Supplementary methods

Additional protocol details

Based on preliminary quantifications of male-male aggression in 60 DGRP lines, we randomly chose 6 Wolbachia-free lines from each quartile of aggression level. We indeed had a wide range of both threat and aggression (Fig. 1). Our quantification of aggression is different from that used in an earlier study [1]. Nevertheless, our lines included lines that had both very low (DGRP 235) and very high (DGRP 808) aggression score in Shorter et al (2015).

Statistics

We fit a longitudinal model using duration as the response, with category of aggression (threat vs. fighting) as the only fixed effect. These categories were allowed to vary according to the random effects of both genotype (to estimate genetic correlations, r_G) and trial (r_T). Additionally, we included day effects nested within block effects in the females present experiment to account for the fact that we had two separate blocks of days, and random effects for day and observer effects in the females absent experiment, which had 2 observers scoring the videos. To account for the zeroes in observations (i.e., no aggressive behaviours observed in some trials: 37% and 7% in the females present and absent experiments, respectively), we used a Tweedie distribution with a log link function. To confirm that our estimates were robust to choice of distribution and link function, we also fit models using a zero inflated Gamma distribution (log link function).

We used Maximum Likelihood (ML) estimates for fixed effects, and refitted the model using REML to get unbiased estimates of variance components. We used the ranef() and coef() functions from glmmTMB to extract conditional means (random effects, with Wald approximations for 95% CI) of each genotype and emmeans v1.7.3 [2] to extract fixed effects and their confidence intervals. We also evaluated model fit using comparisons to parameter constrained models. For the first restricted model, correlations for the random effects of DGRP between aggression categories was fixed to zero ($r_G = 0$) using a diagonal covariance matrix, and compared to the fit of the unconstrained covariance matrices to evaluate r_G using a likelihood ratio (LR) test with *1 df* (for estimation of r_G). Similarly, we compared the fit of the full models to constrained models with only a single shared genetic variance term (i.e. no distinction for genetic effects between threat and physical aggression). These were compared to the unrestricted model with an LR test with *2 df*. Comparison of the full models with models constrained with a single class of aggressive behaviours (i.e., fitting a single genetic variance effect shared for both threat and fighting) is consistent with the full model being a better fit to the data for both experiments (females present: $LR = 11.71$, $df = 2$, $p = 0.003$; females absent: $LR = 25.6$, $df = 2$, $p = 0.0008$).

With our experimental design, the quantitative genetic analysis represents an extension of a "twin" or "clone" design. While the two males within each arena were genetically identical, they were heterozygous between arenas. This means that, in addition to additive & non-additive genetic effects, covariance between (social) environmental effects and genotype cannot be separated out from estimates.

Reference

[1] Shorter, J., Couch, C., Huang, W., Carbone, M.A., Peiffer, J., Anholt, R.R.H. & Mackay, T.F.C. 2015 Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior. *Proceedings of the National Academy of Sciences* **112**, E3555-E3563. (doi:10.1073/pnas.1510104112).

[2] Lenth, R. 2022 emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.3. https://CRAN.R-project.org/package=emmeans.