Biology 708: Final Project Report for Brandon McIntyre and Rajat Bhargava

Biological Background

The genetic background has a known measurable effect on phenotypic expression, but the mechanisms of how the genetic background affects allelic expression is poorly understood. Genetic background effects can be defined as “the genotype of all other related genes that may interact with the gene of interest, and therefore potentially influence the specific phenotype” (Yoshiki & Moriwaki, 2006). Examples of genetic background effects can be seen with canalization. Canalization can be defined as “a process by which the phenotypic variance of a trait is reduced when faced with a given perturbation” (Félix & Barkoulas, 2015). The idea of canalization began with the early work of Waddington where he was able to introduce a heat perturbation to genetically identical lines of *D. melanogaster* which resulted in a crossveinless phenotype (Waddington, 1952). He continued this for several generations such that the phenotype persisted when the perturbation was not longer present (Waddington, 1952). This canalization took advantage of plasticity in a population, suggesting the alleles needed for this characteristic inheritance were within the population but not displayed until perturbation (Waddington, 1942). This implied that canalization depends on a trait’s ability to vary where decanalized traits have increased variability (Félix & Barkoulas, 2015; Green et al., 2017; Waddington, 1942).

Though work has continued with canalization, the mechanisms behind background dependence of genetic mutational effects is not clear. Currently two non-mutually exclusive mechanisms behind background genetic effects on canalization have been proposed. The first posits that canalization is due to individual alleles and/or developmental mechanisms that modulate the effects of other genetic variants (Green et al., 2017). The second posits canalization is due to properties of the evolved developmental network itself (Green et al., 2017). In either case these relationships are represented by a non-linear relationship between gene activity and phenotypic effect (Green et al., 2017).This can be seen in Figure 1 where a threshold of gene activity for a normal wild type trait must be met to ensure phenotypic expression (Green et al., 2017). In canalized traits, this threshold is mediated by compensatory mechanisms in the genetic background in response to perturbations minimizing phenotypic variance from wild type. In this model the canalized region where variance is buffered resulting in wild type phenotypic variation is in the blue shaded regions, while the decanalized region where there is much more variation resulting in mutant phenotypes is seen in the red shaded regions (Green et al., 2017).

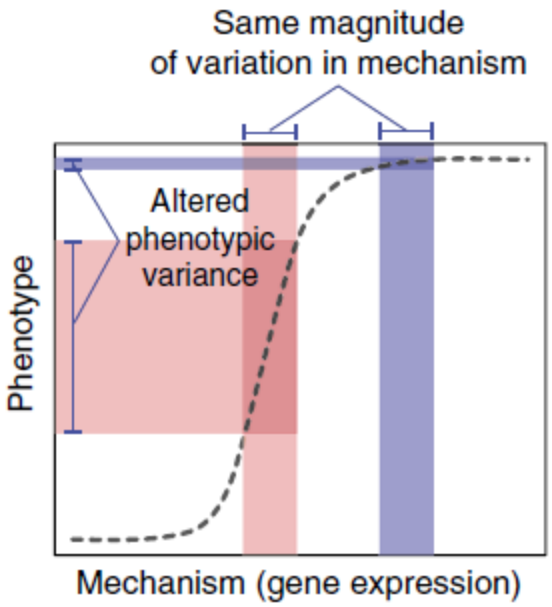


Figure 1: Non-linearities across development modulate variance. This is a general model for a non-linear genotype-phenotype map taken from Green et al. where the amount of a developmental process (gene expression) dictates the mean expressed phenotype (2017). Wild type gene expression can be seen in the blue column while mutant gene expression can be seen in the red column. Of note, the same amount of variation within the mechanisms can lead to different amounts of phenotypic variation. The canalized region where variation is buffered by the genetic background can be seen in the wild type gene expression (blue regions). Decanalization can be observed in the mutant gene expression where variation is not as well buffered by the genetic background (red regions). Due to decanalization phenotypic variance increases (Green et al., 2017).

Recent work by Chandler et al. has discovered a “goldilocks” phenomena where alleles of moderate phenotypic effect on a trait have the most sensitivity to genetic background effects while alleles with weak and severe phenotypic effects on a trait are the least sensitive to genetic background effects (2013). If we were to relate this to the graph by green et al. we could picture these mutant alleles or genetic perturbations as being on the top right of the sigmoidal curve for mild phenotypic effect and at the bottom left of the sigmoidal curve for severe phenotypic effect (2017). Of note the alleles of moderate phenotypic effect would be in the center and would be the most sensitive to background genetic effects as indicated by increased phenotypic variation.

Though we know what canalization is, how to measure variability and determine canalization is not agreed upon. Currently, there are two main methods to characterize the definitions of canalization. The first is reaction norm of the mean, the second is the variation approach (Dworkin, 2005). The reason this is important is because both methods to measure canalization depend on a different metric of measurement. The reaction norm of the mean is primarily concerned with changes of genetic line means and how variability of these line means changes from one genetic environment to the next, as seen in figure 2 (Dworkin, 2005). The second variation approach is fundamentally concerned with not how the line means change but how the a measure of variation around line means changes, as seen in figure 3 (Dworkin, 2005).

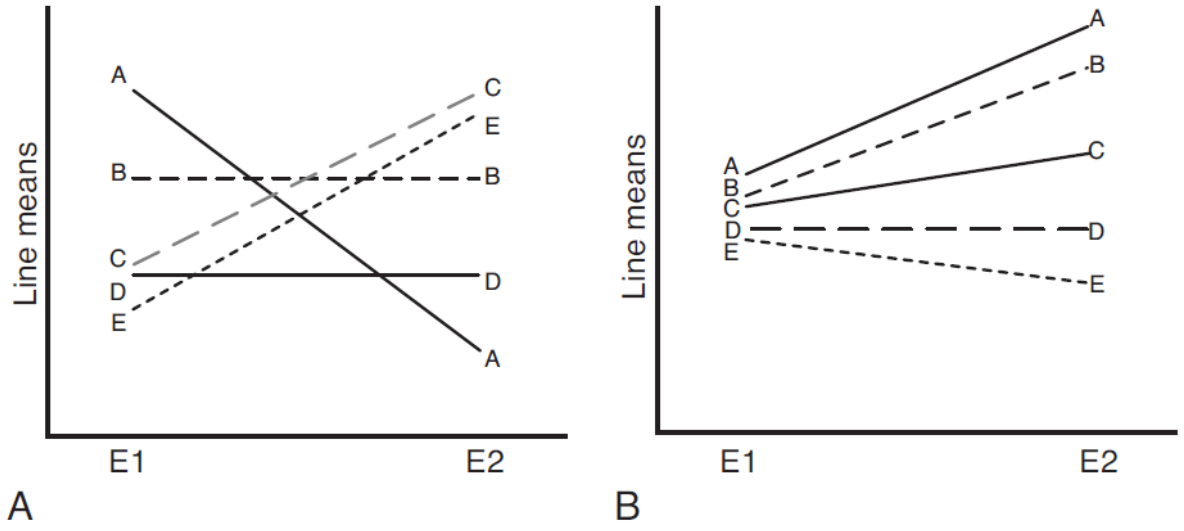


Figure 2: Reaction norm of the mean approach to canalization. This is a figure taken from Dworkin demonstrating phenotypic plasticity as the opposite to canalization (2005). Each line (A-E) represents a genotype where multiple individuals have been sampled in two environments. (E1 & E2) These environments can be thought of as "treatments" whether that be a genetic perturbation via mutant allele or raising of the lines in different temperatures. **A** Each line displays different trait means (wing size or leg length) which can change across environments. Lines' whose trait means remain constant across environments (B & D) are said to be canalized with respect to the environments (E1 & E2) **B** An idealized example of canalization where each line shows little between line variation in one environment (E1) and a significant increase in between line variation in another environment. (E2) This is the goal with reaction norm of the mean approach, where the more canalized a line is the less it's mean should change across environments (Dworkin, 2005).

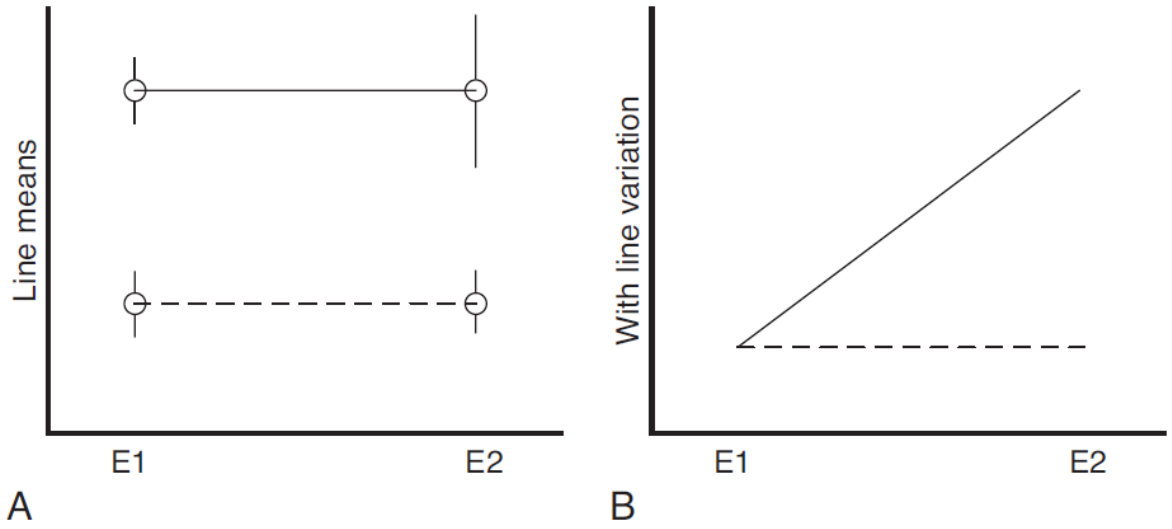


Figure 3: Variation approach to canalization. This is a figure taken from Dworkin demonstrating that Canalization is determined through how the measure of within line variation changes (2005). Two genetic lines are depicted as solid and striped lines. These lines have been sampled in two environments (E1 & E2). **A** The line trait means are depicted on the y-axis with a measure of within line variability depicted by the "error" bars. The genetic line represented by the dashed line displays canalization through no change in the within line variability across environments. **B** The same pattern is displayed, though the y-axis now represents the chosen measure of variability.

