
Figures and figure supplements

Genomic architecture and evolutionary antagonism drive allelic expression bias in the social supergene of red fire ants

Carlos Martinez-Ruiz *et al*

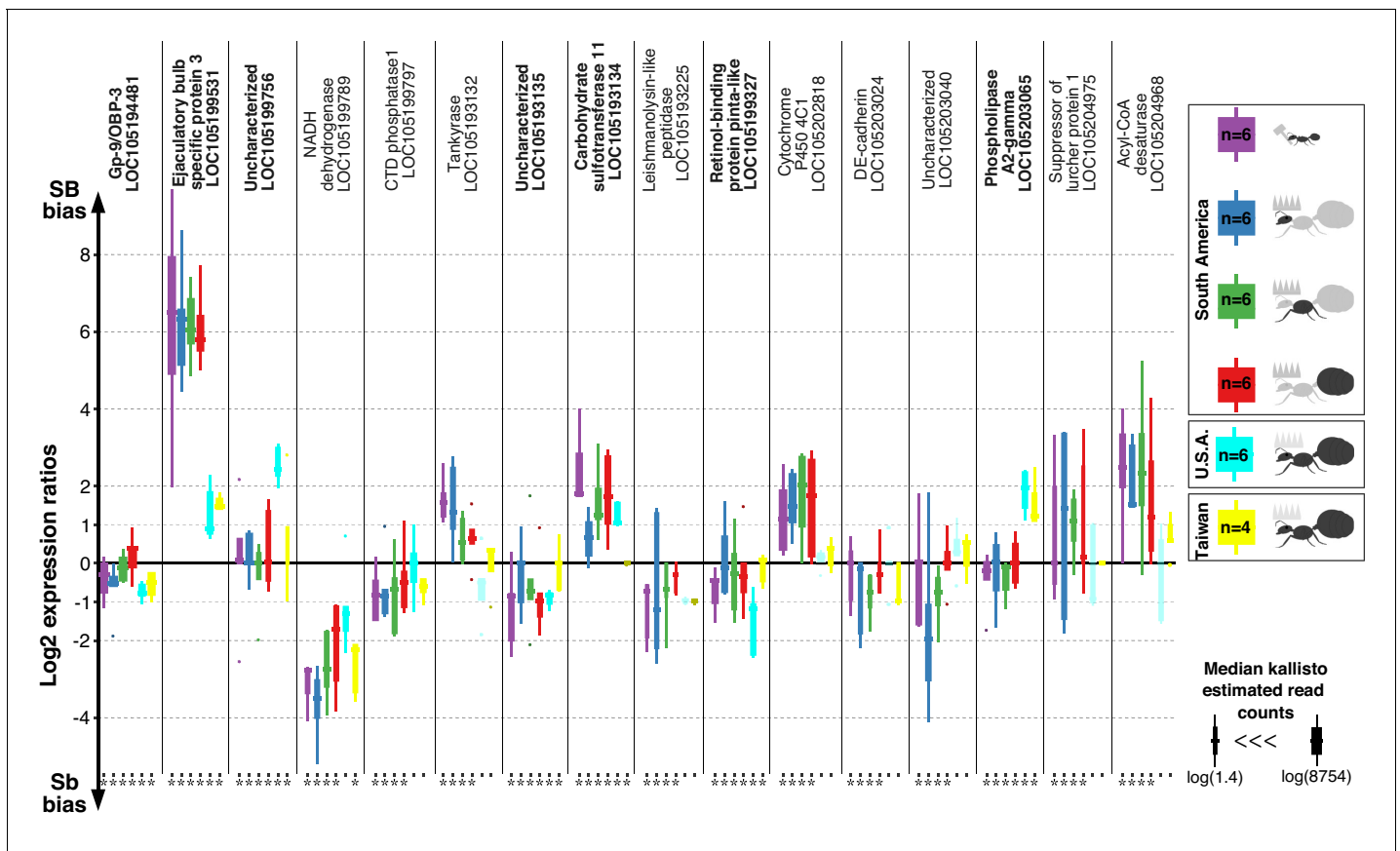


Figure 1. Differences in expression levels between alleles for genes in the fire ant social supergene in heterozygous SB/Sb individuals which exist only in multiple-queen colonies. Differences in expression (y axis) between social chromosome variants in whole bodies of workers from South America, heads, thoraces and abdomens of queens from South America, whole bodies of queens from North America and Taiwan. We show all 16 genes with significant allelic bias in South American populations, and the corresponding biases from the other populations. Bold gene names highlight when allelic bias occurs in all populations (Benjamini-Hochberg (BH) adjusted $p < 0.05$ from the joint linear mixed-effects model). Significance in population-specific comparisons is indicated by an asterisks under each boxplot (BH adjusted $p < 0.05$ from either the joint analysis or DESeq2 Wald tests). Each box shows the distribution of logarithm two expression ratios between SB and Sb in each sample type. Positive values indicate higher SB expression; negative values indicate higher Sb expression. A log2 expression ratio greater than one or smaller than -1 represents a two-fold gene expression difference in either direction. Genes are in chromosomal order.

