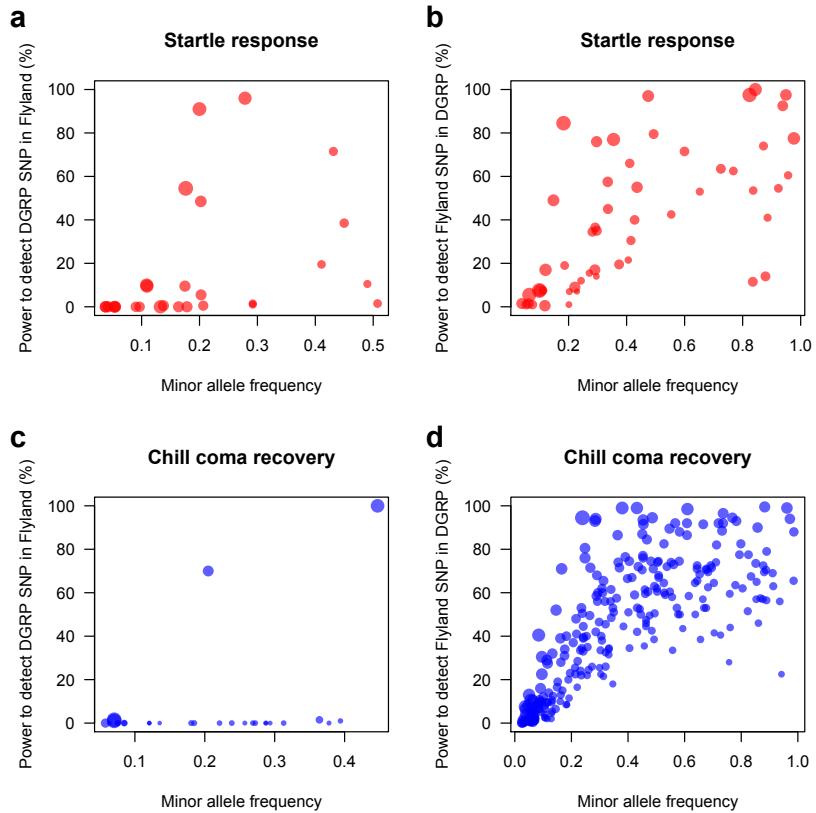
## **Supplemental Figures**



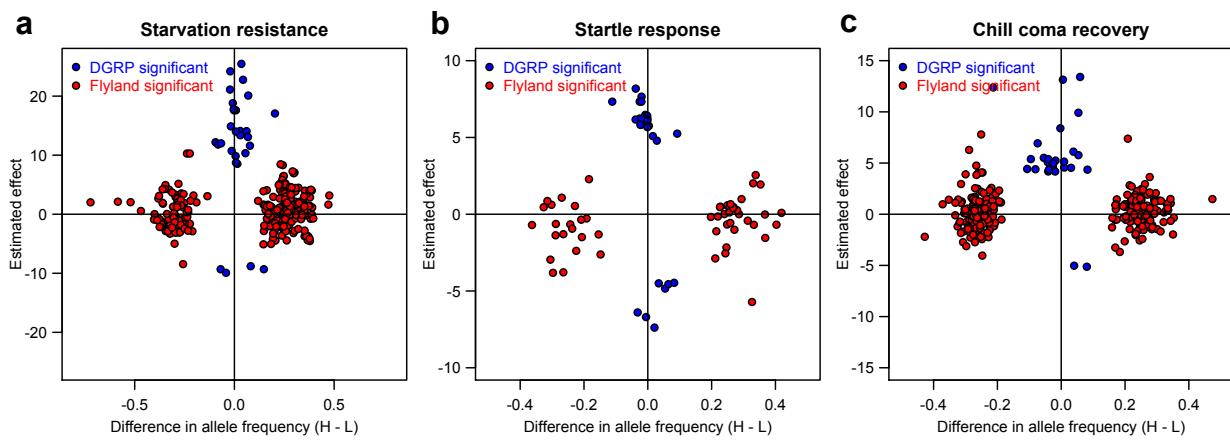
**Figure S1. Relationship between significance and allele frequency differentiation and distribution of effect sizes.** Significance ( $-\log_{10} P$ ) of SNPs are plotted against the absolute difference in allele frequencies for (a) starvation resistance, (b) startle response and (c) chill coma recovery. Points with dark color indicate SNPs meeting the  $P < 0.05$  Bonferroni corrected significance level. Histograms of the absolute difference in allele frequencies are shown for (d) starvation resistance, (e) startle response and (f) chill coma recovery.