Levene's Statistic

The measures of variation and statistical tests for of within-line variation is a second point of contention for measuring canalization. Some common methods of measuring variance include the use of standard deviation and variance. These measurements are most useful within a sample and for comparing samples with normal distributions and equal means but can reflect differences in means or units used (Félix & Barkoulas, 2015; Schultz, 1985). A more common method is utilization of the coefficient of variation which adjusts for the variation for the size effect of the mean and/or unit size (Dworkin, 2005; Schultz, 1985). The downside to utilizing the coefficient of variation is the sample size required for an accurate estimate may be large and for small sample sizes it can be biased (Dworkin, 2005; Schultz, 1985; Sokal & Braumann, 1980). Therefore, a method that does not rely on normality may be useful. One of these methods is the levene's statistic, which under it's median form is the most robust (Dworkin, 2005; Félix & Barkoulas, 2015; Schultz, 1985; Sokal & Braumann, 1980). As such, during our analysis of within-line variation we utilized the levene's statistic in its median form as a measure of within line variability. This can be represented by equation 1, where the absolute difference of each individual is taken from the median of their respective group (Dworkin, 2005; Félix & Barkoulas, 2015; Schultz, 1985).

Equation 1:

$$LS_{ijk} = |x_{ijk} - Med[X_{ijk}]|$$

Where given k lines and j environments LS_{ijk} is the levene's statistic for individual i, for line j, in environment k. $Med[X_{ijk}]$ is the median of the data for all individuals for line j in environment k.

Log transformation

The log transformation is used to rescale the actual measurements from an experiment, to better meet the assumptions associated with a statistical analysis. There are two main ways that log transformation can help with this. First, if the variability (measured by standard deviation) varies roughly proportionately with the mean value of the response variable, a log transformation of the actual observations can equalize the standard deviations. This helps in fulfilling the assumption of homoscedasticity. Second, a log transformation can help meet the assumption of the distribution of actual observations having normality (Curran-Everett, 2018). Curran-Everett (2017) discovered that a one-sample t-test's results are meaningful only if the sample mean's theoretical distribution is approximately normal. So perhaps more often, a log transformation helps meet the normality assumption by making the theoretical distribution of the sample mean more normal (Curran-Everett, 2018).

Is a log transformation always the correct thing to do however? Feng et al. (2013) state that a log transformation does not always reduce skewness, meaning it does not always in moving towards meeting the condition of normality. Curran-Everett (2017) found that if the sample size is big enough, the theoretical distribution of the sample mean could be approximately normal,

regardless of the distribution of the actual observations. In such a case, a transformation would not be warranted. How big is big enough if the population is not normally distributed however? Hesterberg (2015) used simulations to find that a sample size of over 5000 might be required when a population is even moderately skewed. Even in considering this, Curran-Everett (2017) says that there is not hard-and-fast rule regarding size.

A way to determine if a log transformation is appropriate or not is the Box-Cox method. It considers a family of transformations on response variables with positive values. It has an associated likelihood function, which given a sample and a parametric family of distributions that could have formed the sample, associates the probability of observing the given sample to each parameter (Box, 1964). The equation for the associated likelihood function is:

Equation 2:

$$g_\lambda(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases},$$

Where y is the responsible variable, λ is the transformation parameter, and g λ(y) is the log-likelihood value.

The λ value that maximizes the log-likelihood value is chosen as the appropriate transformation parameter. It gives an indication of whether a transformation is appropriate or not, and if it is, which transformation might be most appropriate. The likelihood function, given a sample and a parametric family of distributions that could have formed the sample, associates the probability of observing the given sample to each parameter (Taboga, 2017).

Curran-Everett (2018) used the Box-Cox method on 10,000 observations from a skewed distribution used by Feng et al. (2013). This was done using the "boxcox" function from the "MASS" package on R (Venables & Ripley, 2002). An approximate 95% confidence interval of [0.50, 0.54] for λ was determined, and since this interval excludes 0 (the function becomes the log function when λ is equal to 0), it was concluded that a log transformation of these observations is not valid.

Biological Question

The biological question we are attempting to answer is how the variation around the trait mean of wing size varies between DGRP lines crossed with different mutant alleles. We hypothesize that mutant alleles with moderate phenotypic effects will display the greatest variability in within line wing size variability for dgrp lines crossed with mutant alleles.

Our Dataset

The data we are utilizing for our analysis was collected by MSc Caitlyn Daley utilizing *D. melanogaster* in Dr. Ian Dworkin's lab. Within our dataset we have 5496 individual fly wing size measurements. These individuals are spread across the scalloped (sd) and beadex (bx) allelic series in the Oregon white genetic background. The sd allelic series consists of sd[29.1], sd[1], sd[E3], sd[ETX4], sd[58d] from weak to severe phenotypic effects. The bx consists of bx[1],

bx[2], and bx[3] ranging from slightly moderate to moderately severe phenotypic effects. Therefore, including the Oregon white wild type background our dataset has a total of 9 genetic perturbation environments to test lines in. Our dataset has 20 distinct wild type genetic lines taken from the Drosophila Genetic Reference Panel (DGRP), each line and their corresponding lookup number is provided in table 1. (supplementary tables) The experiment was conducted over 2 replicate blocks initially but due to unforeseen circumstances some crosses and in replicate block 2 were lost necessitating a subset replicate block 2a. Each mutant allele was crossed to each DGRP line for a total of 180 unique genotypes with 3-23 individuals measured per genotype per replicate block.

Statistical analysis

The data utilized for our statistical analysis can be found on our repo: https://github.com/b-mcintyre/the_fly_guys within the file titled "NEW_CD_DGRP_Subset_Data_2019_V2.csv. The code for our statistical analysis can be found in the file titled "Final Draft.R". The code for the function we utilize to calculate levene's stat can be found in the file titled "ID_LeveneStat_V1_2016.R" and was provided by Dr. Ian Dworkin. The graphics from our analysis can be found in the folder titled "graphics".

Data Exploration and Visualization

We began our statistical analysis by visualizing our data by plotting line means against line standard deviations and line coefficients of variation utilizing the ggplot2 package (Wickham, 2016). We did this in aggregate and by alleles as seen in figures 4 and 5. We noticed that the alleles of moderate phenotypic (sd[E3] & sd[ETX4]) effect seemed to display the largest standard deviations. To ensure this was not due to a scaling effect we plotted the coefficient of variation against the means and observe the same pattern of variability with the coefficient of variation in figure 5. Next, we visualized the levene's statistic through a reaction norm of the mean plot of the raw data as seen in figure 6. This graph displays the changes in variability and unlike in the previous plots we can observe bx[2] having increased variability compared to bx[1] and b[3] while also observing increased variability of the scalloped moderate phenotypic effect alleles sd[E3] and sd[ETX4]. We then plotted a reaction norm of the mean graph for a gauge of changes in line means, as seen in figure 7. We can observe increased among line variability of moderate phenotypic effect alleles sd[E3], and sd[ETX4] but not a noticeable difference in bx[2]. To view the increased variability of moderate phenotypic effect alleles we plotted boxplots of wing size in mm by DGRP line and the weak phenotypic effect allele sd[1], moderate phenotypic effect allele sd[E3], and severe sd[58d] allele. This plot can be seen in figure 8 where we can observe increased within-line variation through the drastic change in the interquartile ranges of DGRP line means across the mutant environments with the moderate effect sd[E3] showing the largest interquartile ranges. From this plot we can also view increased among line mean variation across the mutant environments through changes in line means.

