Report

Current Biology

Testing Implications of the Omnigenic Model for the Genetic Analysis of Loci Identified through Genome-wide Association

Highlights

- GWA analysis identifies genes associating with *Drosophila* pupal case length
- Randomly chosen genes can similarly affect pupal case length as those from GWA
- A large number of genes can influence independent quantitative pupal traits
- Current strategies for genetic confirmation of GWA hits need to be revisited

Authors

Wenyu Zhang, Guy R. Reeves, Diethard Tautz

Correspondence

tautz@evolbio.mpg.de

In Brief

Zhang et al. show for a standard GWA mapping approach in *Drosophila* that not only the GWA candidate genes can be confirmed by testing knockouts in isogenic lines, but also a set of randomly chosen genes can affect the phenotype. This supports an implication of the omnigenic model that most genes contribute to most quantitative trait phenotypes.

Zhang et al., 2021, Current Biology *31*, 1092–1098
March 8, 2021
https://doi.org/10.1016/j.cub.2020.12.023

CellPress

CellPress
OPEN ACCESS

Current Biology

Report

Testing Implications of the Omnigenic Model for the Genetic Analysis of Loci Identified through Genome-wide Association

Wenyu Zhang,[1] Guy R. Reeves,[1] and Diethard Tautz[1,2,*]
[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Biology, August-Thienemann-Straße 2, 24306 Plön, Germany
[2]Lead Contact
*Correspondence: tautz@evolbio.mpg.de
https://doi.org/10.1016/j.cub.2020.12.023

SUMMARY

Organismal phenotypes usually have a quantitative distribution, and their genetic architecture can be studied by genome-wide association (GWA) mapping approaches. In most of such studies, it has become clear that many genes of moderate or small effects contribute to the phenotype.[1–4] Hence, the attention has turned toward the loci falling below the GWA cut-off, which may contribute to the phenotype through modifier interactions with a set of core genes, as proposed in the omnigenic model.[5] One can thus predict that both moderate effect GWA-derived candidate genes and randomly chosen genes should have a similar likelihood to affect a given phenotype when they are analyzed via gene disruption assays. We have tested this hypothesis by using an automated phenotyping system for *Drosophila* pupal phenotypes.[6,7] We first identified candidate genes for pupal length in a GWA based on the *Drosophila* Genetic Reference Panel (DGRP)[8,9] and showed that most of these candidate genes are indeed involved in the phenotype. We then randomly chose genes below a GWA significance threshold and found that three-quarters of them had also an effect on the trait with comparable effect sizes as the GWA candidate genes. We further tested the effects of these knockout lines on an independent behavioral pupal trait (pupation site choice) and found that a similar fraction had a significant effect as well. Our data thus confirm the implication that a large number of genes can influence independent quantitative traits.

RESULTS

We used pupal case length as a measure of pupal size in *Drosophila melanogaster*, which can be scored via automated image analysis (Figure 1). Comparable to human height, pupal case length in *Drosophila* is also a highly heritable trait (with h^2 of 0.44–0.50 and H^2 of 0.52–0.71; Table S1), with a polygenic basis.[6] The same approach allows also to score the independent phenotype of pupation site choice (pupation height), which represents a behavioral phenotype with a polygenic genetic architecture.[7] The profiles of pupal case length from 14 wild-type and of 198 *Drosophila* Genetic Reference Panel (DGRP) inbred lines are shown in Figures S1A and S1B, suggesting that DGRP lines capture indeed the existing variation of pupal case length in *Drosophila melanogaster*.

Genome-wide Association Analysis

We used the genetic variants of DGRP freeze 2[9] and excluded the variants with missing values above 20% and minor allele frequency below 5% from further analysis. In order to correct any potential influence from cryptic population structure, we exploited a linear mixed model implemented in the fastLMM[11] program (version 0.2.32) for genome-wide association (GWA) mapping analysis. Several possible covariates were assessed in order to optimize the model for GWA mapping analysis (see STAR Methods). We found 50 significant SNPs ($p \leq 1 \times 10^{-5}$) to be associated with pupal case length in the DGRP strains, corresponding to 67 associating genes that locate within 5 kb up/downstream (default setting in SnpEff)[12] of these genetic variants (Figure 2A; Data S1A). To identify possible additional candidate genes associated with the variants, we examined the long-range linkage disequilibrium (LD) between pairs of detected candidate variants and with other genetic variants found in the DGRP strains. LD blocks were then calculated for each significant genetic variant with a commonly used threshold r^2 = 0.8,[13] and 17 significant LD blocks were found with average block size of 19.3 kb (Data S1B). Combining the additional genes identified in the above LD blocks, we identified in total 90 candidate genes associating with pupal case length variation in *Drosophila melanogaster.*

Conventional Phenotype Confirmations

To test the phenotypic effects from different alleles on pupal case length at single-gene scale, we used transposon insertion mutagenesis lines that had been constructed in common co-isogenic backgrounds.[14] Ten gene disruption lines corresponding to 9 of the GWA-identified associated genes (all are directly associated with the GWA significant SNPs, not from LD blocks) were available for this experiment (Figure 2A; Data S1A). All experimental tests were done via replicated phenotypic

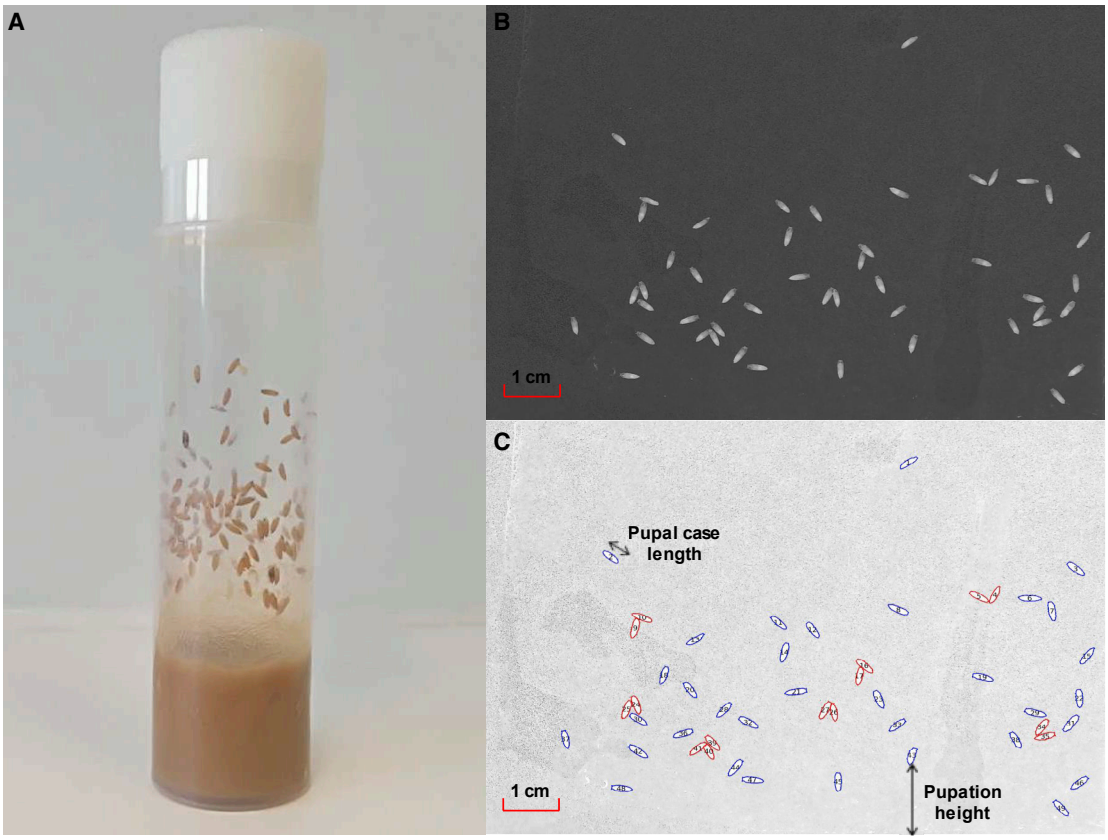

Figure 1. Automated Measurement of Pupal Case Length and Pupation Height

(A) A vial with a large, but not too dense, number of pupae attached to the wall.

(B) A plastic sheet that lined the inner wall was included in the vial and was taken out for being photographed.

(C) The image analysis software CellProfiler[10] was then used to identify the outlines of the pupae and measure their major axis lengths and distances between their pupation sites to the food surface. The pupae that can be reliably measured (singularized) are marked in blue, and the ones with lower confidence scores (aggregated) are marked in red.

comparisons between co-isogenic stocks and homozygous mutant lines (Figure 2C).

Figure 2C shows the profile on the measured pupal case length differences between mutant lines and the respective progenitor stocks. Lines corresponding to eight genes showed a significant difference (Wilcoxon rank-sum test; $p \leq 0.05$). Interestingly, all eight lines show an increased pupal case length ($p = 0.008$; binomial test; two-tailed), suggesting that the effects are not simply due to a general growth deficiency caused by the insertion lines. It should be noted that two independent experimental tests were conducted for gene *Drl-2*, with transposon insertions landing in the different regulatory regions and from different co-isogenic backgrounds (Data S2A). Both showed consistent phenotypic effects, affirming the robustness of the gene disruption assays.

Expression Analysis

Genes involved in determining pupal case length can be hypothesized to be expressed at developmental stages close to the pupation process. To explore this, we analyzed a developmental RNA sequencing (RNA-seq) dataset from *Drosophila melanogaster*,[15] which included the transcriptome of 27 distinct developmental stages covering all the major phases of the *Drosophila* life cycle (Figure 3).