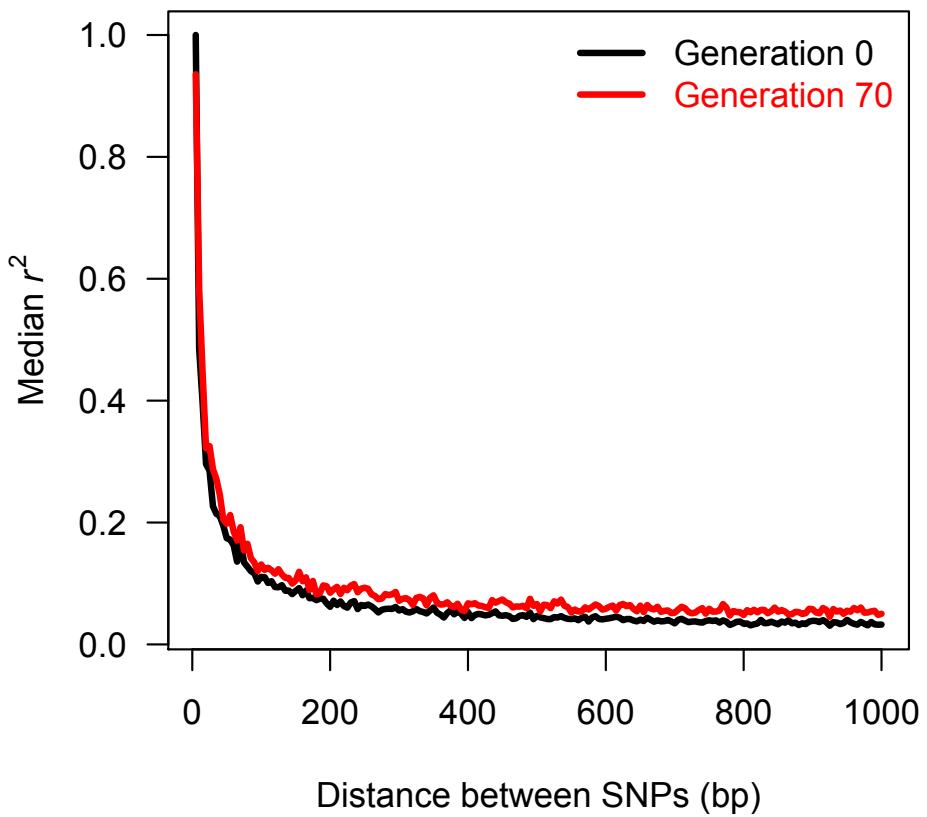
**Figure S2. ‘ROC’ plots of overlap between Flyland and DGRP significant SNPs.**  
 (a) The significance threshold in Flyland is relaxed such that increasing numbers of SNPs were called significant (x-axis). The proportion of these SNPs significant in the DGRP is plotted on the y-axis, giving a graphical representation of overlap similar to a ROC curve. A diagonal line is expected when there is not more overlap than expected by chance. (b) The reciprocal analysis is plotted, showing the proportion of significant Flyland SNPs recovered by relaxing the significance threshold in the DGRP.