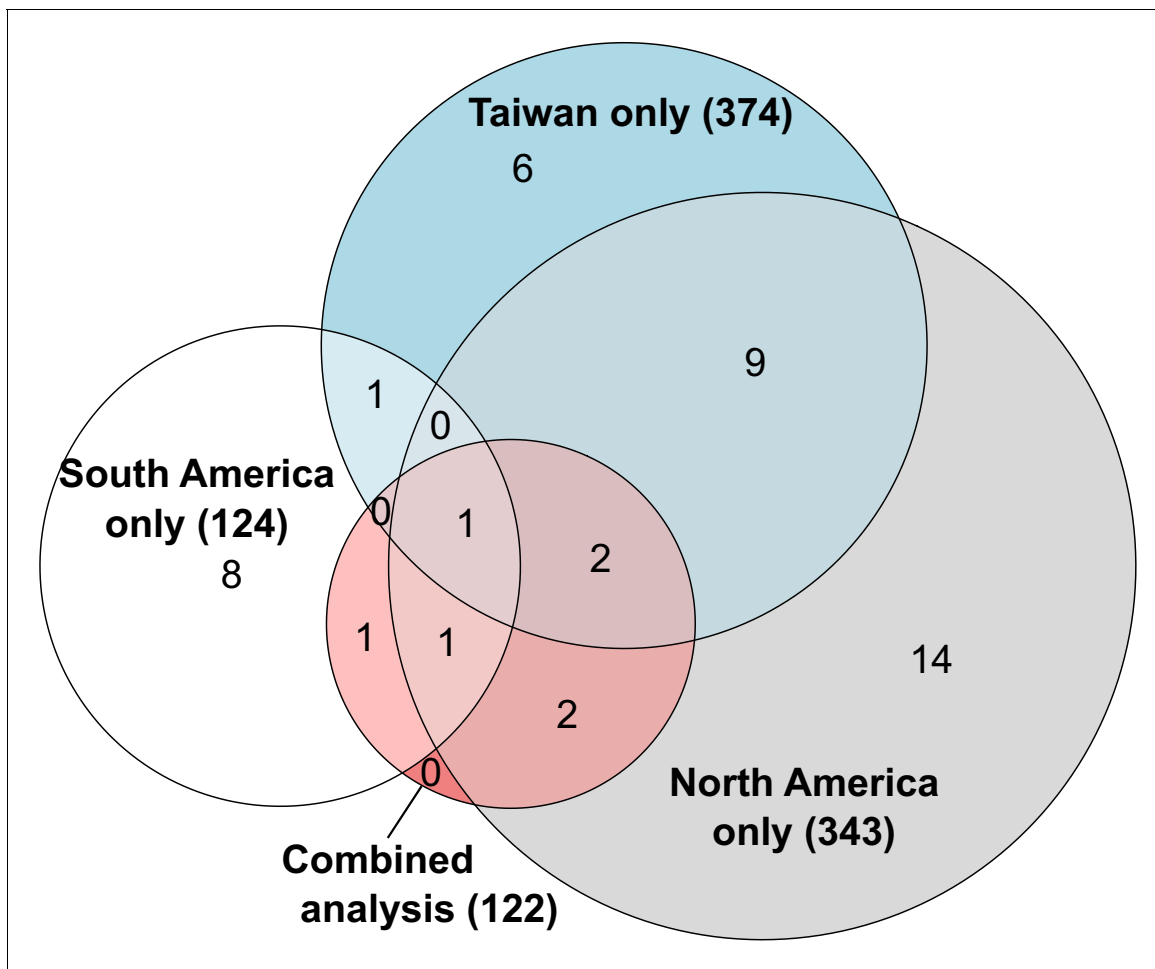


Figure 1—figure supplement 1. Overlapping number of genes with allele-specific expression according to comparisons in each population independently or after combining data from all populations. The combined analysis detected seven genes with allele-specific expression across both populations, three of which were independently detected using only South American populations, six using only North American populations and three using only Taiwanese populations. The numbers in parentheses indicate how many genes analyzed in each comparison.