Compared with a null dataset from 1,000 times of random selection of the same number of coding genes, we observed a moderate positive enrichment of these 90 GWA candidate genes with expression (FPKM > 0) in all 27 tested distinct developmental stages (Figure 3A), suggesting that GWA tends to pick up gene candidates with broad expression. More interestingly, we found these GWA candidate genes are mostly significantly enriched in the developmental period between late larval stage and pupal stage (Fisher's exact test; $p \leq 0.05$; Figure 3B). As the pupation process occurs from the late larval stage, and the genesis of pupal case is completed before eclosion, our results imply that the developmental period between late larval and pupal stage might be the most relevant developmental stages for pupal case length morphogenesis in *Drosophila melanogaster*. However, note that, although the effect is significant, the absolute differences to other stages are small (Figure 3A), implying that many of the earlier expressed genes may also have an effect on later developmental stages.

Phenotypic Effects of Randomly Chosen Genes

The omnigenic model suggests that most, if not all, genes with naturally segregating variants contribute to the heritability of a quantitative phenotype when they are expressed in the relevant

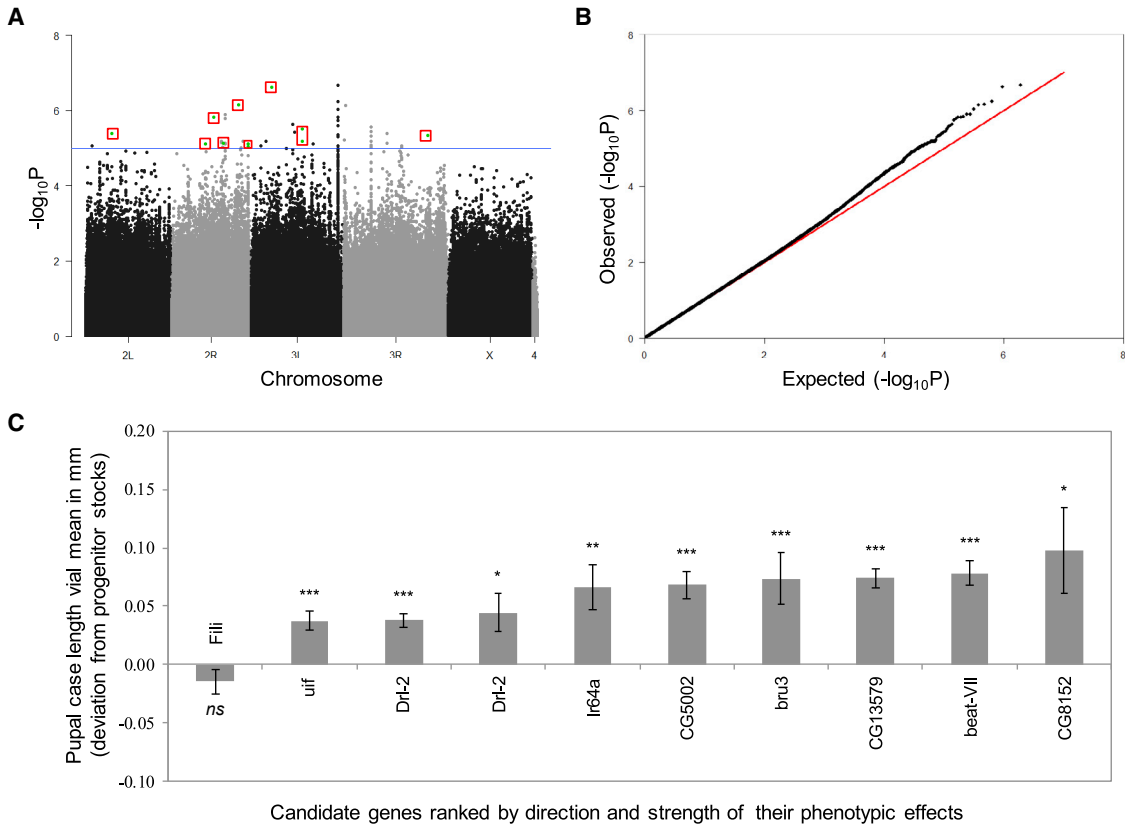

Figure 2. GWA Mapping and Gene Disruption Test Results on Pupal Case Length

(A) Manhattan plot for GWA results. The p values ($-\log_{10}$ transformed) are shown on the y axis. The blue horizontal line marks the nominal p value threshold (1×10^{-5}) used in this study. The genetic variants in red rectangles are the ones selected for experimental validation as shown in (C).

(B) Quantile-quantile (qq) plot for all GWA p value results. The expected and observed p values ($-\log_{10}$ transformed) are represented by the x axis and y axis, respectively.

(C) Gene disruption test results of GWA candidate genes. The phenotypic effects were measured as the deviation of pupal case length of stocks with transposon-based gene disruption compared with that from the corresponding progenitor stocks. The error bars show the standard error of mean (SEM) values. Statistical p values were computed via Wilcoxon rank-sum tests. ***$p \leq 0.001$; **$p \leq 0.01$; *$p \leq 0.05$; not significant (ns), $p > 0.05$. All stocks were tested as homozygotes. See also Data S1.

tissues or stages.[5] This implies that they must in some way functionally contribute to the phenotype, presumably through pleiotropic network interactions. Hence, when gene disruption lines are tested for effects on the phenotype, one can predict that most genes can affect a given quantitative trait. We have therefore set out to use our phenotyping pipeline to test this implication. We selected 47 gene disruption lines (corresponding to 45 disrupted genes) constructed in a co-isogenic background from the panel of *Drosophila* gene disruption stocks[14] using the following criteria: (1) the GWA p value of the focal gene is above the significance threshold (i.e., $>1 \times 10^{-5}$); (2) the homozygous disruption is viable—note that this criterion biases against essential genes (approximately 30% are not homozygous viable); and (3) the corresponding gene has detectable expression in the above defined relevant developmental stages for pupal case length (late larval and pupal stage; FPKM > 0). It should be noted that all these 9 experimental tested GWA genes also have expression in these relevant developmental stages (Data S2A); therefore, the random genes and GWA genes were picked in an equivalent way for comparison.

Figure 4A compares the phenotypic effect sizes of all strains tested in this study with their respective pupal case length GWA study (GWAS) p values. It shows that the genes picked because of their GWA significance have not necessarily the largest phenotypic effects as disruption lines and that there is no significant difference on the overall phenotypic effects between GWA candidate genes and randomly chosen genes (Wilcoxon rank-sum test p = 0.16; Figure S2A). Overall, around 76% of random genes (34 of 45) showed significant effects ($p \leq 0.05$) on the pupal case length phenotype (Figure 4B), in contrast with the 89% of GWA candidate genes (8 out of 9), but these numbers are not significantly different (p = 0.67; Fisher's exact test).

Strains with different co-isogenic backgrounds should have a similar probability of showing a phenotype for the tested trait. This is indeed the case; we find that 4 out of 5, 34 out of 44, and 7 out of 8 of the isogenic progenitor stocks 3605, 5905, and 6326, respectively, show a significant phenotypic effect on pupal case length (Data S2A; $p > 0.05$; Fisher's exact test for any pairwise comparison). For three genes, we could do the direct comparison, because independent disruption stocks in

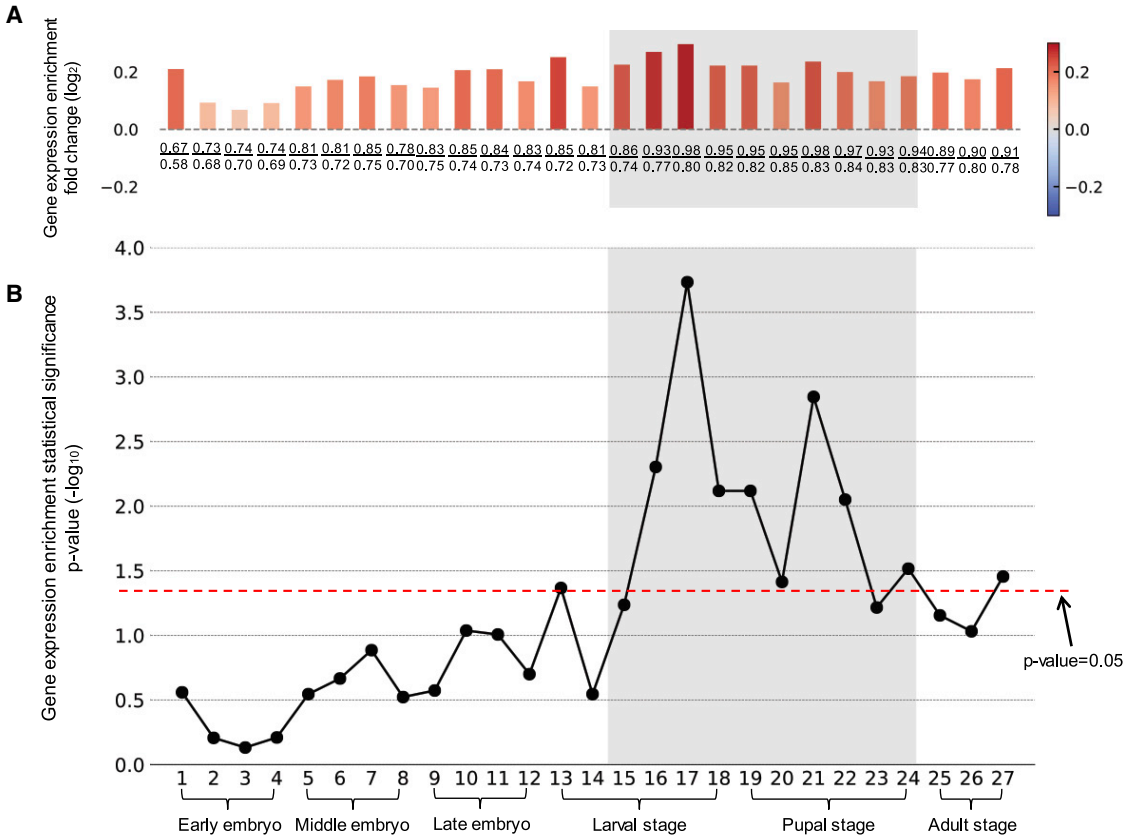

Figure 3. Gene Expression Enrichment Patterns of GWA Candidate Genes across Different Developmental Stages

The expression enrichment statistics is described in STAR Methods.