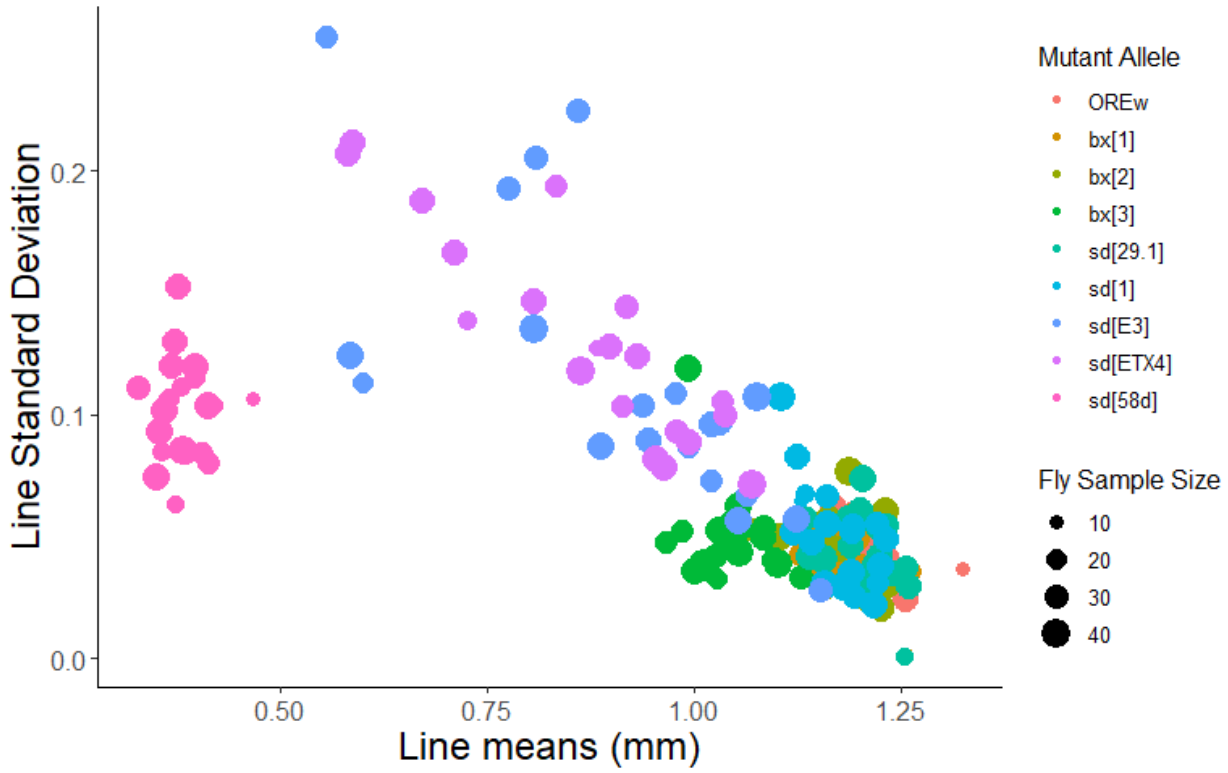


Figure 4: Line standard deviations plotted against line means in mm. The fly sample size represents the number of individuals within each genotype. Individual mutant alleles are colour coded. From this plot we can observe that the moderate phenotypic effect alleles seem to have the largest standard deviation.

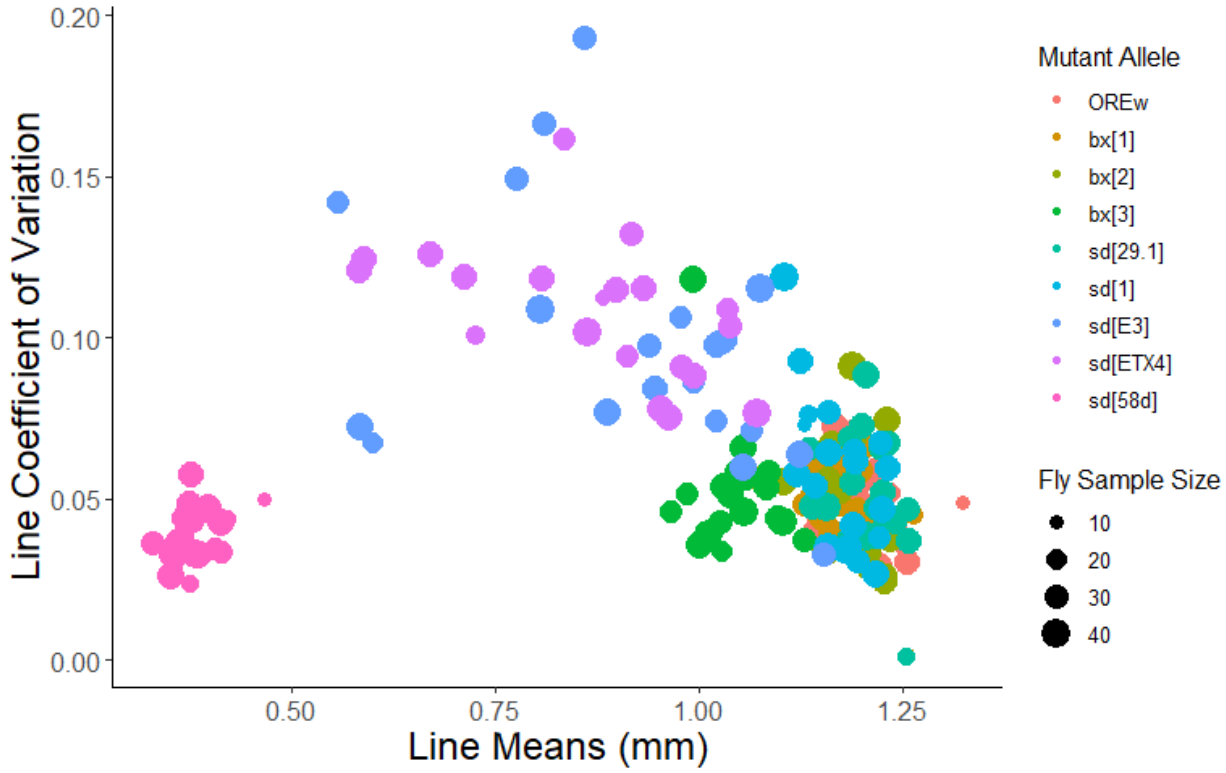


Figure 5: Line coefficient of variation plotted against line means in mm. The fly sample size represents the number of individuals within each genotype. Individual mutant alleles are colour coded. We can observe the moderate phenotypic effect alleles display increased coefficients of variation.

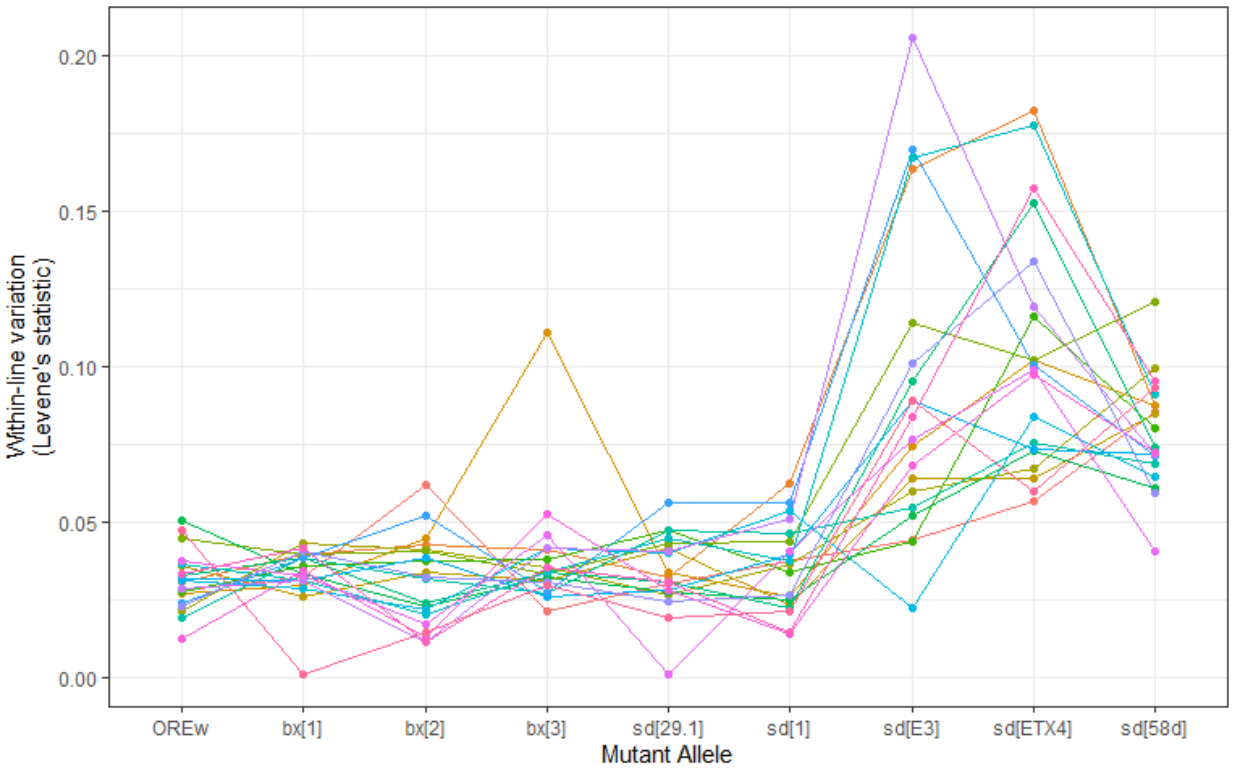


Figure 6: Reaction norm of the mean graph of raw levene's statistic. We can observe increased within-line variability of moderate phenotypic effect alleles bx[2], sd[E3], and sd[ETX4]. Each coloured line represents a distinct DGRP strain. The mutant alleles are represented on the x-axis with wild type followed by the bx allelic series and sd allelic series from weak to severe phenotypic effects.