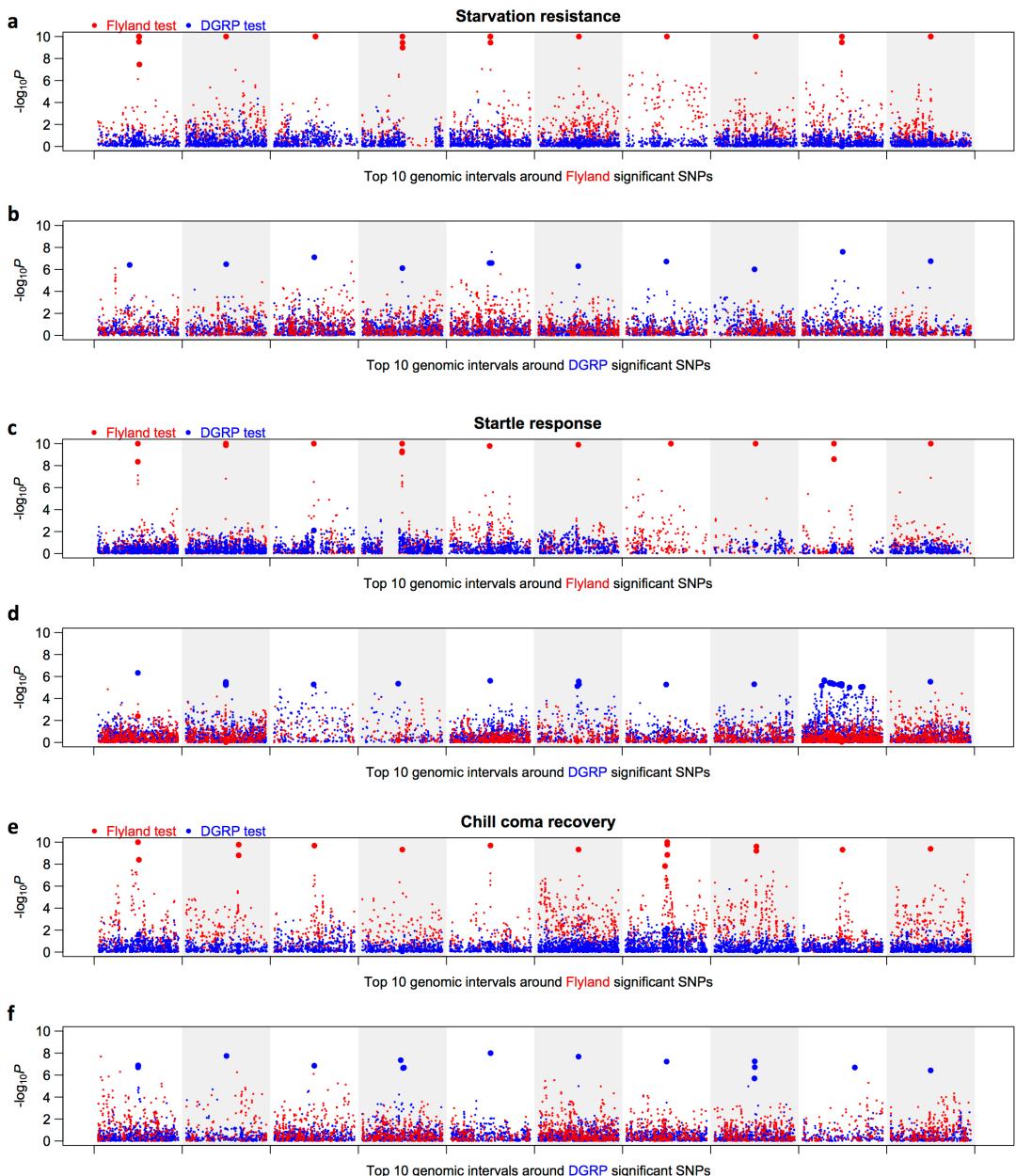
**Figure S3. Power to replicate in one population significant SNPs discovered in a second population.** (a) Phenotypes of individual flies are simulated based on SNP effects estimated in the DGRP and allele frequencies in the Flyland. Allele frequencies among the top and bottom 15% flies are estimated and compared by assuming realistic sequencing depth observed in Flyland. Significance thresholds are the same as in the real Flyland experiment. Power is calculated by counting the number of simulation replicates in which the SNPs are correctly detected among 200 simulations. (b) To estimate SNP effects in the Flyland population, we performed a simulation and searched for the minimum effect size at which at least 50% of 100 simulations produce allele frequency difference as large as the observed value. We then simulated the fly phenotypes based on SNP effects estimated in Flyland and allele frequencies in the DGRP. We tested for association exactly as done for the DGRP GWAS. We calculated power by counting the number of simulation replicates in which the SNPs are correctly detected among 200 simulations. (c) Results of similar analyses for chill coma recovery to assess power in Flyland and (d) power in the DGRP (d). Sizes of the symbols are proportional to their effect sizes. Starvation resistance is not analyzed because flies were pooled without measuring individual phenotypes for this trait.