(A) The gene expression enrichment fold changes (i.e., the ratios of the fractions of expressed genes of GWA candidate genes and the null dataset on $\log_2$ scale) between GWA candidate genes and the null dataset. Below each bar, we show the fractions of expressed genes (fragments per kilobase of exon model per million reads mapped - FPKM > 0) for GWA candidate genes on the top and for the null dataset on the bottom.

(B) The statistical significances (p values on $-\log_{10}$ scale; Fisher's exact test) of the differences of fractions of expressed genes between GWA candidate genes and the null dataset. The dashed red line specifies the p value threshold of 0.05, and the shaded area indicates the "relevant" developmental stages of pupal case length morphogenesis.

two different backgrounds were available (*Drl-2*, *CG14007*, and *CG42260*). All three show very similar p values in the pairwise comparisons (*Drl-2*: 0.008 and 0.04; *CG1400:* 0.002 and 0.002; *CG42260*: 0.008 and 0.004; Data S2A).

Effect sizes should not have a direction, i.e., the likelihood of shorter pupae should be the same as for longer pupae for the different genes. However, for the set of GWA candidate genes discussed above, we found mostly longer pupae for gene mutagenesis stocks (Figure 2C), suggesting the existence of such a bias. But when compared for the set of randomly tested genes, we found also a good number of cases with shorter pupae (longer: 30; shorter: 17; Data S2B). Although there was still a bias for longer pupae, the differences are not significant (p = 0.08; binomial test; two tailed). The same conclusion holds for the pupation height analysis as an independent trait as discussed below (p = 0.18; Data S2B).

Finally, we asked whether there could be differences with respect to disruption types in different gene regions (coding sequence [CDS], intron, or untranslated region [UTR]). However, no significant difference was detected (Figure S2B). Similarly, we found no effect on where the transposon was inserted within the length of the CDS (Figure S2C). This latter observation is a bit surprising, because one could have assumed that a direct loss of the CDS should have more effect, but the transposon disruption could have also additional effects on the RNA expression or transcript stability, i.e., it does not necessarily act on the CDS alone.

Phenotypic Effects on an Independent Phenotype

If most genes have an effect on one quantitative trait, one would expect that the same is true for a second trait. As discussed above, pupation site choice is such a second independent trait, which we measured with the same setup. Given the sharing of gene expressions in the common relevant developmental stages (Data S2A), all the 54 disrupted genes on the basis of pupal case length can therefore serve as random gene disruption tests for pupation site choice. Indeed, we observed that around 69% of the randomly tested genes (37 of 54) actually showed significant effects ($p \leq 0.05$) on the pupation site choice trait (Figure 4C). This fraction of effects is not significantly different from the fraction of effects of randomly tested genes for the pupal case length phenotype reported above (p = 0.51; Fisher's exact test). This supports the notion that the majority of genes with expression

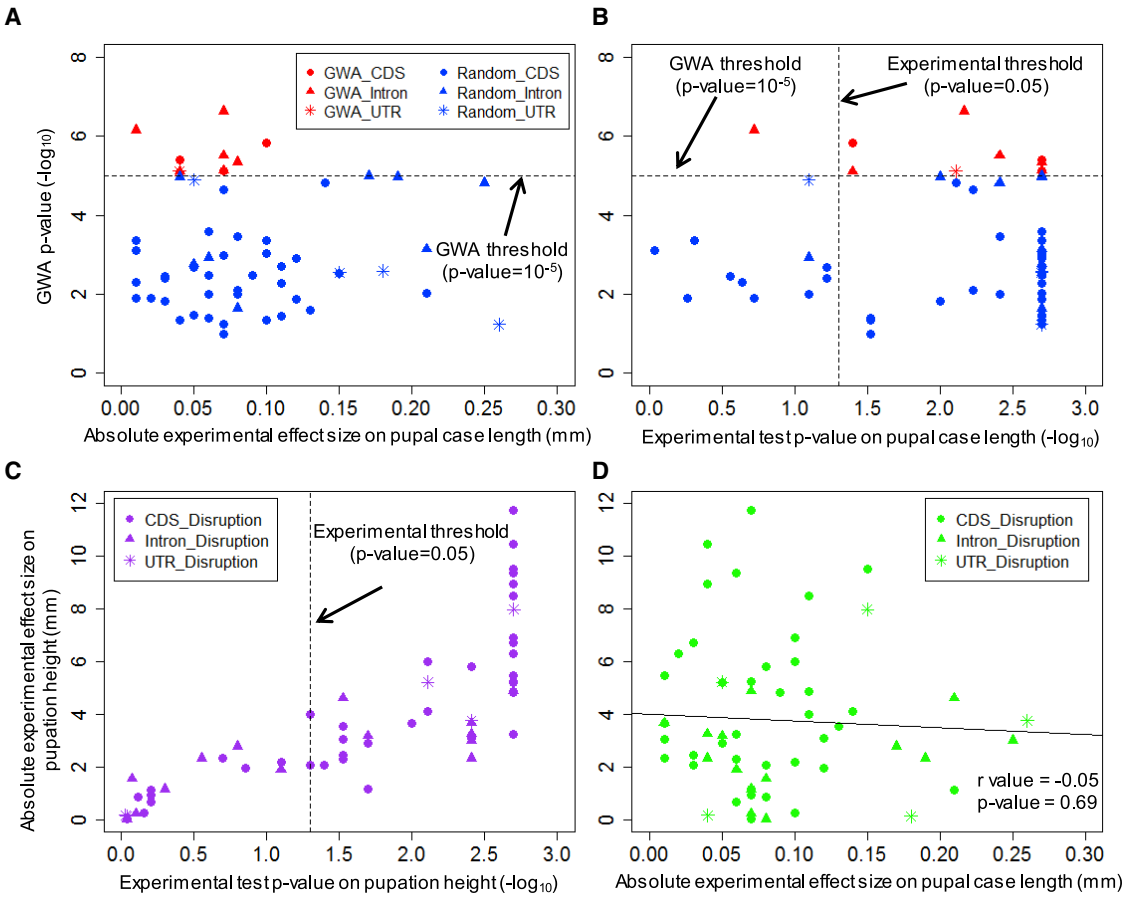

Figure 4. The Impact of Gene Disruption on Pupal Case Length and Pupation Height

(A and B) Comparison between GWA p values and absolute experimental effect sizes and statistical p values for gene disruption of GWA candidate genes (labeled in red) and randomly selected genes (labeled in blue) on pupal case length are shown in (A) and (B), respectively.

(C) Pupation site choice (height) comparison between absolute experimental effect sizes and statistical p values for gene disruption of all tested genes. The horizontal dashed line marks the GWA nominal p value threshold (1×10^{-5}), and the vertical dashed line indicates the p value threshold for experimental tests (0.05). The most rightward vertical stripes in (B) and (C) indicate the maximum statistic for the given sample size in the rank order test.

(D) Correlation between the absolute experimental effect sizes on pupation height and pupal case length. Each dot represents one gene disruption line. Genes with transposon insertion mutagenesis in different genic regions (CDS, intron, or UTR) are labeled with various dot shapes.

See also Data S2.

in the relevant developmental stages can have effects on a given quantitative trait at this stage. Interestingly, the effect sizes do not correlate for the two phenotypes tested (Figure 4D), implying that the effects are channeled through different networks.

DISCUSSION

There are two distinct views of genetics with their roots going back to Mendel and Galton.[4] One focuses on the effects of single loci, determined through studying the effects of (usually artificial) loss-of-function alleles in isogenic backgrounds. The other focuses on natural variation by studying the heritability of quantitative traits, usually in population samples and a polygenic context. One of the paradigms that unites these fields is the assumption that a gene that is identified through a natural variant in a quantitative trait genetics approach, such as a GWAS, should point to a locus that is causally involved in the phenotype under study. Furthermore, such a causal involvement could then be tested in a knockout approach for this gene. But the evidence from extensive GWASs suggests now that almost every gene may segregate heritable variation contributing to a phenotype under study.[5] This implies that knockout variants of almost every gene should alter any given phenotype. Our results support this notion.

One can formulate omnigenic/polygenic models in two ways: one emphasizing a possible role of core genes that are detected via GWA hits versus modifier loci that come out below the threshold[5] and the other via emphasizing a spectrum of effect sizes across all genes, with the notion that the ones falling below the GWA threshold still contribute a substantial proportion to the overall heritability of the phenotype.[16,17] Both views are not mutually exclusive, and they agree that a very large number of genes are expected to contribute to the heritability of any given quantitative trait. Most of the evidence for the omnigenic/polygenic model has so far come from human studies, which are often focused on disease questions and their associated special considerations and limitations.[17] But for well-developed genetic model systems,

such as *Drosophila*, one can do direct genetic experiments to test these implications. Note that the approach to use more or less randomly selected transposon insertion lines to test for an effect of a complex trait has been done before (e.g., for neural development,[18] aggressive behavior,[19] or lifespan[20]), but not in a direct contrast to a corresponding GWAS (Figures 4A and 4B).