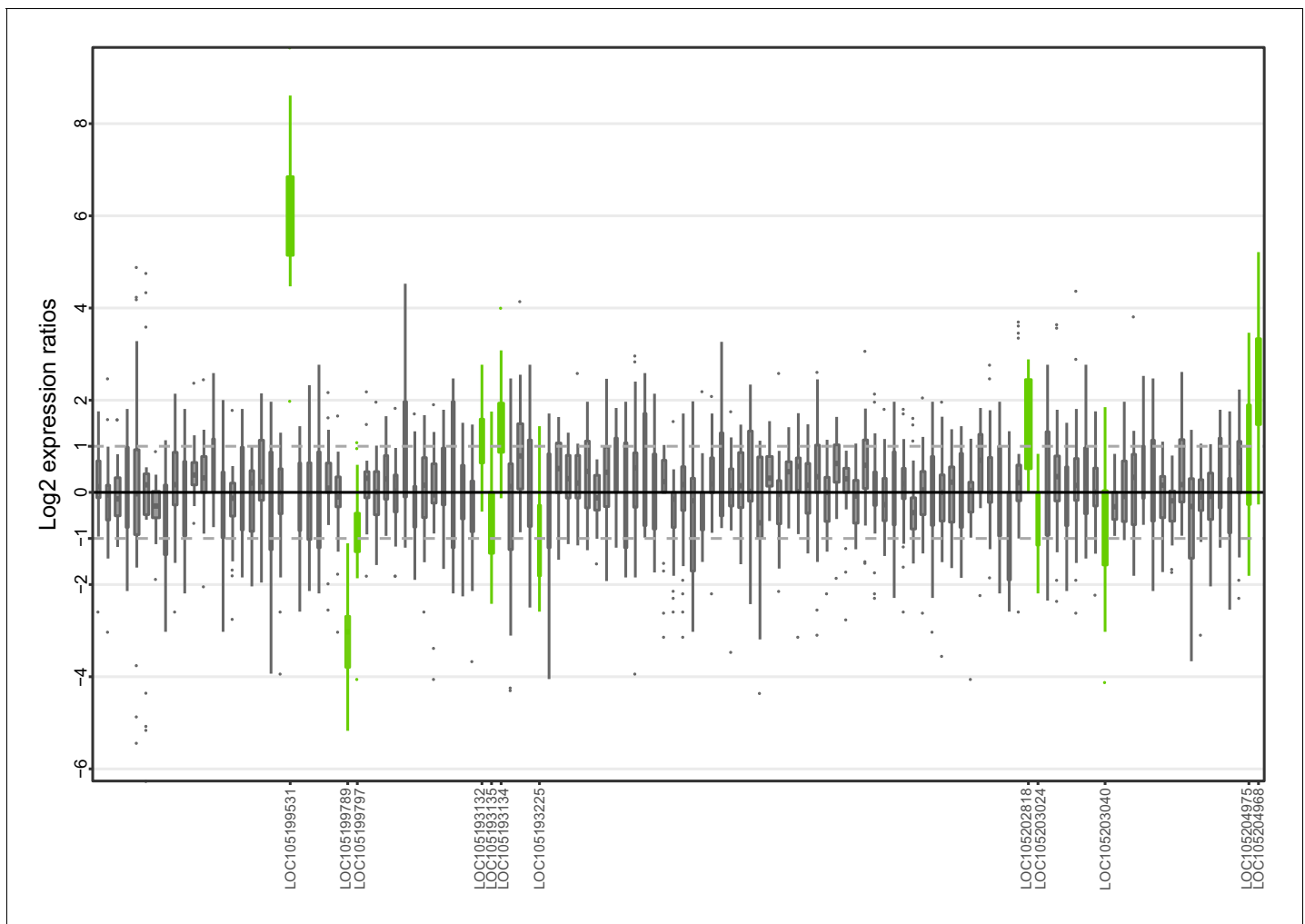


Figure 1—figure supplement 2. Allele-specific expression for genes in the fire ant social supergene for South American samples (information from body parts of queens and whole bodies of workers merged together). Differences in allele-specific expression between variants (y axis) for all genes in the supergene with enough expression information (x axis). Significant expression differences (BH adjusted $p < 0.05$ from Wald test in DESeq2) are in green and the RefSeq ID of the significant gene is outlined in the x axis. Non-significant differences are marked by grey boxes. Within each plot, each box shows the distribution of log2 expression ratios between SB and Sb. Positive values indicate higher expression in SB; negative values indicate higher expression in Sb. The dashed line shows log2 expression ratios = 1, that is a two-fold gene expression difference in either direction. Genes are in chromosomal order, showing that genes with allelic biases in either direction are distributed throughout the supergene, rather than localized in a manner that would be expected by strata of differentiation. Thicker boxes have higher median read counts than thinner boxes.