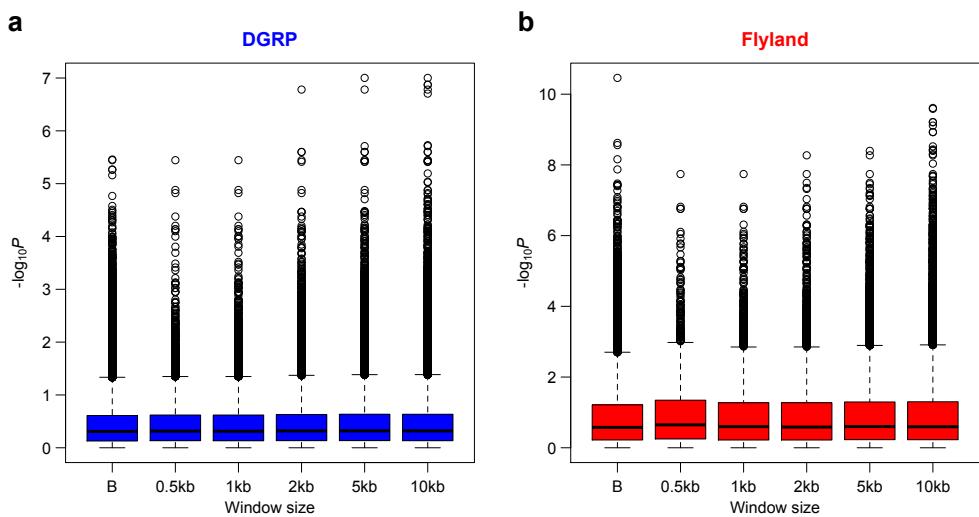
**Figure S4. No correlation between allelic effects estimated in Flyland and the DGRP.**  
 Estimated allele substitution effects in the DGRP are plotted against the difference in frequency of the same allele between high and low pools for (a) starvation resistance, (b) startle response and (c) chill coma recovery in Flyland.



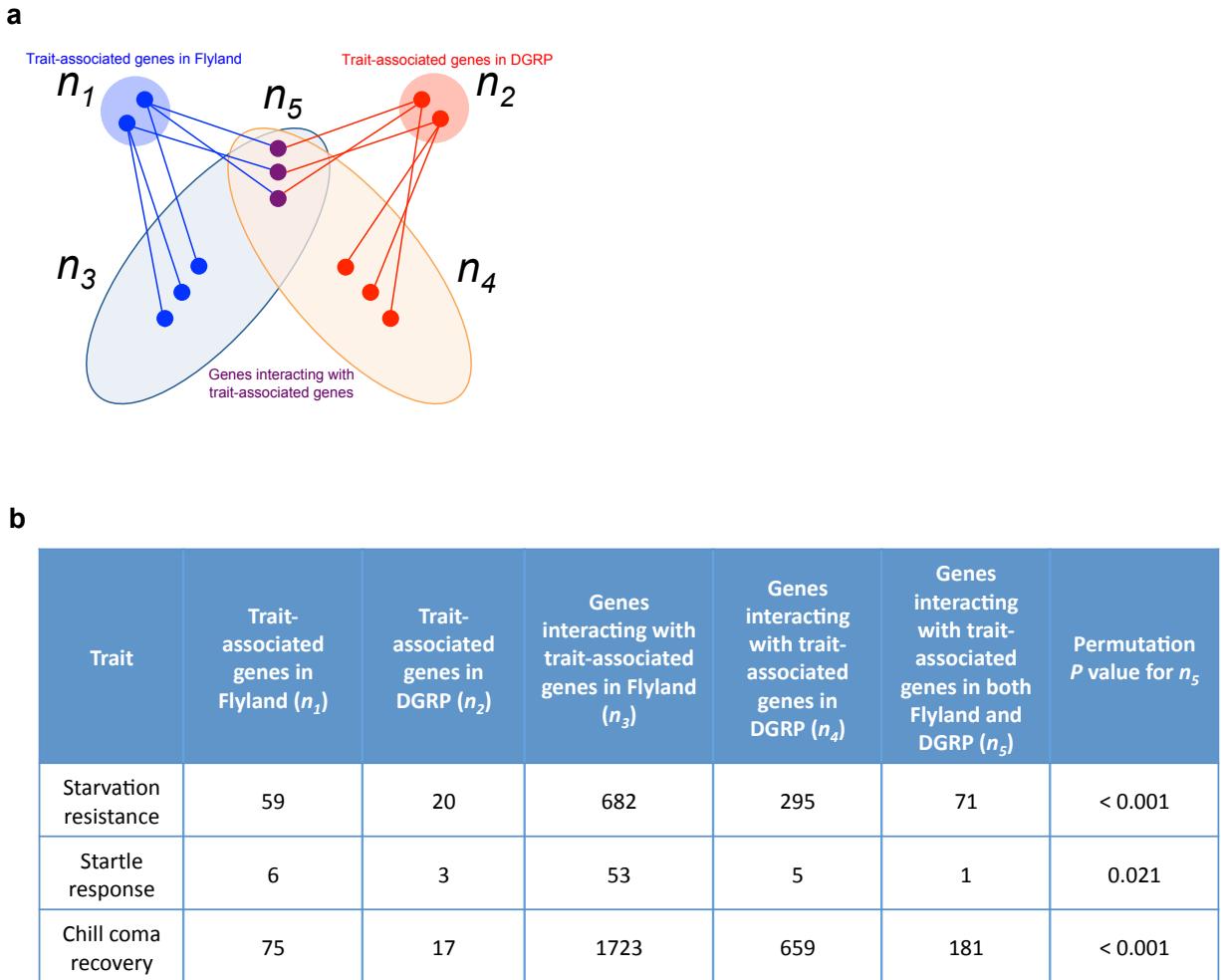
**Figure S5. Change of linkage disequilibrium after 70 generations of random mating.**  
We simulated population expansion and maintenance exactly as described in the methods section using actual genotypes and assuming a medium genome-wide recombination rate of 2.9cM/Mb. Only regions around significant SNPs in Flyland for starvation resistance were simulated. Results are summarized by median pair-wise  $r^2$  with increasing SNP-SNP physical distance.