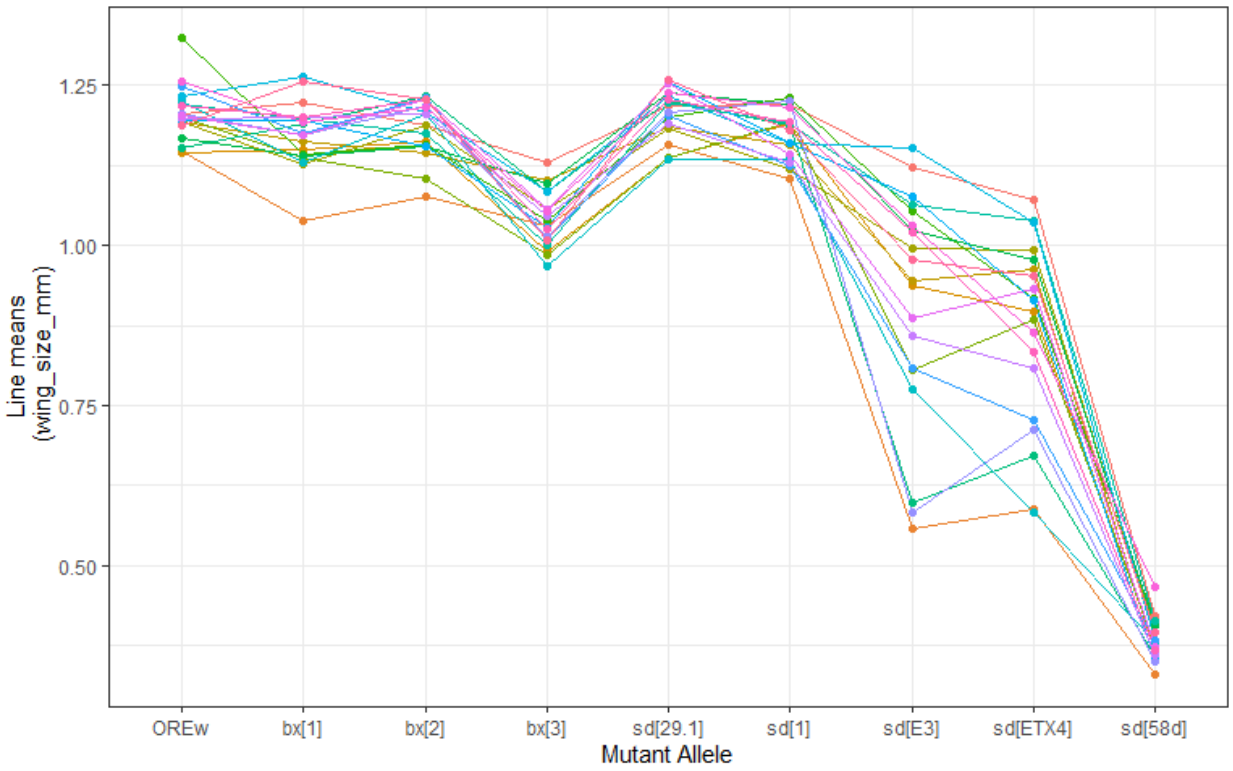


Figure 7: Reaction norm of the mean graph of raw line means. We can observe increased among line variability of moderate phenotypic effect alleles sd[E3], and sd[ETX4]. Each coloured line represents a distinct DGRP strain. The mutant alleles are represented on the x-axis with wild type followed by the bx allelic series and sd allelic series from weak to severe phenotypic effects.

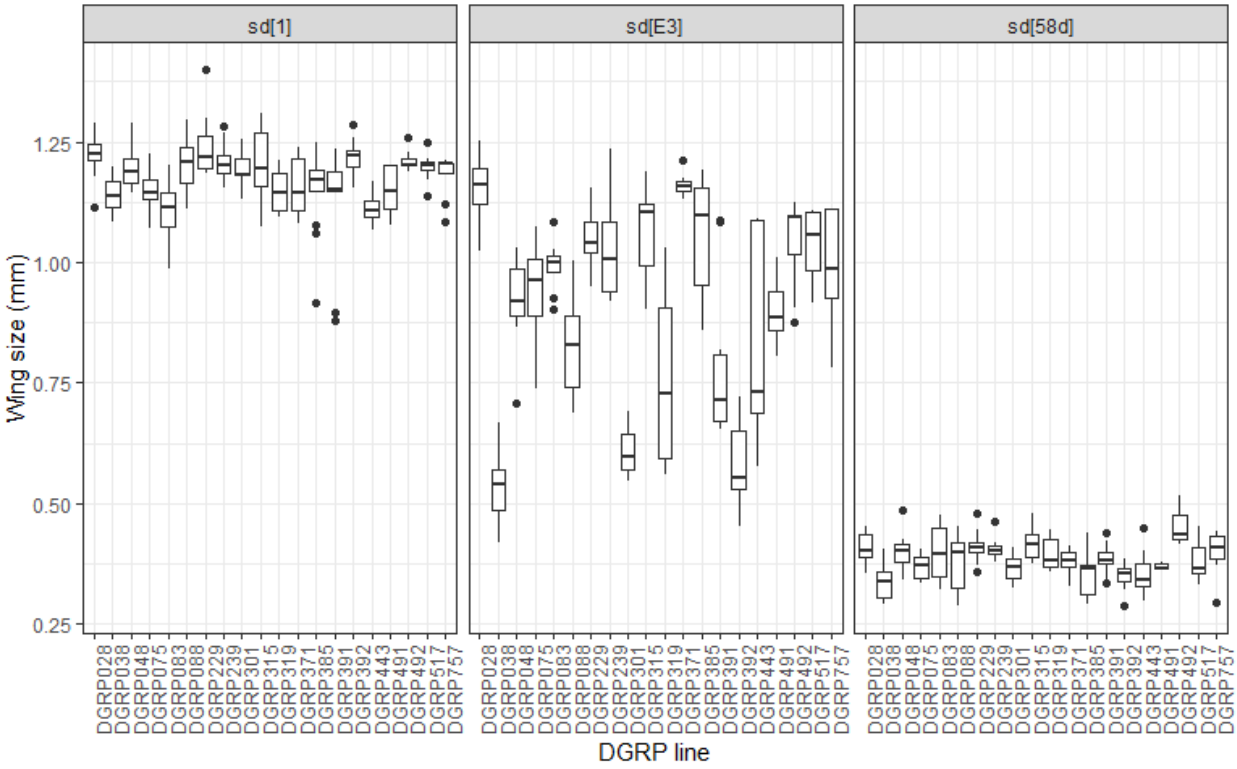
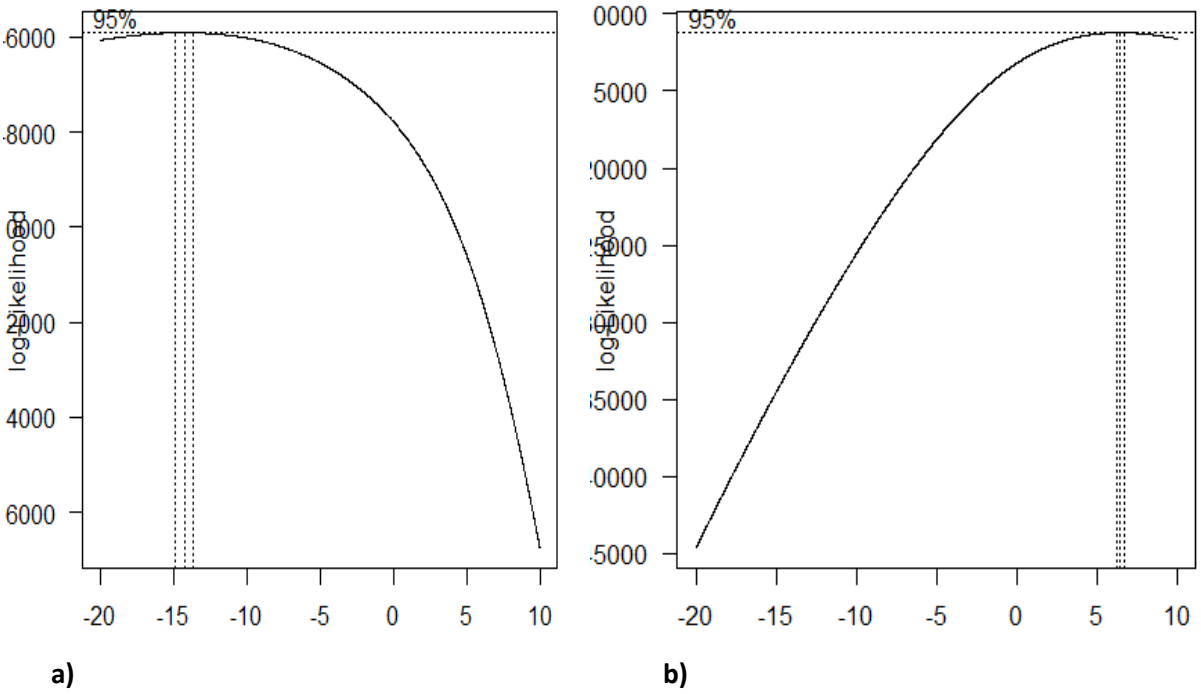


Figure 8: Boxplot of weak, moderate, and severe mutant alleles. Can observe the increase in within line variability as each of DGRP lines' interquartile range drastically increases in the moderate phenotypic effect environment sd[E3] relative to the weak phenotypic effect environment sd[1] and severe phenotypic effect environment sd[58d].

To Log Transform, or Not to Log Transform?