Using the DGRP lines that represent segregating wild-type alleles, we were able to identify a set of 90 candidate genes for the pupal size phenotype (Data S1). We found that eight out of nine tested GWA candidate genes showed significant effects ($p \leq 0.05$) on pupal size. These include two loci known to be involved in developmental processes, such as *uninflatable* (*uif*), which modulates Notch signaling,[21] and the receptor tyrosine kinase *Derailed 2* (*Drl-2*) that transduces Wnt5 signaling.[22] Accordingly, the GWA did identify loci that could be part of core pathways of the phenotype, similarly to what we showed before for the behavioral pupation site choice phenotype[7] and as has been concluded for other studies using the DGRP.[23]

We thus set out to use gene disruption experimental tests on randomly chosen genes with expression in the relevant developmental stages. The GWA p value for these 45 randomly chosen genes is below any threshold that one would normally consider using. Accordingly, none of them would have been identified as GWA candidate genes. However, our results showed that approximately three-quarters of them actually had significant effects ($p \leq 0.05$) on the pupal case length phenotype. Intriguingly, a similar fraction, but only partially overlapping loci, also showed an effect on the behavioral pupation site choice phenotype, thus confirming another implication of the omnigenic model, namely a high degree of pleiotropy in gene effects.

One can ask why these randomly chosen genes, especially the ones showing significant phenotypic effects on pupal phenotypes, were not picked up by GWA. The most likely reason is that they do not include segregating variants of sufficient effect size in the population from which the DGRP was derived. In fact, the power of detection for small or moderate effect size variants in the DGRP is low, due to the limited number of strains.[23] It is very possible that at least some of the other genes could have been detected in an analysis of higher power, provided they are segregating relevant natural variation. But the problem of detection power plagues GWA analyses in DGRP lines also with respect to the uncertainty whether any of the detected variants are actually true associations or just consequences of random fluctuations among a very large number of tests. We have discussed above that additional criteria can be used to support true associations. But the validity of this question disappears anyway when both GWA hits, as well as non-hits, are actually commonly involved in the phenotype, at least when tested as knockout alleles.

This raises the question of whether a single GWA is ever likely to be sufficient to understand a complex trait? Hence, it should be of special interest to assess whether GWA p values, potentially when viewed across multiple studies, provide at least a guide toward identifying the core networks required to generate the phenotype, as predicted by the omnigenic model.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.12.023.

ACKNOWLEDGMENTS

We thank the lab members and anonymous reviewers for helpful discussions and suggestions. We thank Anita Moeller, Elke Blohm Sievers, and Michaela Schwarz for their excellent technical help with conducting this experiment. This work was supported by institutional funding through the Max Planck Society, Germany.

AUTHOR CONTRIBUTIONS

W.Z., G.R.R., and D.T. designed the study. W.Z. conducted the experiment, analyzed the data, and performed the statistical analysis, with support from G.R.R. W.Z. and D.T. wrote the paper. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 16, 2020
Revised: November 19, 2020
Accepted: December 15, 2020
Published: January 7, 2021

REFERENCES

1. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.
2. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. *43*, 519–525.

3. Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J., Kutalik, Z., et al.; Electronic Medical Records and Genomics (eMEMERGEGE) Consortium; MIGen Consortium; PAGEGE Consortium; LifeLines Cohort Study (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. *46*, 1173–1186.
4. Tautz, D., Reeves, G., and Pallares, L.F. (2020). New experimental support for long standing concepts of polygenic genetics implies that the Mendelian genetic paradigm needs to be revised. In The New (Old) Genetics, A. Wittinghofer, and H. Jäckle, eds. (NAL-live).
5. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: from polygenic to omnigenic. Cell *169*, 1177–1186.
6. Reeves, R.G., and Tautz, D. (2017). Automated phenotyping indicates pupal size in *Drosophila* is a highly heritable trait with an apparent polygenic basis. G3 (Bethesda) *7*, 1277–1286.
7. Zhang, W., Reeves, G.R., and Tautz, D. (2020). Identification of a genetic network for an ecologically relevant behavioural phenotype in Drosophila melanogaster. Mol. Ecol. *29*, 502–518.
8. Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., Casillas, S., Han, Y., Magwire, M.M., Cridland, J.M., et al. (2012). The Drosophila melanogaster Genetic Reference Panel. Nature *482*, 173–178.
9. Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ràmia, M., Tarone, A.M., Turlapati, L., Zichner, T., Zhu, D., Lyman, R.F., et al. (2014). Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Res. *24*, 1193–1208.
10. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100.
11. Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. Nat. Methods *8*, 833–835.
12. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.
13. Pallares, L.F., Harr, B., Turner, L.M., and Tautz, D. (2014). Use of a natural hybrid zone for genomewide association mapping of craniofacial traits in the house mouse. Mol. Ecol. *23*, 5756–5770.
14. Bellen, H.J., Levis, R.W., He, Y., Carlson, J.W., Evans-Holm, M., Bae, E., Kim, J., Metaxakis, A., Savakis, C., Schulze, K.L., et al. (2011). The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics *188*, 731–743.
15. Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473–479.
16. Thoday, J.M., and Thompson, J.N. (1976). The number of segregating genes implied by continuous variation. Genetica *46*, 335–344.
17. Wray, N.R., Wijmenga, C., Sullivan, P.F., Yang, J., and Visscher, P.M. (2018). Common disease is more complex than implied by the core gene omnigenic model. Cell *173*, 1573–1580.
18. Norga, K.K., Gurganus, M.C., Dilda, C.L., Yamamoto, A., Lyman, R.F., Patel, P.H., Rubin, G.M., Hoskins, R.A., Mackay, T.F., and Bellen, H.J. (2003). Quantitative analysis of bristle number in Drosophila mutants identifies genes involved in neural development. Curr. Biol. *13*, 1388–1396.
19. Edwards, A.C., Zwarts, L., Yamamoto, A., Callaerts, P., and Mackay, T.F.C. (2009). Mutations in many genes affect aggressive behavior in Drosophila melanogaster. BMC Biol. *7*, 29.
20. Magwire, M.M., Yamamoto, A., Carbone, M.A., Roshina, N.V., Symonenko, A.V., Pasyukova, E.G., Morozova, T.V., and Mackay, T.F.C. (2010). Quantitative and molecular genetic analyses of mutations increasing Drosophila life span. PLoS Genet. *6*, e1001037.
21. Loubéry, S., Seum, C., Moraleda, A., Daeden, A., Fürthauer, M., and Gonzalez-Gaitan, M. (2015). Uninflatable and notch control the targeting of sara endosomes during asymmetric division. Curr. Biol. *25*, 817–818.
22. Sakurai, M., Aoki, T., Yoshikawa, S., Santschi, L.A., Saito, H., Endo, K., Ishikawa, K., Kimura, K., Ito, K., Thomas, J.B., and Hama, C. (2009). Differentially expressed Drl and Drl-2 play opposing roles in Wnt5 signaling during Drosophila olfactory system development. J. Neurosci. *29*, 4972–4980.
23. Mackay, T.F.C., and Huang, W. (2018). Charting the genotype-phenotype map: lessons from the Drosophila melanogaster Genetic Reference Panel. Wiley Interdiscip. Rev. Dev. Biol. *7*, e289.
24. Richardson, M.F., Weinert, L.A., Welch, J.J., Linheiro, R.S., Magwire, M.M., Jiggins, F.M., and Bergman, C.M. (2012). Population genomics of the Wolbachia endosymbiont in Drosophila melanogaster. PLoS Genet. *8*, e1003129.
25. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
26. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.
27. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915.
28. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
29. Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. https://doi.org/10.1101/005165.
30. Schmidt, J.M., Battlay, P., Gledhill-Smith, R.S., Good, R.T., Lumb, C., Fournier-Level, A., and Robin, C. (2017). Insights into DDT resistance from the drosophila melanogaster genetic reference panel. Genetics *207*, 1181–1193.
31. Zeh, J.A., Bonilla, M.M., Adrian, A.J., Mesfin, S., and Zeh, D.W. (2012). From father to son: transgenerational effect of tetracycline on sperm viability. Sci. Rep. *2*, 375.
32. Attrill, H., Falls, K., Goodman, J.L., Millburn, G.H., Antonazzo, G., Rey, A.J., and Marygold, S.J.; FlyBase Consortium (2016). FlyBase: establishing a Gene Group resource for Drosophila melanogaster. Nucleic Acids Res. *44* (D1), D786–D792.
33. Hoffmann, A.A., and Rieseberg, L.H. (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu. Rev. Ecol. Evol. Syst. *39*, 21–42.
34. Dembeck, L.M., Böröczky, K., Huang, W., Schal, C., Anholt, R.R., and Mackay, T.F. (2015). Genetic architecture of natural variation in cuticular hydrocarbon composition in Drosophila melanogaster. eLife *4*, e09861.
35. Lee, Y.C.G., Yang, Q., Chi, W., Turkson, S.A., Du, W.A., Kemkemer, C., Zeng, Z.B., Long, M., and Zhuang, X. (2017). Genetic architecture of natural variation underlying adult foraging behavior that is essential for survival of Drosophila melanogaster. Genome Biol. Evol. *9*, 1357–1369.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Broad range Quick DNA Marker	New England Biolabs	Cat #: N0303
Tetracycline	MilliporeSigma	Cat #: 87128
Critical Commercial Assays		
DNeasy blood and tissue kit	QIAGEN	Cat #: 69504
Deposited Data		
DGRP freeze 2 genetic variant calling data	[9]	http://dgrp2.gnets.ncsu.edu/
Drosophila melanogaster developmental RNA-seq data	[15]	NCBI SRA: SRP001065
Experimental Models: Organisms/Strains		
Drosophila melanogaster: 14 wild strains	EHIME or UC Davis	Data S3A in this paper
Drosophila melanogaster: 198 DGRP inbred strains	BDSC	Data S3B in this paper
Drosophila melanogaster: 57 transposon insertion strains and 3 corresponding progenitor stocks	BDSC	Data S3C in this paper
Oligonucleotides		
Primer wsp81F: 5′-tggtccaaaatgtgagaaac-3′	[24]	N/A
Primer wsp691r: 5′-aaaattaaacgctactcca-3′	[24]	N/A
Software and Algorithms		
CellProfiler (v2.1.0)	[10]	https://cellprofiler.org/
SPSS (v22)	IBM	IBM SPSS Statistics (v22)
PLINK (v1.90)	[25]	https://zzz.bwh.harvard.edu/plink/
fastLMM (v0.2.32)	[11]	https://pypi.org/project/fastlmm/
SnpEff (v4.3t)	[12]	https://pcingola.github.io/SnpEff/
Fastp (v0.20.0)	[26]	https://github.com/OpenGene/fastp
HISAT2 (V2.1.0)	[27]	http://daehwankimlab.github.io/hisat2/
featureCounts (v1.6.3)	[28]	http://subread.sourceforge.net/
R software (v4.0.2)	CRAN	https://cran.r-project.org/
R package: qqman	[29]	https://cran.r-project.org/web/packages/qqman/
Other		
Onset data logger	HOBO®	Cat #: 10328732
OHP Transparency Film	nobo®	Cat #: 33638237
Incubator for flies rearing	POL-EKO APARATURA	ST-CHL 500-700-1200-1450
Canon Camera	Canon inc.	Canon EOS-1100D
NanoDrop spectrophotometer	Thermofisher	ND1000 spectrophotometer