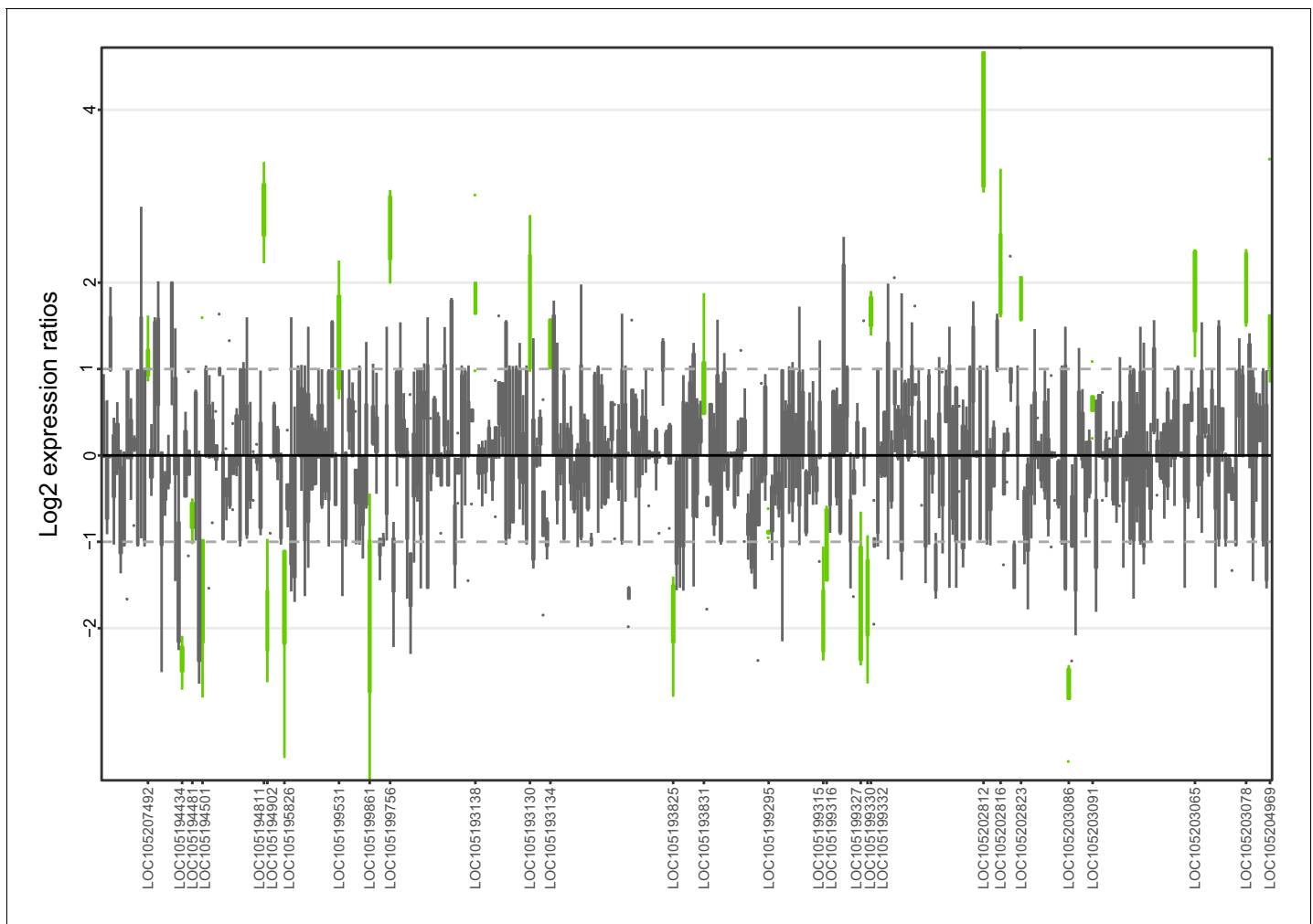


Figure 1—figure supplement 3. Allele-specific expression for genes in the fire ant social supergene for whole bodies of North American queens. Differences in allele-specific expression between variants (y axis) for all genes in the supergene with enough expression information (x axis). Significant expression differences (BH adjusted $p < 0.05$ from Wald test in DESeq2) are in green and the RefSeq ID of the significant gene is outlined in the x axis. Non-significant differences are in gray. Within each plot, each box shows the distribution of log2 expression ratios between SB and Sb. Positive values indicate higher expression in SB; negative values indicate higher expression in Sb. The dashed line shows log2 expression ratios = 1, that is a two-fold gene expression difference in either direction. Genes are in chromosomal order, showing that genes with allelic biases in either direction are distributed throughout the supergene, rather than localized in a manner that would be expected by strata of differentiation. Thicker boxes have higher median read counts than thinner boxes.

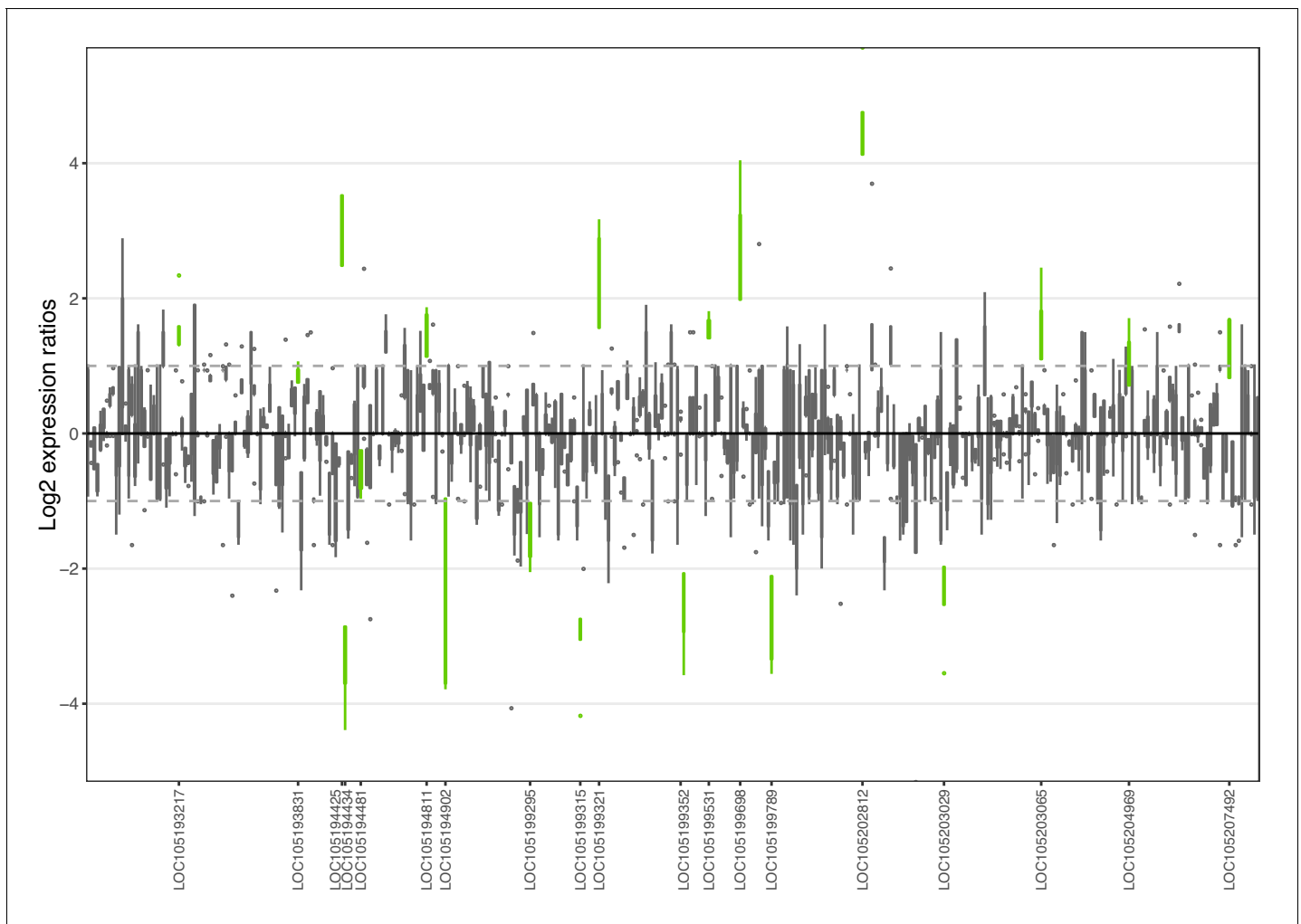


Figure 1—figure supplement 4. Allele-specific expression for genes in the fire ant social supergene) for whole bodies of Taiwanese queens. Differences in allele-specific expression between variants (y axis) for all genes in the supergene with enough expression information (x axis). Significant expression differences (BH adjusted $p < 0.05$ from Wald test in DESeq2) are in green and the RefSeq ID of the significant gene is outlined in the x axis. Non-significant differences are in gray. Within each plot, each box shows the distribution of log2 expression ratios between SB and Sb. Positive values indicate higher expression in SB; negative values indicate higher expression in Sb. The dashed line shows log2 expression ratios = 1, that is a two-fold gene expression difference in either direction. Genes are in chromosomal order, showing that genes with allelic biases in either direction are distributed throughout the supergene, rather than localized in a manner that would be expected by strata of differentiation. Thicker boxes have higher median read counts than thinner boxes.