Before we could go ahead fit a model, we needed to see if it is appropriate to log transform the values in our dataset. We followed the R code provided by Curran-Everett (2018) for the Box-Cox analysis on our actual values. We found that for our levene's statistic values, the 95% confidence interval of [-14.83, -13.66] for λ did not include a value of 0. Thus, we decided it was best not to do a log transformation with the values. The plot for this analysis can be seen in figure 9a. For the analysis using our wing size values, the confidence interval of [6.24,6.69] for λ did not include a value of 0, and thus we decided it was best not to do a log transformation with these values. The plot for this analysis can be seen in figure 9b. We are aware that these values are outside of the -5 to 5 range for λ that would be indicative of the proper transformation to perform. We could not however find any clear solution for how to transform the data for values that fall outside of this range, and thus proceeded with not transforming our data in any way.



a) b)

Figure : Log-likelihood curves derived from Box-Cox transformations of a) levene's statistic values and b) wing size values. λ is plotted along the x axis, and log-likelihood is plotted along the y-axis. The outer vertical dotted lines indicate the boundaries for the confidence interval of the λ value that maximizes log-likelihood. The center vertical line indicates a point estimate for the λ value that maximizes log-likelihood.

Linear Model Fitting

We then attempted to model both wing size and within-line variation to provide estimates of the changes in among line and within-line variation. We utilized lme4 with the lmer call, stating that the mutant allele was a fixed effect while DGRP lines and replicates were random effects in two separate models (Bates, Mächler, et al., 2015; Bolker, 2015). Our decision to utilized mixed models is based on the design of our dataset. Due to our dataset containing 9 mutational environments with 20 DGRP groups per environment and an uneven level of individual counts per mutant per DGRP per replicate ranging from 3-23 we believed a mixed model approach would be the best choice (Bolker, 2015; Bolker et al., 2009). Our initial model resulted in a singular fit. We feel it is appropriate to talking about the theory behind singular fitting here before proceeding.

Singular Fits and Principle Component Analysis

A singular fit occurs when a mixed model is overfitted. This occurs when there are a large number of variance-covariance parameters. This variance-covariance matrix ends up having some zero eigenvalues, which less technically is the same thing as saying that some of the variances are estimated to be zero. The end up being less than full rank, and is described as rank deficient (Bolker, 2015). For scalar random effects such as intercept-only models, or 2-dimensional random effects such as intercept+slope models, singularity is fairly simple to determine because it results in random-effect variance estimates of zero or nearly zero, or

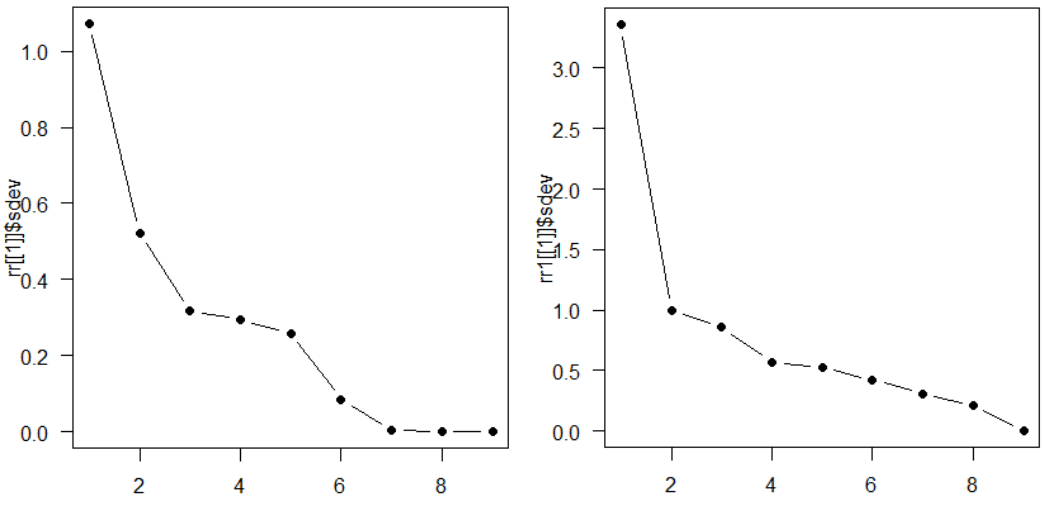
correlation estimates that are exactly or almost exactly -1 or 1. However, for more complex models (variance-covariance matrices of dimension >=3), a singular fit can be harder to detect. Models can often be singular without any of their individual variances being close to zero or correlations being close to +/-1 (Bates, Kliegel, et al., 2015).

PCA (Principal Component Analysis), is a dimensionality-reduction method. It performs "singular value decomposition", which can be used to resolve the rank deficiency problem that was mentioned earlier. In essence, it is used to reduce the dimensions of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set. principal components are new, uncorrelated variables that are computed from the eigenvectors and eigenvalues of a variance-covariance matrix. PCA tries to put maximum information in the first principle component, the second most in the second component, and so forth (Holland, 2008). rePCA (random-effects Principal Components Analysis) takes a model fit by lmer and produces a list of principle components (Bates, Kliegel, et al., 2015). One can attempt to resolve a singular fit issue by removing principal components from a model. General recommendations for removing principal components include ignoring principal components at the point at which the next one offers little increase in the total explained variation, and ignoring the last principal components whose variation are all roughly equal. It is useful to use a skee plot to visualize the principal components (Holland, 2008).

Linear Model Fitting Continued

Initially we were suggested to treat the replicates, which only had three levels, as fixed effects. This is because a singular fit can result from treating a variable with five or less levels as a random effect, and this is a common prescription for resolving this (Bolker, 2015). However, when we attempted to run this model, we still encountered a singular fit.

After this Dr. Bolker created a reduced rank model for the levene's statistic utilizing 5 principle components utilizing the glmmTMB package (Brooks et al., 2017). This involved using rePCA analysis to determine the number of principal components that should be removed. This involved dusing a skee plot, which can be seen in figure 10a. We then fit a similar model using 3 principal components for the wing size. The relevant skee plot for this can be seen in figure 10b.



a) **b)**

Figure 10: Skee plots of PCA analysis, for a) the model fit using the levene's statistic, and b) the model fit using wing size area. Standard deviation, which is a measure of variance, is on the y-axis, and principal component number is on the x-axis. Using these plots, we were able to determine that 5 principal components and 3 principal components would be appropriate for the levene's statistic and the wing size respectively.

Once we ran the reduced rank-model we conducted a type 3 anova on the within-line model utilizing the cars package (Fox & Weisberg, 2019). We decided on the type 3 anova because we believe the interactions to be significant. The results from the anova show that for the fixed effects intercepts and mutant background are significant but the replicates are not. We repeated this process with the wing size model and found that the fixed effects intercepts, mutant background, and replicate were significant.

Next, we created a correlation matrix of each of the mutant alleles with regards to within-line variation utilizing the corrplot package (Wei & Simko, 2017). This is an indication of how similar the responses are for the DGRP lines across the mutant alleles. In other words, if two alleles are highly correlated and the within-line variation (levene statistic) was larger in one mutant environment for a DGRP line then they tended to be larger in the contrasted mutational environment in the same line as well. From figure 11 we can observe the effects of the to moderate phenotypic effect alleles sd[E3] and sd[ETX4] are a moderately positively correlated relationship. The strongest correlation occurs between the weaker phenotypic effect alleles sd[29.1] and sd[1]. We repeated this process for correlation with respect to wing size, as see in figure 12. From figure 12 we can observe that most alleles are positively strongly correlated with respect to wing size. The strongest observed correlation occurs between bx[1] and sd[29.1]. The correlation between sd[E3] and sd[ETX4] was found to be 0.88.

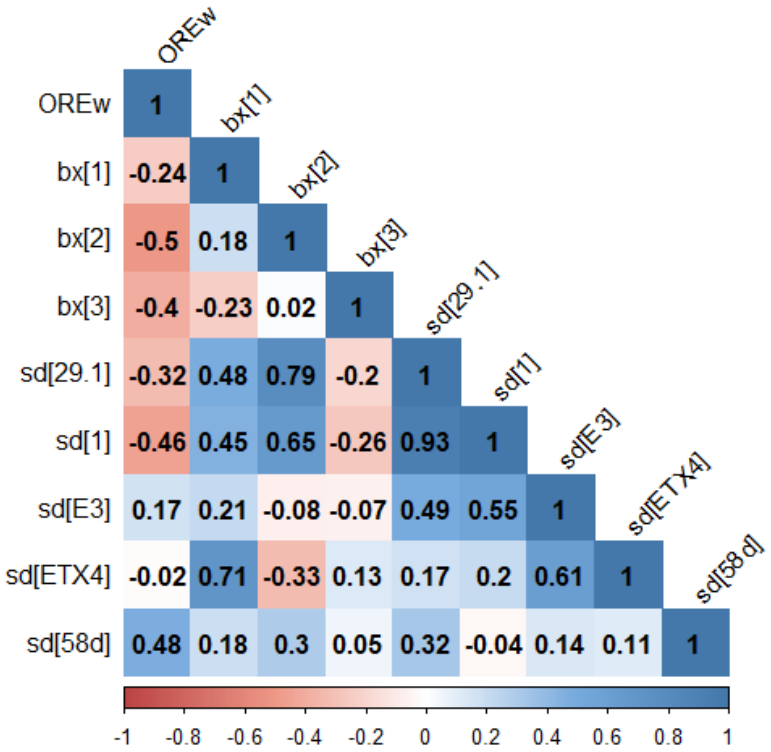


Figure 11: Correlation matrix for the mutant allele's effects on DGRP backgrounds with respect to within-line variation. Each mutant allele's effects correlated to other mutant allele effects. Can observe that the moderate phenotypic effect alleles sd[E3] and sd[ETX4] have positive moderate correlation.