RESOURCE AVAILABILITY

Lead Contact

Further information and resource requests should be directed to and will be fulfilled by the Lead Contact, Diethard Tautz (tautz@evolbio.mpg.de).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The raw individual pupal measurement data generated in this study can be accessed via Mendeley Data (https://data.mendeley.com/datasets/m7bv36r2xf/1). All the analyses in this study were performed by using public software and algorithms.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Drosophila strains

In total, 14 wild-type stocks, 198 DGRP inbred lines, 57 transposon insertion mutagenesis stocks (disruption for 54 annotated protein coding genes) and 3 corresponding progenitor stocks were assayed in this study. Almost all these *Drosophila melanogaster* strains are public available through either Bloomington *Drosophila* Stock center (https://bdsc.indiana.edu/) or EHIME *Drosophila* stock center (https://kyotofly.kit.jp/cgi-bin/ehime/index.cgi), while one wild-type stock was obtained from UC DAVIS. Detailed information of these stocks can be found in the Data S3.

Flies rearing

All flies were reared in vials containing cornmeal–molasses–agar medium at 24°C, 55% −78% relative humidity, and a 12:12-h light-dark cycle. A HOBO® onset data logger was placed in the incubator to monitor and record any potential environmental changes, including temperature, light and humidity.

METHOD DETAILS

Phenotyping of pupal case length

We conducted the measurement of pupal case length by adopting an earlier established image-analysis based automated phenotyping pipeline as described in[6,7]. The entire phenotyping pipeline is implemented with five successive procedures:

1) *Vial preparation.* Standard cornmeal–molasses–agar food was dispensed into standard 28.5 mm diameter, 95 mm height vials (Genesee Scientific) to a depth of approximately 20mm. Once the food vials had fully cooled, 10.1 cm x 10.5 cm squares of transparent film (nobo, plain paper copier film, 33638237) were slid into the bottom of each vial lining their entire vertical wall. Inserting the film does not change the storage properties of the food. Prior to introducing adult flies, a very small amount of live yeast paste was dotted on the food surface. A barcode with custom printed semi-transparent label was affixed to the outside of each vial as the unique identifier.
2) *Mating flies.* Approximately ten 2-5 days old healthy female flies (fifteen for inbred stocks to compensate the reduced fertility) and five similar-aged male flies were introduced into each vial, under the incubation condition as mentioned above. These adult flies were cleared from the vials after 1-2 days introduction and vials were kept in the same incubation condition for another 8-9 days to allow them to reach pupation stage. In case of observing that most of offspring in the vials were present as pupae attached to the transparent film, the film was then gently taken out from each vial, and the food from the lower part was scrapped away and any viable larvae were also removed.
3) *Film Photographing.* Once removed from the vial, the film (with the barcode label affixed) was then placed into a custom plastic frame, which holds the film flat for further photographing. Frames were then photographed in a light tight box with a sliding door, which provided illumination only from underneath the frame, effectively silhouetting the pupae while minimizing tangential shadows. Every photograph included a 1 euro cent coin (16.25 mm diameter), for the control of camera coordinate changes and the conversion of measurements from pixels to millimeters (mm). Batches of the resulting images were then introduced into the following image analysis procedure.
4) *Image analysis.* We applied the open-source image analysis software CellProfiler (v2.1.0)[10] for the recognition of pupae and measurements of a variety of attributes, with a customized pipeline adopted from[7]. First, by using the CellProfiler module ‘‘identify primary objects,’’ we identified any ‘‘primary object’’ with significant distinction from the background without restriction on their sizes. Second, the above identified objects composing of multiple touching pupae were disentangled into distinct pupae (module ‘‘Untangle Worms’’). Third, the resulting candidate pupae were shrunk and re-propagated outward for a more precise detection of the edges of each pupa based on boundary changes in pixel intensity (module ‘‘Identify secondary objects’’). Lastly, distinct attributes for the pupae were calculated and a specific confidence class was assigned for each pupa based on its size attribute. Based on manual curation on 40 randomly selected films, we found that the above CellProfiler pipeline can reach to a 96% of sensitivity (fraction of identified true pupae), but a modest false discovery rate of 19% for identified putative pupae[7]. To further reduce the false discovery rate, we refined an additional criterion setting based on the size attributes of ‘‘true’’ pupae based on manual curation[7]. Applying this setting of new criteria, the false discovery rate for pupae detection dropped to around 0.15%, with only a tiny fraction (< 0.7%) of loss for true positive results.
5) *Pupal case length measurement.* The pupal case length is defined as the length of the major axis of the ellipse that has the same normalized second central moments as the region of identified pupae, measured with the ‘‘Areashape_MajorAxisLength’’ index in CellProfiler. Based on its ratio to the diameter measurement of 1 euro cent coin (16.25 mm diameter) included in the photograph, the measurement of case length of each pupa was converted from pixels to mm. The mean of all the measurements of pupal case length for all the pupae in the food vial was taken as representative for the focal vial. To further reduce the potential bias from low sampling effect, we only included vials with a pupal density of a minimum of 15, and required each stock should include at least 6 such reliable vial measurements.

Repeated-measurements of control stocks

Throughout the experiments for wild-type and DGRP inbred strains measurements, two wild-type stocks (S-314 and S-317) were continually re-measured in the manner described above to control for environmental effects, especially small fluctuations in humidity. Such re-measurement of control stocks were not applied for the functional validation experiments as shown below, as the experimental tests on each pair of gene disruption stock and progenitor stock were conducted in the same experimental condition. The impacts of the incubator relative humidity changes on the pupal case length measurement (across all rounds of DGRP inbred stock experiments) were directly examined by using Pearson's correlation tests for two control stocks separately. Given the roughly consistent measurements on pupal case length of these two control stocks throughout all DGRP line experiments (Figure S1C) and the only minor impact on pupal case length from the change of relative humidity across experiments (Figure S1D), no special action was taken to adjust the pupal case length measurements.

Phenotyping on wild-type strains and DGRP inbred lines

To assess whether the *Drosophila* Genetic Reference Panel (DGRP) lines[8,9] reflect the scope of natural variation, we measured also the pupal case length phenotypes of 14 natural wild-type *Drosophila melanogaster* strains collected from different geographic regions, including two from Africa (strain details in Data S3A and S3B). The profiles of pupal case length from the wild-type and of 198 DGRP inbred lines are shown in Figures S1A and S1B, suggesting that DGRP lines capture indeed the existing variation of pupal case length in *Drosophila melanogaster*.

Estimates of pupal case length heritability

We estimated the broad sense heritability (H^2) of pupal case length with the variance components of a linear model of the form: Phenotype = Population mean + Line effect + error[30]. We computed the total phenotypic variance as Genetic Variance (G_v) + Environmental Variance (E_v), and the H^2 as G_v/G_v+E_v. This calculations for both wild-type and DGRP inbred stocks were implemented by using IBM SPSS Statistics (version 22), with pupal case length measurement as the dependent variable and DGRP stock names as a random factor. Meanwhile, we compared the estimates of H^2 computed with the same methodology, but from different *Drosophila melanogaster* strains datasets[6]. The estimates on the narrow sense heritability (h^2) of pupal case length based on mid-parent regression were directly retrieved from[6]. The detailed estimates on pupal case length can be found in Table S1.

Wolbachia infection effect test

Wolbachia pipientis is a maternally transmitted endosymbiotic bacterium that was reported to infect around 53% of DGRP inbred stocks (Data S3B)[9]. We examined the possible effects of *Wolbachia* infection on pupal case length through two different approaches. First, we directly compared the pupal case length values between infected strains and non-infected strains, and found no significant difference between these two groups (Wilcoxon rank-sum test: p value = 0.90, Data S4A).