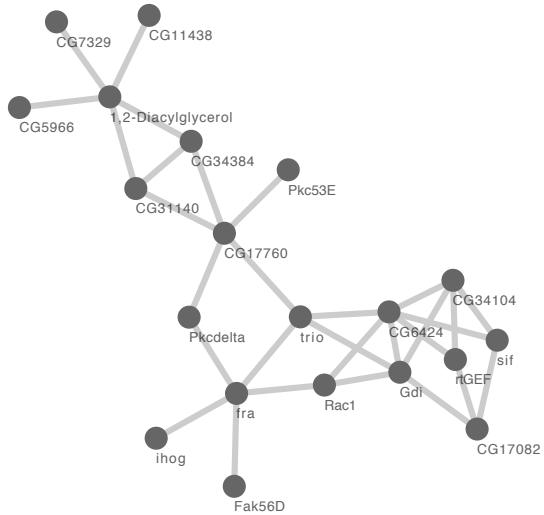
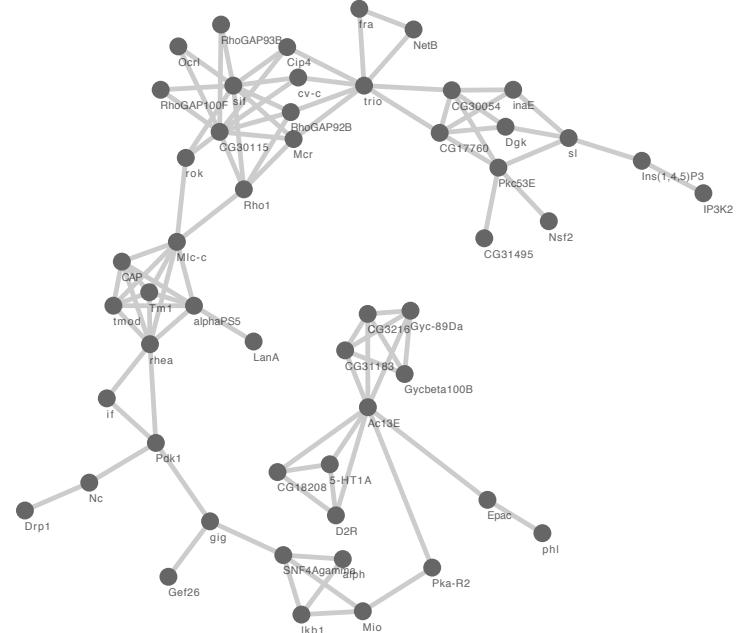
**Figure S6. Association of SNPs around significant SNPs in Flyland and the DGRP.** For each of starvation resistance (a-b), startle response (c-d), and chill coma recovery (e-f), SNPs around (10kb) significant SNPs in Flyland for starvation were tested in Flyland and the DGRP (a, c, e). The same analysis was performed for SNPs selected around significant SNPs in DGRP and tested in Flyland and the DGRP (b, d, f).