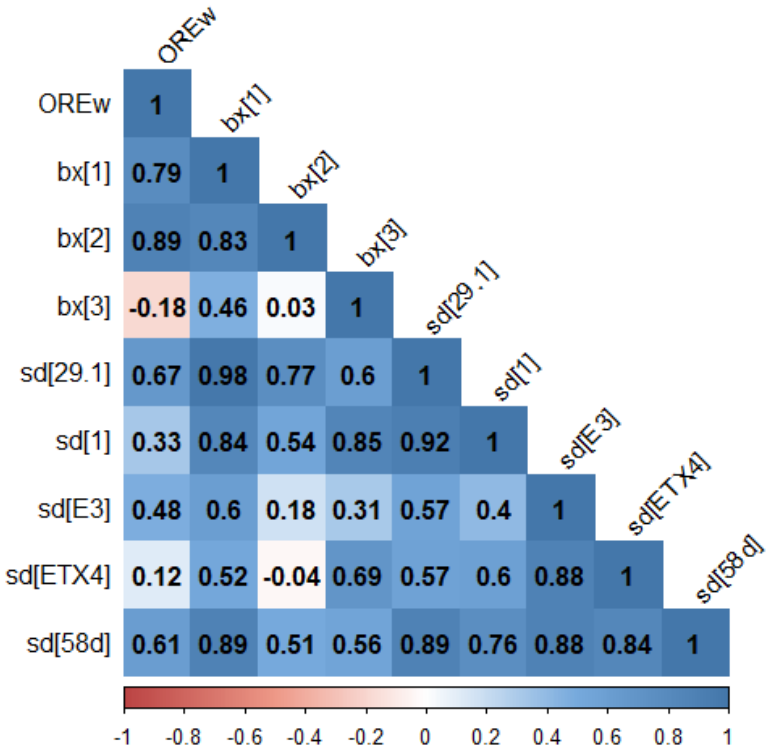


Figure 12: Correlation matrix for the mutant allele's effects on DGRP backgrounds with respect to among line mean variation. Each mutant allele's effects correlated to other mutant allele effects. Can observe that majority of effects are strongly positively correlated. The strongest observed correlation occurs between bx[1] and sd[29.1]. The correlation between moderate phenotypic alleles sd[E3] and sd[ETX4] was found to be 0.88.

Next we computed the estimated marginal means for the mutant alleles with their respective replicate trials and plotted them in figure 13 and 14 using the emmeans package for both the levene statistic and wing size (Lenth, 2021). From figure 13 we can observe that within line-variability is highest in the moderate phenotypic effect alleles. Difference in within line variability is statistically clear when comparing weak to moderate phenotypic effect alleles but is not clear comparing severe to moderate phenotypic effect alleles. Furthermore, this pattern is observed in the scalloped allelic series but is not statistically clear in the beadex allelic series. This pattern appears to hold true across each of the replicate trials. From figure 14 we can observe that the largest estimated marginal mean in OREw or Sd[1], it is not statistically clear which. It appears that bx[2] has slightly less of an effect on mean wing size than bx[1] but this is not statically clear. As expected, the sd[58d] and bx[3] clearly have the largest effects on mean wing size for their respective allelic series.

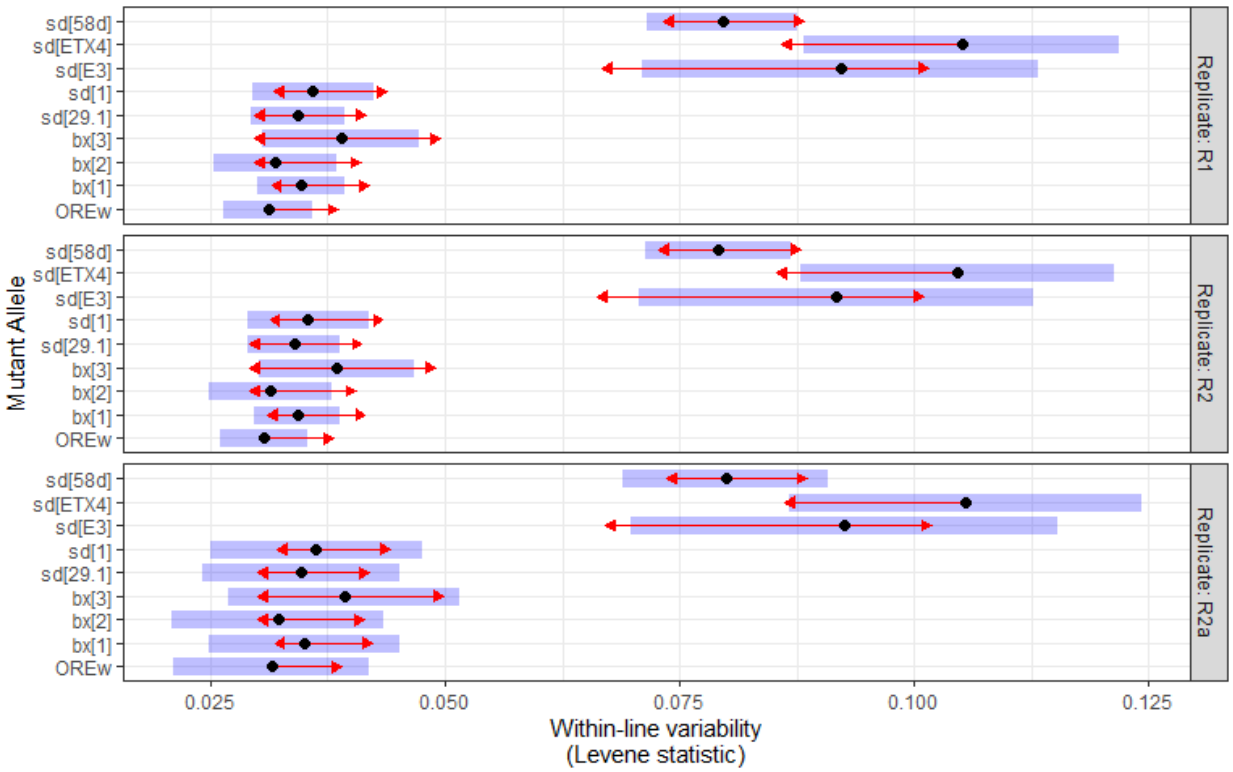


Figure 13: Estimated marginal means plot of the mutant alleles by fixed effect replicate blocks utilizing levene's statistic. Observe each estimated marginal mean(emmean) of each allele across replicate trials. The confidence intervals are depicted as blue bars and comparison arrows depicted to compare significance. Can observe the largest emmean in moderate phenotypic effect alleles sd[E3] and sd[ETX4], followed by sd[58d]. There is statistically clarity that these moderate phenotypic effect alleles have increased within-line variability relative to the weak phenotypic effect alleles. It is not statistically clear if these alleles have increased within-line variability relative to the severe phenotypic effect allele. The severely moderate effect bx[3] allele appears to have the largest within-line emmean's variability but it is not statistically clear from the other bx alleles in the series.