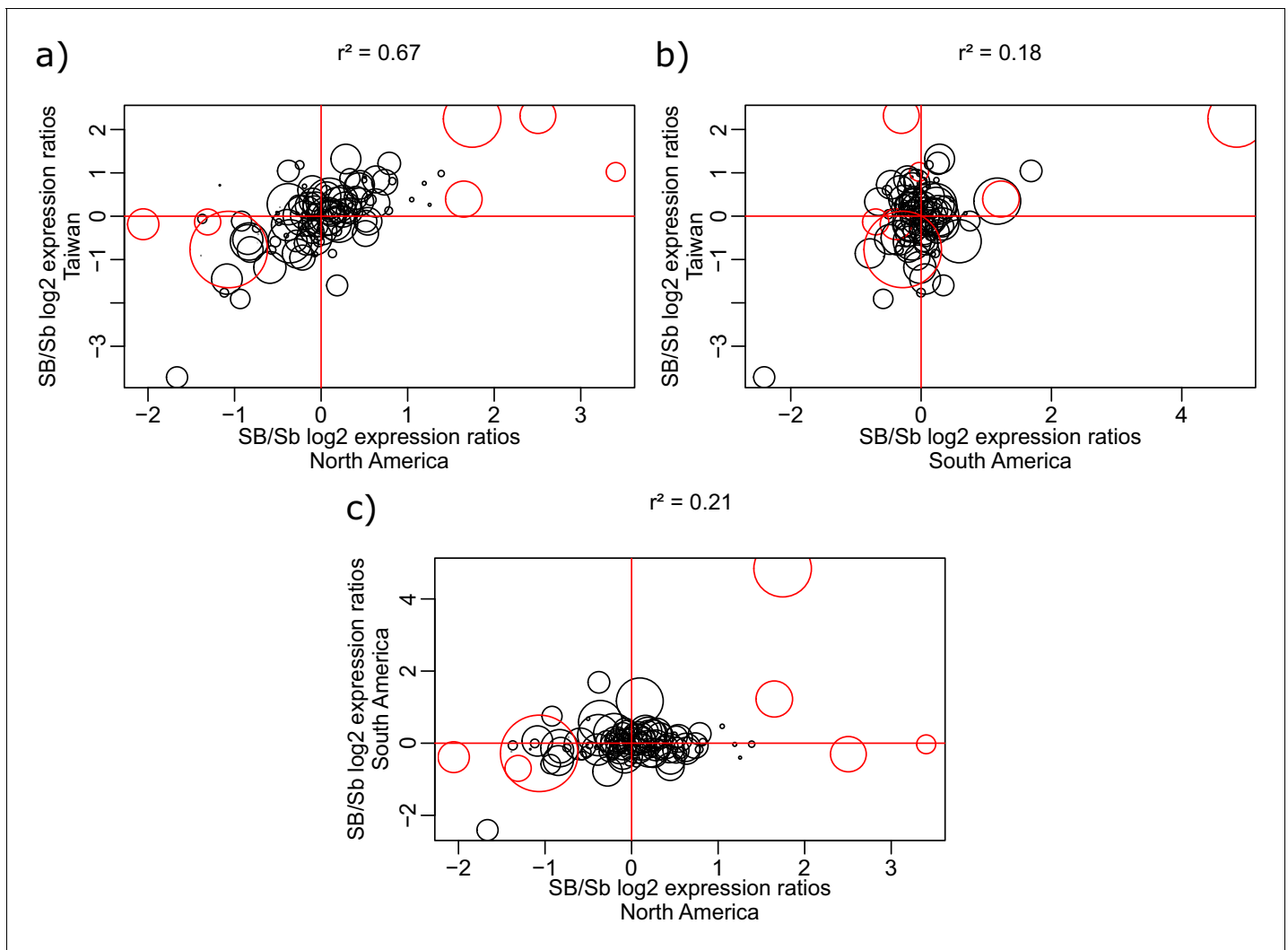


Figure 1—figure supplement 5. Correlation of log₂ allele-specific expression ratios between the SB and Sb variants in heterozygous queens from three populations: South American data we generated, North American data (from Morandin et al., 2016), and Taiwanese data (from Fontana et al., 2020). We show correlations: (a) between Taiwanese and North American populations; (b) between Taiwanese and South American populations; and (c) between South American and North American populations. Positive values represent higher expression of the SB allele, negative values represent higher expression of the Sb allele. Each dot represents a single gene within the supergene region. The sizes of the dots are proportional to the average gene expression level. Red dots represent the genes detected by the linear mixed effects model as significantly differentially expressed between SB and Sb across populations. The correlation r^2 between each pair of populations was calculated using the Spearman method, with each gene being weighted by mean expression level (read counts).