Second, we conducted a direct experimental test to compare the changes of pupal case length measurement between *Wolbachia*-infected stocks and *Wolbachia*-free stocks, which were created though two generations of tetracycline treatment on the infected stocks (by adding an appropriate volume of 100 mg/ml of tetracycline suspended in 99% ethanol to the surface of the solid prepared food) and then reared for at least another two generations with standard food to avoid any detrimental parental effects[31]. We first ruled out the possibility that tetracycline treatment could have an influence on pupal case length, by comparing the pupal case length measurement changes between three randomly selected *Wolbachia*-free stocks (DR_14, DR_45, DR_106), and the same stocks with the above mentioned tetracycline treatment procedure. Then we randomly selected three *Wolbachia* infected stocks (DR_16, DR_21, DR_67), and half of the files were treated with tetracycline as mentioned to create *Wolbachia*-free stocks, and the rest half the flies from same three strains were also reared with standard food across the experiment as controls. We extracted genomic DNA from the above 6 stocks individually by using DNeasy blood and tissue kit (QIAGEN), and measured the purity and concentration of the resulting DNA with NanoDrop ND-1000 spectrophotometer (Thermofisher). A diagnostic PCR to test for the presence of the *Wolbachia wsp* gene was performed by using the primers wsp81F (5′-tggtccaaaatgtgagaaac-3′) and wsp691r (5′-aaaattaaacgctactcca-3′)[24], under the reaction condition of 35 cycles of 94°C for 15 s, 55°C for 30 s and 72°C for 1 minute. A standard (1%) agarose gel electrophoresis was used to test for the presence of the PCR product (~630 bp), with the broad range Quick DNA Marker (NEB #N0303) as loading ladder. Pupal case length measurement between the three *Wolbachia*-infected and created *Wolbachia*-free lines were then measured and compared by using Wilcoxon rank-sum tests. No significant statistical differences on pupal case length were observed for all tested strains (Figure S3). Based on these analyses, we concluded that the *Wolbachia* infection on DGRP strains had only minor, if any, influence on pupal case length, and consequently we did not incorporate the *Wolbachia* infection status in the association analysis below.

Principal component analysis and genomic inversion effect test

The DGRP lines were generated in a way that population structure impacts should be minimized, but some genetic relatedness leading to cryptic population structure might still exist[23]. To examine whether any cryptic population structure could contribute to the observed pupal case length variation of DGRP inbred stocks, we exploited PLINK[25] to identify major principal components (PCs) of genetic variants in the DGRP strains. We retrieved all the genetic variant calling data from DGRP freeze 2 [9], for which the coordinates were based on Flybase version 5[32], and only kept the genetic variants with < 20% missing values and $\geq$ 5% minor allele frequency (MAF) (corresponding to 1,903,028 genetic variants) for this analysis. We examined the correlation between pupal case

length values and the top 20 PCs, and found three out of these 20 PCs to be significantly associated with pupal case length values (Data S4B, Pearson's correlation test, p value $\leq$ 0.05), suggesting a possible impact on pupal case length from cryptic population structure. Hence, we decided to use a linear mixed model implemented in the fastLMM[11] program (version 0.2.32) for GWA mapping analysis as shown below.

Approximately 45% of the DGRP strains harbor at least one type of major genomic inversion[9] (Data S3B), and these major genomic inversions might contribute to the observed population structure and have impact on pupal case length as well. We systematically tested the correlations between genomic inversion status in DGRP strains and the top 2 PCs based on genetic variants as mentioned above, and found significant effects from In(2L)t and In(3R)Mo (Pearson's correlation test, p value $\leq$ 0.05, Data S4C), hinting toward potential roles in population divergence[33]. Moreover, we also observed a significant association between pupal case length and In(3R)Mo (Pearson's correlation test, p value = 0.0002). Accordingly, the presence status of In(3R)Mo in DGRP lines was incorporated as a covariate of the linear mixed model used for GWA mapping analysis on pupal case length.

Genome-wide association analysis

We performed genome-wide association (GWA) analysis on pupal case length for the above filtered genetic variants from DGRP freeze 2 [9], by using the FastLMM[11] program (version 0.2.32), with the major genomic inversion In(3R)Mo status being included as a covariate for the GWA analysis. This program fitted a linear mixed model that could control for population stratification effects.

We defined GWA significant associated genetic variants by using p value $\leq 1 \times 10^{-5}$, which is a nominal threshold frequently used in *Drosophila* quantitative trait genetic studies[7,34,35]. The R package "qqman"[29] was exploited for the visualization of GWA results in a Manhattan plot and qq-plot.

We predicted the effects of GWA significant genetic variants by using SnpEff[12] with default parameters, and taken all the corresponding affected protein-coding genes as associating genes. In short, all the protein-coding genes within 5 kb up/down-stream of focal genetic variant were taken as its associating genes. We also tested the genotypic linkage disequilibrium (LD) for each GWA significant genetic variant and other genetic variants by calculating the squared correlation estimator r^2 with PLINK v1.90[25]. A significant genetic region (QTL) was defined by the position of the most distant downstream and upstream genetic variants showing a minimum r^2 of 0.8 to the focal GWA significant genetic variant. The combined associating genes from SnpEff predictions and the genes within the QTL regions were considered as GWA candidate genes (Data S1).

Expression analysis

We downloaded the *Drosophila melanogaster* developmental raw Illumina paired-end RNA-Seq data from Graveley et al.[15] (NCBI SRA accession: SRP001065). This dataset included the transcriptome of 27 distinct developmental stages covering all the major steps in *Drosophila* life cycle, which were further collapsed into six super-developmental stages: *1) Early embryo*: including 0-2 hr. embryo, 2-4 hr. embryo, 4-6 hr. embryo, and 6-8 hr. embryo; *2) Middle embryo:* including 8-10 hr. embryo, 10-12 hr. embryo, 12-14 hr. embryo, and 14-16 hr. embryo; *3) Late embryo:* including 16-18 hr. embryo, 18-20 hr. embryo, 20-22 hr. embryo, and 22-24 hr. embryo; *4) Larval stage:* including L1 stage, L2 stage, L3 stage of 12 hr. post-molt, L3 stage of dark blue gut (PS1-2), L3 stage of light blue gut (PS3-6), and L3 stage of clear blue gut (PS7-9); *5) Pupal stage:* including white prepupa (WPP), 12 hr. after WPP (P5), 24 hr. after WPP (P6), 2 days after WPP (P8), 3 days after WPP (P9-10), and 4 days after WPP (P15); *6) Adult stage:* male/female after 1-day eclosion, male/female after 5-day eclosion, and male/female after 30-day eclosion.

We trimmed and filtered the low-quality raw Illumina RNA-Seq reads sequenced from the sample of each developmental stage by using Fastp program v0.20.0[26], and only included the sequencing reads with minimum length of 20bp and average quality score of 20 for further analysis. The filtered RNA-Seq reads were aligned to *Drosophila melanogaster* v6 reference genome sequence with HISAT2 v2.1.0[27], taking advantage of the protein-coding gene annotation in Flybase v6.32[32] by using the–exon option of the hisat2-build. Then we counted the fragments mapped to the annotated genes with featureCounts v1.6.3[28], and calculated the expression level of each annotated coding gene from the sample of each developmental stage in the unit of FPKM (Fragments Per Kilobase of transcript per Million mapped reads).

To get the null dataset of the number of expressed genes, we randomly selected 90 genes (same number as GWA candidate genes) from all coding genes annotated in Flybase v6.32[32] for 1,000 times, and counted the number of genes with expression (FPKM > 0) for each developmental stage independently. The average number (round to the closer integer) of expressed gene from these 1,000 times of permutation was taken as the null number of expressed genes for each developmental stage. We took the average number of expressed genes of males and females in the adult stages as for representation. We calculated the gene expression enrichment fold changes across these 27 developmental stages as the ratios of the fractions of genes with expression for GWA candidate genes and that of null dataset. The statistical significances (p values) of on the difference of the fractions of genes with expression for GWA candidate genes and that of null dataset were computed by using Fisher's exact tests.

Functional validation experiments

We validated the phenotypic effects of 9 GWA candidate genes from 10 gene disruption mutagenesis stocks (The gene *Drl-2* was tested twice with different disruption locations), based on their stock availability in the *Drosophila* gene disruption project[14] (Data S3C). All these mutagenesis stocks were conducted with transposon insertion disruption, for which each focal gene was disrupted by a transposon (Minos, P-element or PBac) insertion in the gene region (either coding or regulatory region). The exact landing site of the transposon was determined based on the sequencing reads of both sides of flanking regions[14]. The detailed information about

these gene mutagenesis stocks and their co-isogenic progenitors is provided in Data S3C. All these 10 stocks are homozygous insertion viable, and thus were tested as homozygotes (In case of semi-lethal, only the adult flies with no balancer maker were used for experimental tests). For each test, at least 8 vials (15 females and 5 males introduced to each vial) were set up for each insertion and its progenitor stock, and their phenotypic values (pupal case length and pupation height) were measured by following the above phenotyping pipeline. The experimental tests on each pair of gene disruption stock and progenitor stock were conducted in the same experimental round, thus any potential influence (food preparation, incubator environment et al.) on the phenotype measurements can be avoided. The experimental effect size of each gene disruption was calculated as the average deviation of phenotypic values of gene disruption stock from that of its progenitor stock. The statistical differences of the phenotypic measurements between gene disruption stocks and their progenitor stocks were performed by using Wilcoxon rank-sum tests.