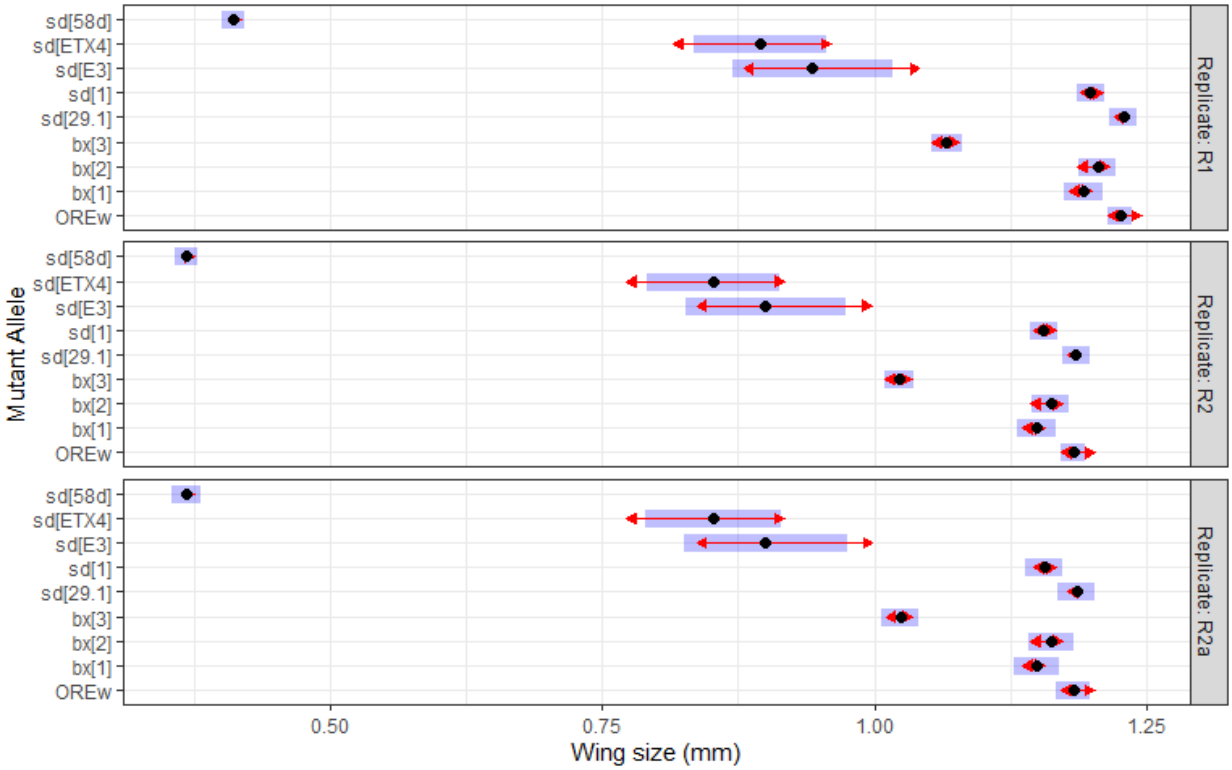


Figure 14: Estimated marginal means plot of the mutant alleles by fixed effect replicate block utilizing wing size (mm). Observe each estimated marginal mean (emmean) of each allele across replicate trials. The confidence intervals are depicted as blue bars and comparison arrows depicted to compare significance. Can observe the largest emmean in OREw and Sd[1]. It appears that bx[2] has slightly less of an effect on mean wing size than bx[1] but this is not statically clear. As expected, the sd[58d] and bx[3] have the largest effects on mean wing size.

Finally, we attempted to create a reaction norm of the mean graph for the levene's statistic utilizing the coefficients from our rank reduced model. We ran into the problem where the coefficients for the sd[1], sd[E3], sd[ETX4] and sd[58d] DGRP lines were identical. Dr. Bolker then created a new data frame with each of the mutant and DGRP combinations, which was then filled with the predicted levene's statistic utilizing the rank reduced models. Included in this was the confidence interval which was calculated as two times the predicted standard error in either direction of the predicted means. The results of these predictions were plotted as a reaction norm of the mean plot as seen in figure 15. From these predictions we can observe that the moderate phenotypic effect alleles have increased within-line variability relative to the weak phenotypic effect alleles. It appears that the moderate phenotypic effect alleles have some lines that are greater in variability than the severe phenotypic effect alleles though this trend is not as clear. We repeated this process for the line means as seen in figure 16. From these predictions we can observe that the moderate phenotypic effect alleles for the scalloped allelic series have increased among line variability relative to the weak and severe phenotypic effect alleles. This trend is not as clear when examining the beadex allelic series.

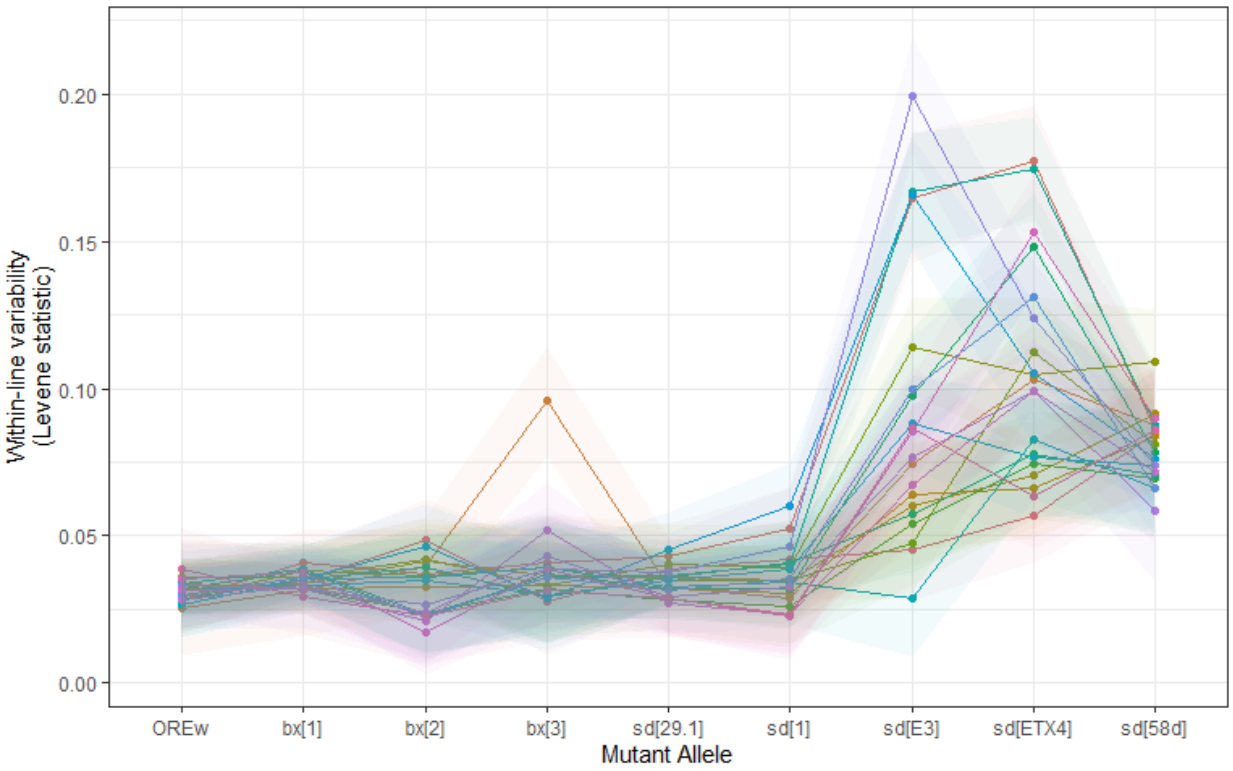


Figure 15: Reaction norm of the mean of predicted levene's statistic. Each DGRP line is represented as a distinct colour. The confidence intervals of each line are correspondingly colour coded as transparent shading around each line. The trend for moderate phenotypic effect alleles increasing in within-line variability for the scalloped allelic series and is less apparent in the beadex allelic series.

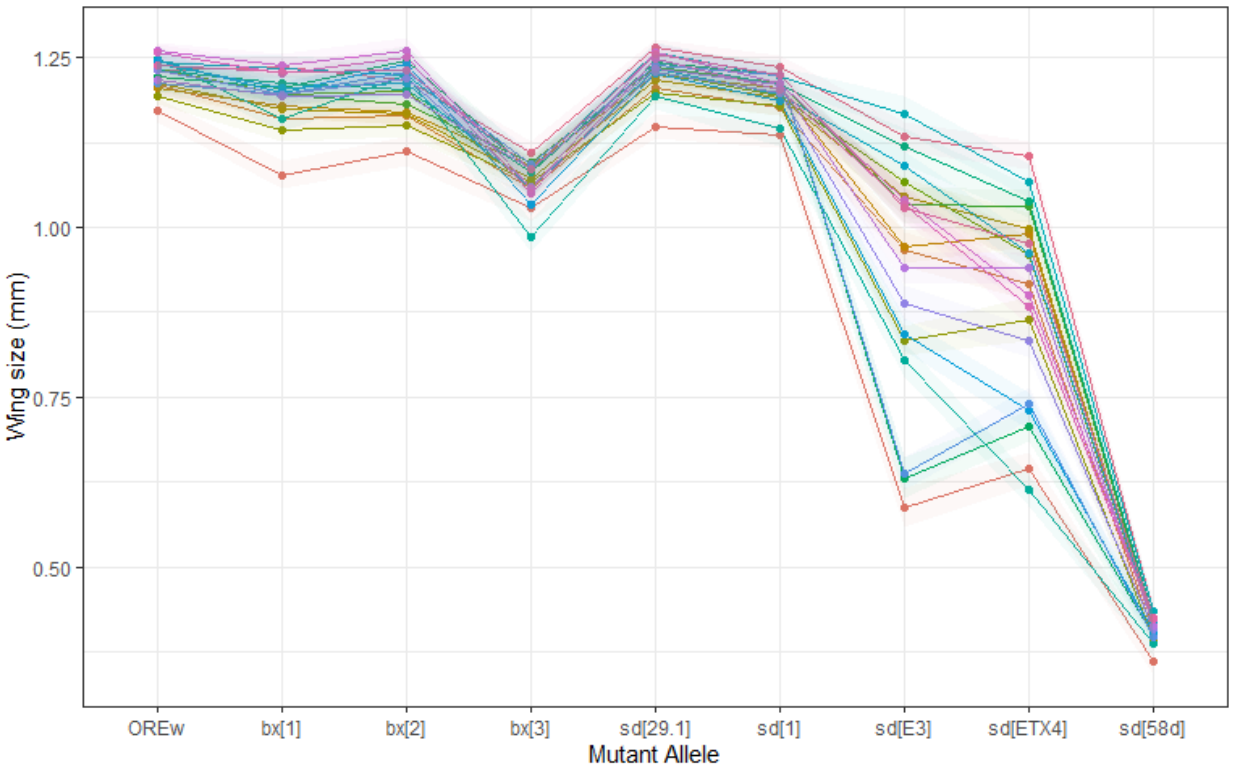


Figure 16: Reaction norm of the mean of predicted line means. Each DGRP line is represented as a distinct colour. The confidence intervals of each line are correspondingly colour coded as transparent shading around each line. The trend for moderate phenotypic effect alleles increasing in among line variability though this trend is not as evident within the bx allelic series.