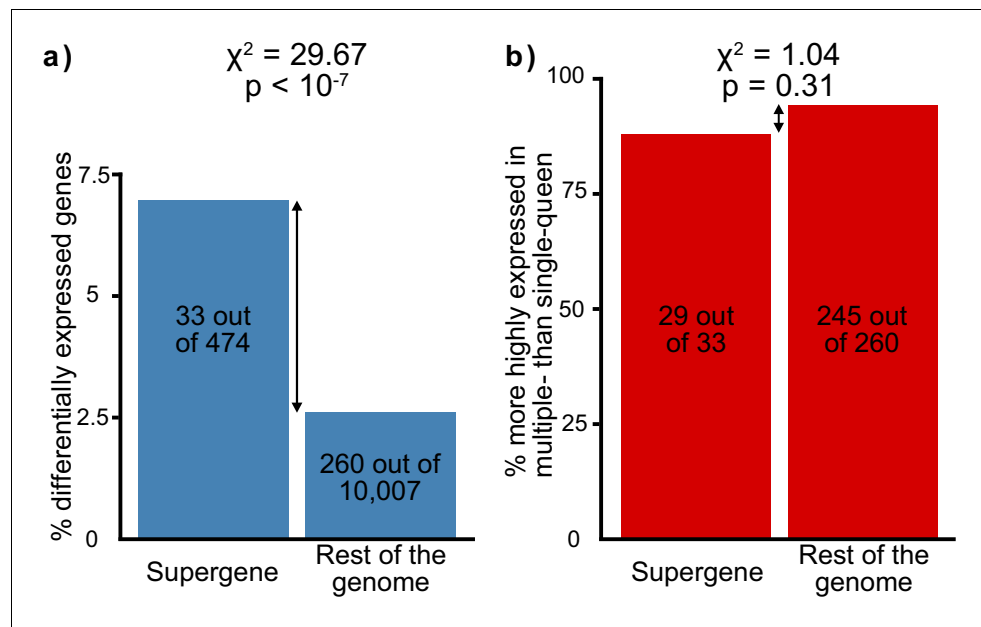


Figure 2. Distribution of socially biased genes in the genome of the red fire ant within (left bars) and outside (right bars) the supergene region. (a) The supergene region is significantly enriched in genes with differences between social forms, a pattern that could indicate the effect of antagonistic selection. (b) Most genes with differential expression between social forms are more highly expressed in multiple-queen colonies. This expression bias is observed across the genome and is not a unique feature of the supergene.