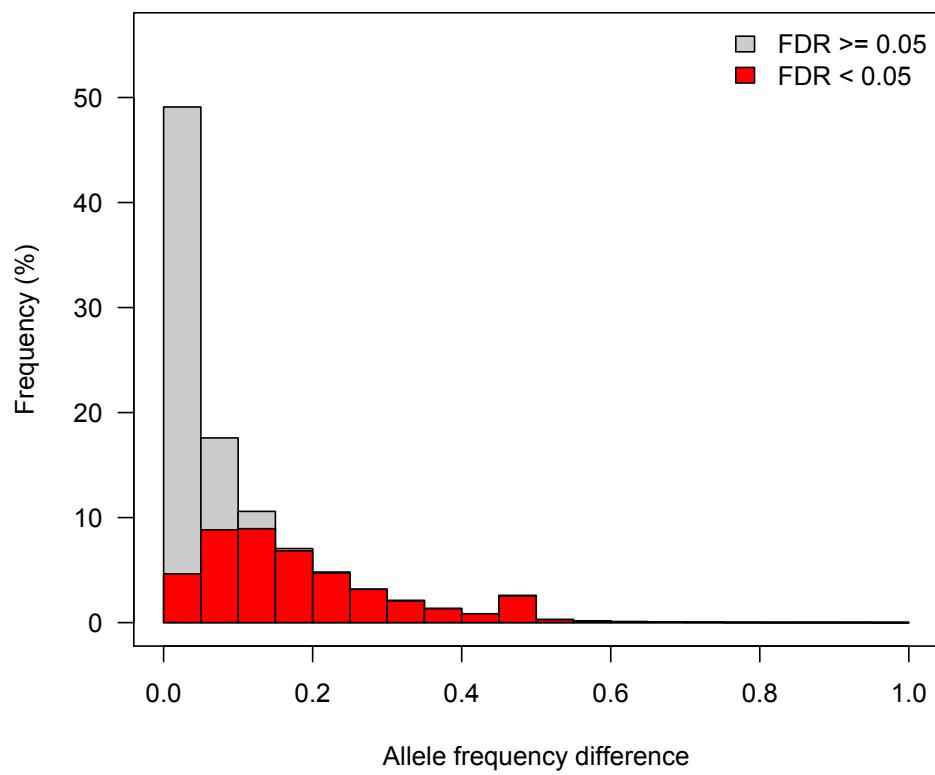
**Figure S7. Significance of SNPs around association signals.** (a) SNPs were selected within 0.5, 1, 2, 5 and 10 kb of Flyland signals or randomly selected from the genome (B: background) and tested in the DGRP. (b) SNPs were selected within 0.5, 1, 2, 5, 10 kb of DGRP signals and tested in Flyland.



**Figure S8. Common genes interacting with distinct trait-associated genes in Flyland and the DGRP.** (a) Schematic showing different genes associated with a quantitative trait in Flyland ( $n_1$ ) and in the DGRP ( $n_2$ ). These genes interact with  $n_3$  and  $n_4$  genes, respectively, with  $n_5$  interacting genes in common. (b) The table gives values of  $n_1$ ,  $n_2$ ,  $n_3$ ,  $n_4$  and  $n_5$  for starvation resistance, startle response and chill coma recovery. The significance of the overlap ( $n_5$ ) was tested by randomly sampling gene lists 1000 times and counting the number of times when overlapping genes are equal or greater than  $n_5$ .

**a****b**

**Figure S9. Sub-networks of epistatically interacting genes enriched for signaling pathways.** For (a) starvation resistance and (b) chill coma recovery, genes involved in epistatic interactions were used to identify significantly enriched sub-networks using R-spider (22). The P-value for enrichment of both traits is  $P < 0.005$ .



**Figure S10. Changes in allele frequencies between Flyland and the DGRP.** Histogram of absolute allele frequency differences between Flyland and the DGRP. Significant and non-significant changes are indicated by red and grey, respectively.