Modeling Deviations from wild type wing size and within-line variability

To attempt to answer our biological question on how mutant allele severity affects within line variation we looked at the relationship between the within line variability (levene's statistic) and deviations of mutant alleles from their respective wild types in their respective DGRP lines. We calculated this deviation by subtracting the wild type line means from the mutant means per DGRP line. The plot to visualize this information can be seen in figure 17 where we observe a linear trend of increased within-line variation as deviations away from the wild type increases. We calculated a Pearson correlation coefficient and found it to be extremely high at 0.995. We then attempted to model this utilizing the mutant allele and deviation estimates as fixed effects and the DGRP lines as random effects. When attempting this we encountered convergence failures and attempted to rectify this by checking our optimizers and scaling and centering our deviations variable no success. We then tried simplifying our model by removing Genotype and still had convergence failures. Finally, we fit a linear model, but when we examined the summary received a warning message of a near perfect fit. We believe this is due to utilizing the two models previously to create predicted values that was used for our model.

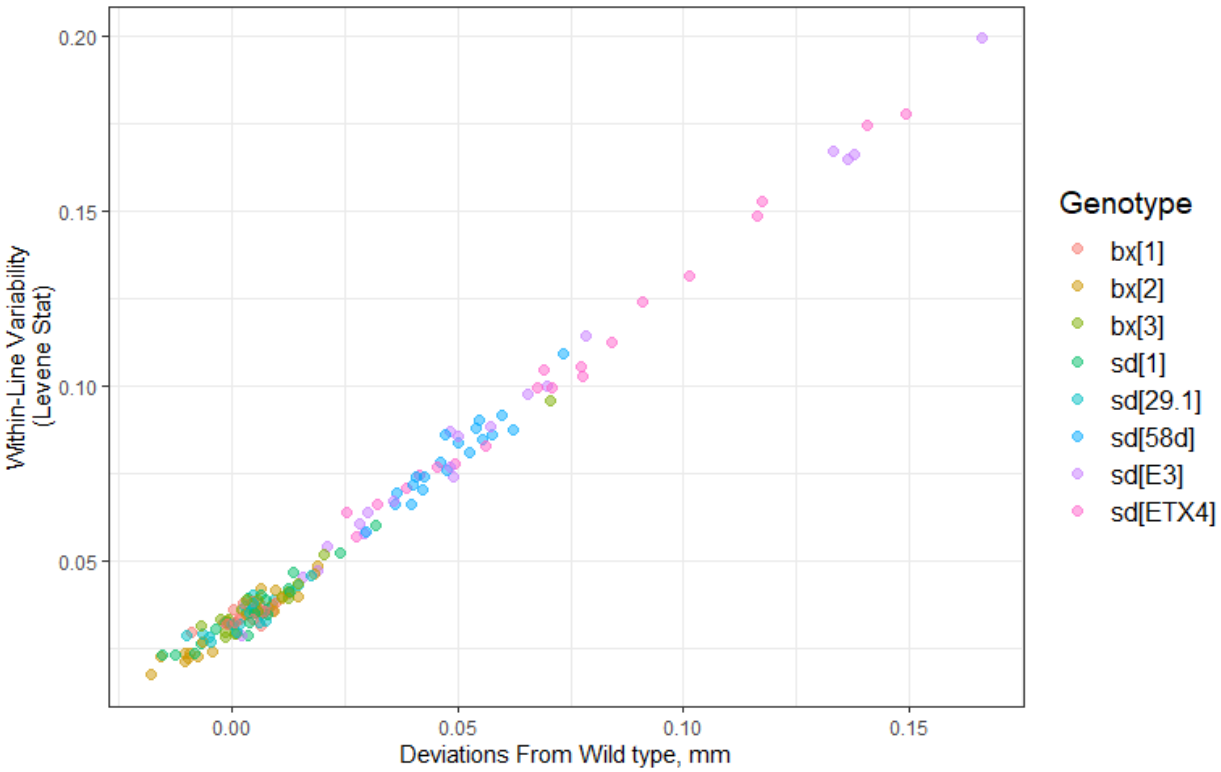


Figure 17: **Within-line variability plotted with Deviations from wild type.** The levene's statistic for each mutant allele DGRP line and its associated deviations from their respective wild type. Pearson correlation coefficient = 0.995

Interpretations

From these results we believe that within-line variability increases with respect to moderate phenotypic effect alleles in the scalloped allelic series. There is some indication that the moderate phenotypic effect alleles in the beadex series are increasing in variability relative to the weak phenotypic effect alleles though this trend will require further clarification. Likewise, there is some indication that moderate phenotypic effect alleles in the scalloped and beadex series display increased within-line variation in comparison to severe phenotypic effect alleles, though this is again not as clear as the weak to moderate comparisons. Furthermore, it appears that the moderate phenotypic effect alleles are moderately correlated for their effects on within-line variability across the DGRP backgrounds. Furthermore, we can qualitatively observe an increase in among-line variation through the variation of changes in means of DGRP lines in mutational environments. Though these are more qualitative results and future steps would be to investigate how the exact magnitude of the perturbation is related to the quantitative magnitude of within-line variability. A future direction could be utilizing a Bayesian approach that allows for specifying an informative prior for the variance-covariance matrix to avoid the singularity issues we encountered (Bolker, 2015). We could do this utilizing MCMC with half-Cauchy priors in WinBUGS (Hadfield, 2010).

References

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bolker, B. M. (2015). Linear Mixed and Generalized Linear Mixed Models. In and V. J. S. G. A. Fox, S. Negrete-Yankelevich (Ed.), *Ecological Statistics: Contemporary theory and application* (pp. 310–334). Oxford University Press. https://doi.org/10.1201/9781351165761-5

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution*, *24*(3), 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211-243.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal*, *9*(2), 378–400. https://doi.org/10.32614/rj-2017-066

Chandler, C. H., Chari, S., & Dworkin, I. (2013). Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends in Genetics*, *29*(6), 358–366. https://doi.org/10.1016/j.tig.2013.01.009

Curran-Everett, D. (2017). Explorations in statistics: the assumption of normality. *Advances in physiology education*, *41*(3), 449-453.

Curran-Everett, D. (2018). Explorations in statistics: the log transformation. *Advances in physiology education*, *42*(2), 343-347.

Dworkin, I. (2005). Canalization, cryptic variation, and developmental buffering: A critical examination and analytical perspective. In B. & H. K. B. Hallgrímsson (Ed.), *Variation* (pp. 131–158). Academic Press. https://doi.org/10.1016/B978-012088777-4/50010-7

Holland, S. M. (2008). Principal components analysis (PCA). *Department of Geology, University of Georgia, Athens, GA*, 30602-2501.

Félix, M. A., & Barkoulas, M. (2015). Pervasive robustness in biological systems. *Nature Reviews Genetics*, *16*(8), 483–496. https://doi.org/10.1038/nrg3949

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Green, R. M., Fish, J. L., Young, N. M., Smith, F. J., Roberts, B., Dolan, K., Choi, I., Leach, C.

L., Gordon, P., Cheverud, J. M., Roseman, C. C., Williams, T. J., Marcucio, R. S., & Hallgrímsson, B. (2017). Developmental nonlinearity drives phenotypic robustness. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-02037-7

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R package. *Journal of Statistical Software*, *33*(2), 1–22. https://doi.org/10.18637/jss.v033.i02

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, *69*(4), 371-386.Lenth, V. R. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means* ({R package version 1.5.5-1). https://cran.r-project.org/package=emmeans

Lenth, V. R. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means* ({R package version 1.5.5-1). https://cran.r-project.org/package=emmeans

Schultz, B. B. (1985). Levene's Test for Relative Variation. *Systematic Zoology*, *34*(4), 449. https://doi.org/10.2307/2413207

Sokal, R. R., & Braumann, C. A. (1980). Significance tests for coefficients of variation and variability profiles. *Systematic Biology*, *29*(1), 50–66. https://doi.org/10.1093/sysbio/29.1.50

Taboga, M. (2017). *Lectures on probability theory and mathematical statistics*. CreateSpace Independent Publishing Platform.

Venables, W.N., Ripley, B.D. (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, https://www.stats.ox.ac.uk/pub/MASS4/.

Waddington, C. H. (1942). Canalization of Development and the Inheritance of Aquired Characters. *Nature Publishing Group*, *150*, 563–565.

Waddington, C. H. (1952). Selection of the Genetic Basis for an Aquired Character. *Nature*, *169*, 278.

Wei, T., & Simko, V. (2017). *R package "corrplot": Visualization of a Correlation Matrix* (0.84). https://github.com/taiyun/corrplot

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Yoshiki, A., & Moriwaki, K. (2006). Mouse phenome research: Implications of genetic background. *ILAR Journal*, *47*(2), 94–102. https://doi.org/10.1093/ilar.47.2.94

Supplementary table

Table 1: DGRP genotype numbers and their respective stock numbers.

DGRP Genotype	DGRP stock number
DGRP028	28124
DGRP038	28125
DGRP048	55016

DGRP075	28132
DGRP083	28134
DGRP088	28135
DGRP229	29653
DGRP239	28161
DGRP301	25175
DGRP315	25181
DGRP319	55018
DGRP371	28183
DGRP385	28191
DGRP391	25191
DGRP392	28194
DGRP443	28199
DGRP491	28202
DGRP492	28203
DGRP517	25197
DGRP757	28226