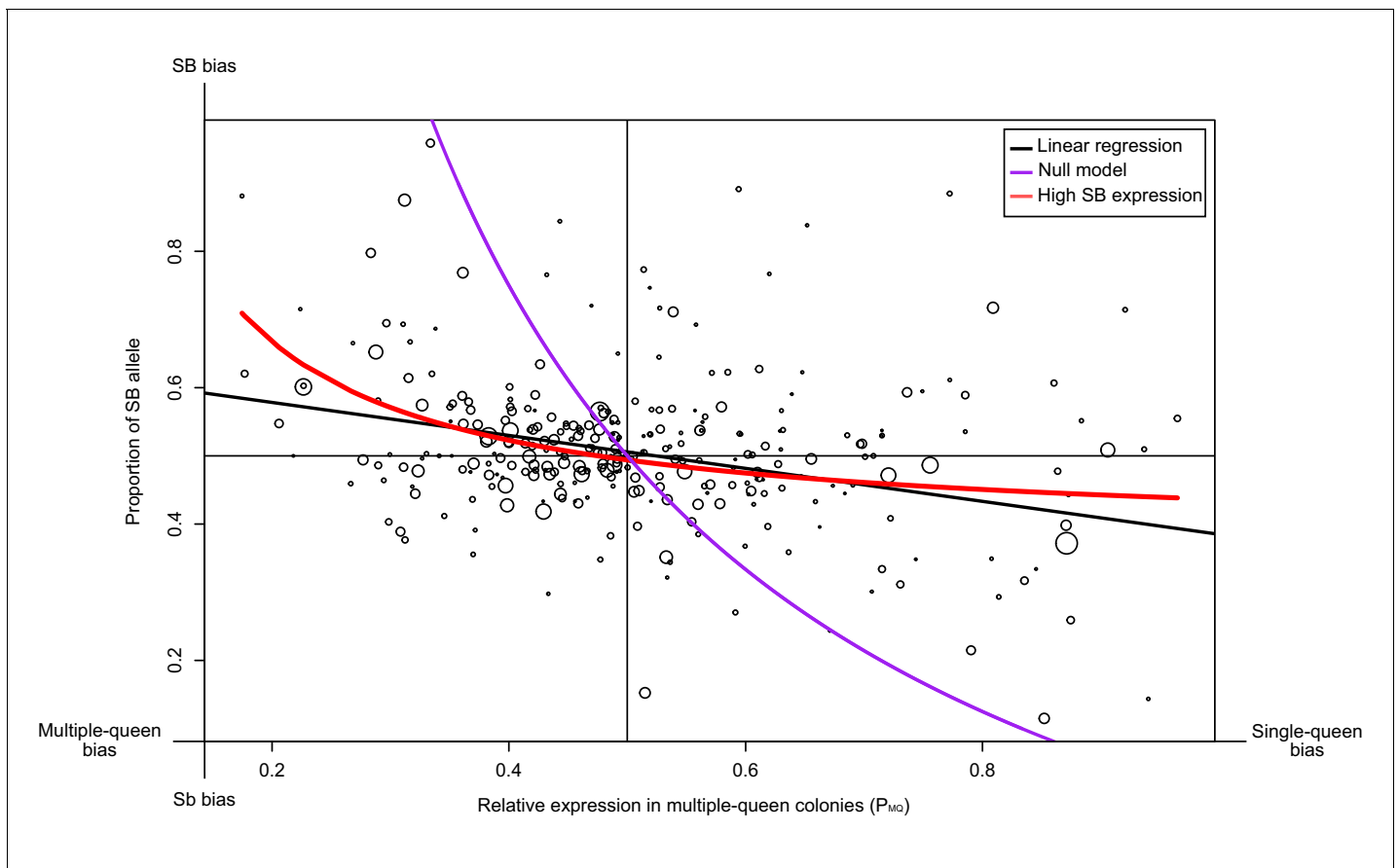


Figure 3. Relationship between measures of bias in allelic expression (P_B) and between social forms (P_{MQ}). Each point is one of 294 genes within the supergene (North American data). Point size is proportional to the mean expression in queens from multiple-queen colonies. The values were calculated as $P_B = x_B / (x_B + x_b)$ and $P_{MQ} = x_{MQ} / (x_{SQ} + x_{MQ})$; where x_B and x_b are the expression of SB and Sb alleles, and x_{SQ} and x_{MQ} are the expression in single-queen and multiple-queen colonies. Values of P_{MQ} below 0.5 therefore indicate higher expression in SB/SB queens from single-queen colonies; values above 0.5 indicate higher expression in SB/Sb queens from multiple-queen colonies. Values of P_B above 0.5 indicate allelic bias towards SB; values below 0.5 indicate bias towards Sb. The straight black line shows a linear regression. The purple line shows the predicted null relationship if the pattern of expression was due to Sb degeneration alone $P_{MQ} = (1 / (2P_B + 1))$. This model is a poor predictor of the data. The red line assumes gene-specific dosage compensation, where a decrease in expression of Sb leads to increased SB expression $P_B = (1 - (P_{MQ}/2)) / P_{MQ}$. The model including dosage compensation fits the data better than the null model; both models are significantly different (analysis of variance between models $p < 10^{-5}$). The observed enrichment of multiple-queen genes in Sb is therefore unlikely due to Sb degeneration alone.