Additionally, we also experimentally tested the phenotypic effects of 45 selected genes, on the basis of the following criteria: 1) the GWA p value on the focal gene falls out the significance threshold (1×10^{-5}), 2) the focal gene has homozygous transposon insertion disruption viable stock(s) from the panel of *Drosophila* gene disruption project[14], and 3) the focal gene has detectable expression in relevant development stages for pupal case length (late larval and pupal stage; FPKM > 0). The second criterion biases against essential genes (approximately 30% are not homozygous viable), but otherwise the selection was essentially random. Two of the above chosen genes (*CG14007*, and *CG42260*) have 2 qualified transposon insertion stocks with disruption in different locations (while one disruption stock for each of the rest 43 genes), and thus in total 47 transposon insertion stocks were applied for the experimental tests. The phenotyping test experiments were conducted with the same procedure as stated above. The GWA association p value was taken as the lowest p values from all the genetic variants within the gene and the 5kb up/down-stream of target gene. The detailed results of all the functional validation experiments can be found in Data S2A.

Furthermore, we analyzed the pattern on absolute pupal case length phenotypic effect size (absolute deviation of the phenotypic values of gene disruption stock from that of progenitor stock) of the disruption of the genes, based on their categories: 1) GWA candidate gene or randomly chosen gene (Figure S2A); 2) the gene region of the transposon insertion landing site (coding region, intron or UTR, Figure S2B); 3) the percentage of disruption on gene's coding region (Figure S2C). For the latter one, only the gene disruptions in the coding region were included for analysis, and the average percentage of disruption was taken as representative, in case of multiple transcripts for a focal gene. The statistical difference on the absolute pupal case length phenotypic effect size between two categorized groups were tested by using Wilcoxon rank-sum tests, and the correlation between the absolute pupal case length phenotypic effect size and percentage of gene's coding region disruption with Pearson's correlation test.

Automatic measurement of pupation site choice

We also assessed the pupation height (an indicator of pupation site choice) for the gene disruption stocks used in the experimental validation tests as shown above. The measurement of pupation height follows the procedure described in[7]. In brief, the pupation height was defined as the distance from the vertical coordinate of pupation site (pupal center) to the food surface in the vial in millimeter (mm). As the pupation height measurements are sensitive to the pupal density in the vial, the raw measurements for pupation height were further adjusted according to the equation described in[7]:

$$\text{Pupation height corrected for pupal density} = O - (D - M) * \beta \quad \text{(Equation 1)}$$

O: pupation height vial mean to be corrected
D: automated estimate of pupal density in the vial to be corrected
M: average vial density across all the experiments (set as 70)
β: average slope of regression of density against pupation height mean vial (set as 0.145)

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of two independent pupal traits

The measurements of pupal case length and pupation height are described in the above text. For each tested *Drosophila* strain, we only included vials with a pupal density of a minimum of 15, and required each stock should include at least 6 such reliable vial measurements. The detailed values of vial and pupa measurements for all tested strains can be found in Data S3.

Statistical analysis

All information concerning the statistical details is provided in the main text and the above Method section correspondingly. Boxplots using the following standards: the midline dot represent the median of the data, with the upper and lower limits of the box being the third and first quartile1 (75th and 25th percentile) respectively, and the whiskers extend up to 1.5 times the interquartile range from the top (bottom) of the box to the furthest datum within that distance. Bar and line plots with error bars using the following standard: the error bar represents the standard error of mean (SEM) values for each data group. Scattered plot with regression line using the following standard: the regression line represents fitted linear regression line based on the data.

Statistical analysis for GWA was performed by using FastLMM[11] program (version 0.2.32) (Figure 2A), and all other statistical tests were conducted by using R software (v4.0.2). The Wilcoxon rank sum test was used to analyze the significant difference between two data groups (Figures 2C, S2A, S2B, S3B, and S3D; Data S2A and S4A). The Fisher's exact test was used to analyze the significant difference between two categorical variables (Figure 3). The Pearson's correlation test was used to analyze the significant association of two variables (Figures 4D, S1D, and S2C; Data S4B and S4C). The Binomial test was used to analyze whether the observation rate of one event is significant different from randomness of 0.5 (Data S2B). SNP variants with statistical p value $\leq 1x10^{-5}$ for GWA were reported as significant, otherwise not significant. For all other statistical tests, statistical significances were represented as follow: ***p value $\leq$ 0.001; **p value $\leq$ 0.01; *p value $\leq$ 0.05; ns (not significant): p value > 0.05.

Current Biology, Volume *31*

Supplemental Information

Testing Implications of the Omnigenic Model for the Genetic Analysis of Loci Identified through Genome-wide Association

Wenyu Zhang, Guy R. Reeves, and Diethard Tautz

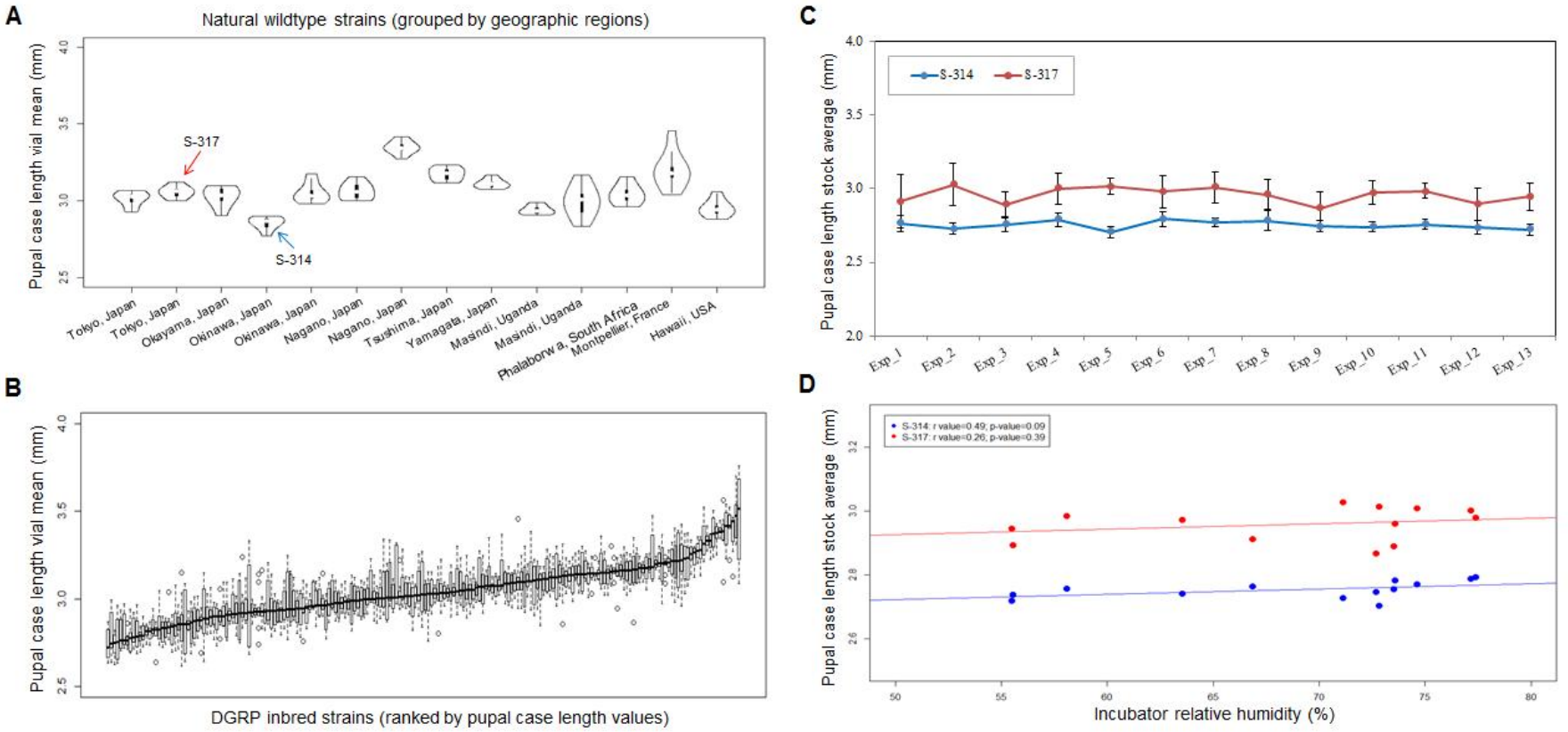

Figure S1: Pupal case length measurement of two control stocks and different *Drosophila melanogaster* strains, Related to STAR Methods. The pupae case length distribution of natural wild derived stocks (A) is grouped by geographic regions, and (B) that of DGRP inbred strains is ranked by pupal case length values. (C) Two control stocks (S-314: in blue; S-317: in red) were repeated measured as controls across all thirteen rounds of DGRP inbred stock experiments. The error bar represents standard error of mean value for each round of measurement. (D) The impact of incubator relative humidity on the measurement of pupal case length for two control stocks across all thirteen rounds of DGRP inbred stock experiments. The statistical tests were performed via Pearson's correlation tests.

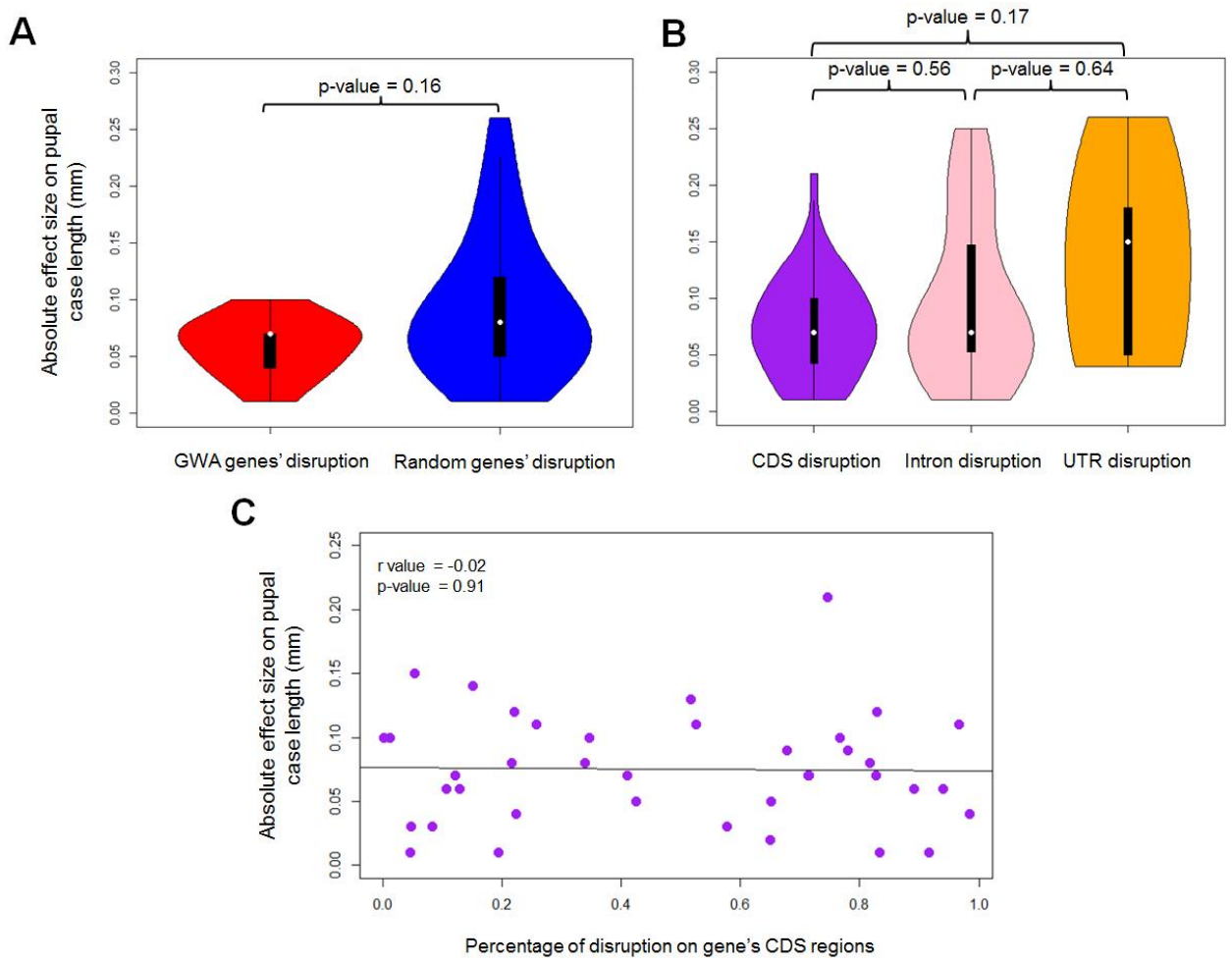

Figure S2: Gene disruption effect sizes on pupal case length based on their categories, Related to STAR Methods. (A) Comparison of absolute gene disruption effect sizes on pupal case length for GWA genes and random selection genes. (B) Comparison of absolute gene disruption effect sizes on pupal case length based on the disruption regions. The statistical p-values between two groups were computed with Wilcoxon rank-sum test. (C) Correlation between the absolute gene disruption effect sizes on pupal case length and the percentages of disruption on gene's CDS regions. This result was based on the analysis of the gene disruption in the CDS only. In case of multiple transcripts for a gene, the average percentage of disruption was included for analysis.

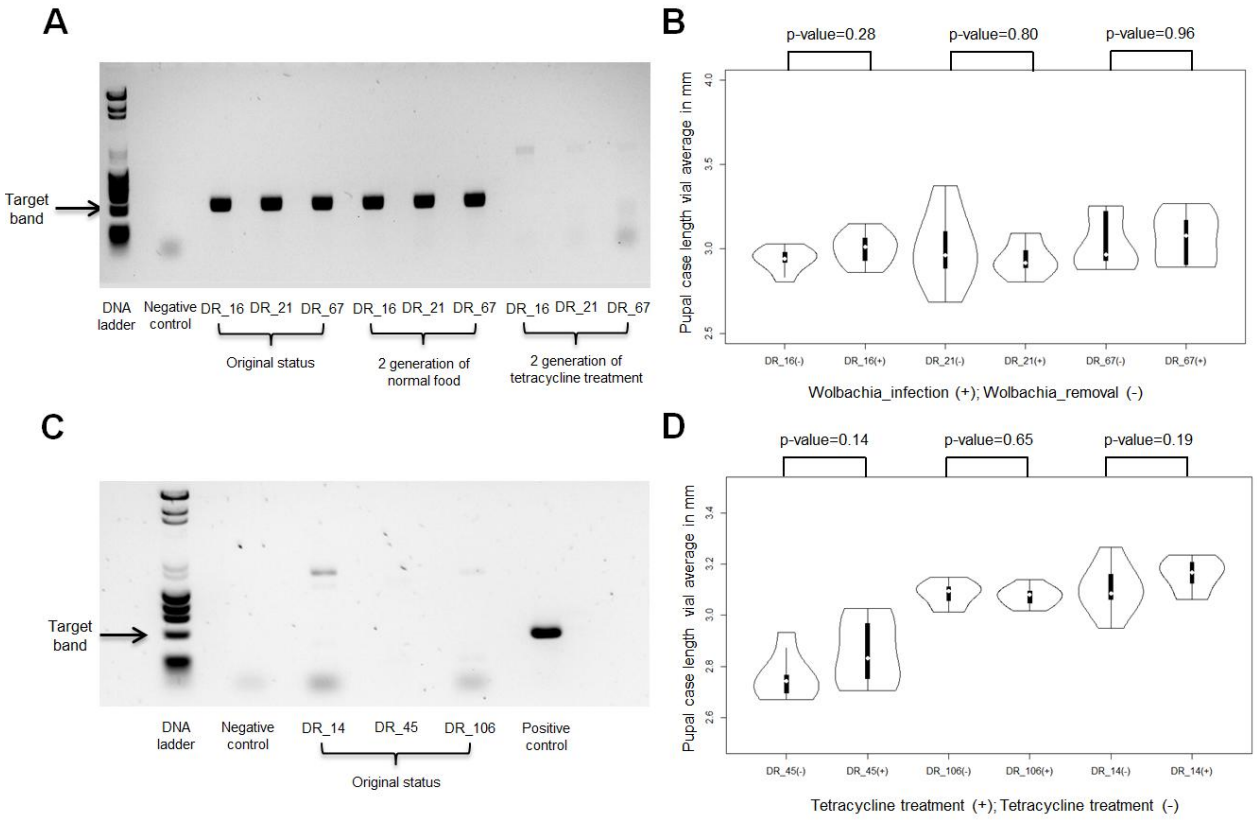

Figure S3: Effect of *Wolbachia* infection on pupal case length, Related to STAR Methods. The *Wolbachia*-infection status of randomly selected *Wolbachia*-infected (A) and *Wolbachia*-free (C) stocks were confirmed with standard agarose gel electrophoresis after PCR amplification. The comparison of pupal case length for *Wolbachia*-infected stocks and *Wolbachia*-free stocks and those after tetracycline treatment are shown in (B) and (D), respectively. The significance p-values were computed with Wilcoxon rank sum tests.

Heritability	Method	Dataset	Heritability Estimate
H^2	IBM SPSS (Variance component)	[a]wild-type strains	0.71
		[b]DGRP inbred strains	0.52
		[c]Four-Way	0.58
		[d]Eight-Way (DSPR)	0.61
h^2	Mid-parent regression	[e]Four-Way	0.44 ± 0.04 SE
		[f]Eight-Way (DSPR)	0.50 ± 0.09 SE

Table S1 Statistics for the heritability estimates of pupal case length, Related to STAR Methods.

[a] Number of stocks = 14; Average number of replicates per stock = 7.8 ± 0.6 SD; Average number of measured pupae per vial = 69 ± 24 SD;
[b] Number of ILs = 198; Average number of replicates per stock = 8.2 ± 1.6 SD; Average number of measured pupae per vial = 40 ± 17 SD;
[c] Number of RILs = 81; Average number of replicates per stock = 5.5 ± 0.8 SD; Average number of measured pupae per vial = 59 ± 24 SD;
[d] Number of RILs = 195; Average number of replicates per stock = 6.7 ± 2.4 SD; Average number of measured pupae per vial = 70 ± 34 SD;
[e] Number of single-pair crosses = 363; Average number of measured pupae per vial = 62 ± 17 SD;
[f] Number of single-pair crosses = 67; Average number of measured pupae per vial = 47 ± 15 SD;
[c-f] Datasets were retrieved from Reeves G 2017 [S1];
ILs: Inbred lines; RILs: recombinant inbred lines; DSPR: *Drosophila* Synthetic Population Resource;
SD: standard deviation; SE: standard error;

Supplemental References

S1. Reeves, R.G., and Tautz, D. (2017). Automated Phenotyping Indicates Pupal Size in Drosophila Is a Highly Heritable Trait with an Apparent Polygenic Basis. G3 (Bethesda). *7*, 1277–1286.