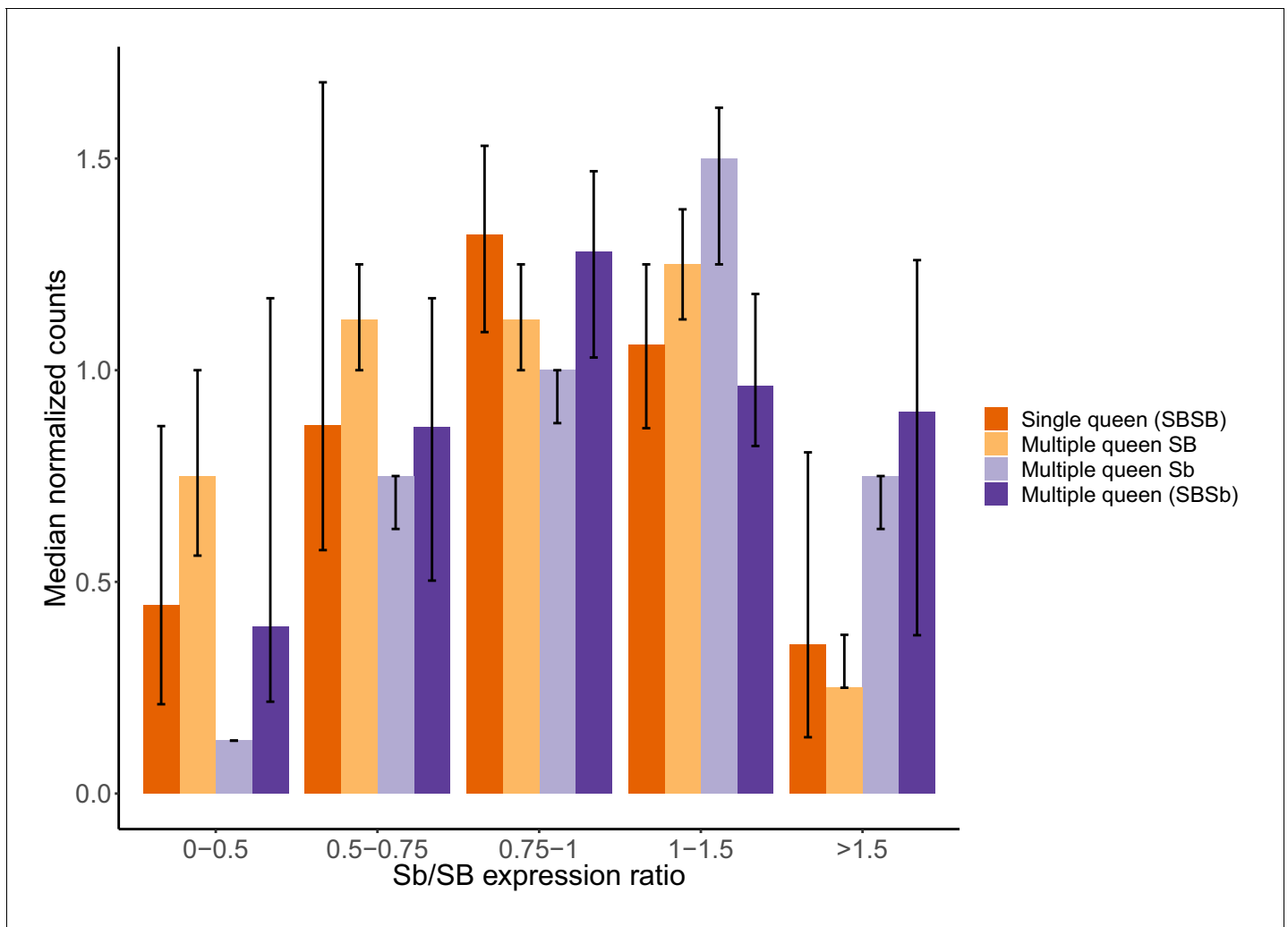


Figure 3—figure supplement 1. Gene expression levels in single-queen and multiple-queen individuals for different levels of SB and Sb expression levels in multiple-queens. The plot represents the overall expression levels for all genes analyzed in the supergene for which there was allele-specific and social form expression data in North American populations. To account for the potential effects of antagonistic selection, nine genes with significant expression biases towards Sb or with high SB expression in multiple-queen colonies were removed. As a result, 193 genes in total were included in this analysis. Each bar represents the median normalized expression within group, the error bars are the 95% CI interval around the median (estimated from a 5000x median bootstrapping). The expression levels within each expression group (Single-queens, multiple-queens, SB and Sb expression) is normalized by the total number of reads in that group. This normalization allows the comparison across different datasets with different levels. The differences in expression between single-queen and multiple-queen individuals remain non-significant (Wilcoxon test $p > 0.05$) across varying levels of SB-Sb expression differences. Only when Sb expression is much higher than SB expression (Sb/SB expression ratio > 1.5) does it seem to be an increase in gene expression for multiple-queen individuals, but the difference between social forms remained non-significant (Wilcoxon test $p = 0.08$). Without dosage compensation, we would expect that allelic bias would invariably result in changes in expression between social forms: As the allelic expression levels of SB increases, we would observe an increase in expression of single-queen colonies. Instead, expression levels in single-queen and multiple-queen individuals remain similar. These results are therefore consistent with SB expression compensating for low Sb expression.

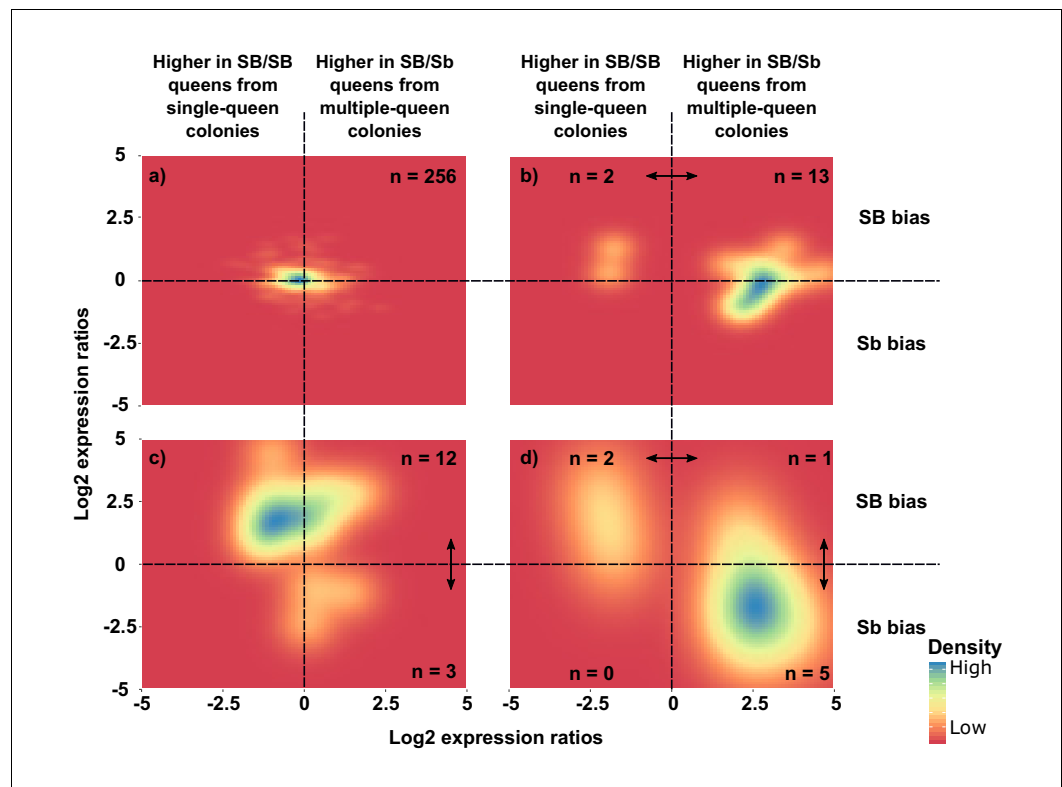


Figure 4. Distribution of differences in gene expression between social forms and between supergene alleles. X axes indicate ratios of expression between SB/Sb queens and SB/SB queens. Y axes indicate allelic expression ratios in SB/Sb queens. Both ratios use a log2 scale whereby log2 = 0 indicates absence of differences. Colors are proportional to numbers of genes. Double-headed arrows indicate significant expression differences. Panel (a) shows expression patterns for genes showing no difference in either comparison. The remaining three panels summarize expression patterns for: (b) genes with significant expression differences between SB/Sb and SB/SB queens only – these are biased towards higher expression in multiple-queen colonies (13 multiple-queen vs. two single-queen, binomial test, $p=0.007$); (c) genes with significant expression differences only between SB and Sb alleles within SB/Sb individuals – these are biased towards higher expression in the SB variant, in line with a dosage compensation mechanism (12 SB vs. 3 Sb, binomial test, $p=0.03$); (d) genes with significant expression differences between SB/Sb and SB/SB queens and between the SB and Sb variants in SB/Sb queens – the genes with higher expression of the Sb allele ($y < 0$) tend to be more highly expressed in queens from multiple-queen colonies ($x > 0$), in line with evolutionary antagonism between social forms (5 out of 8 Sb biased genes with bias towards multiple-queen colonies, compared with 1 out of 15 for SB biased genes $\chi^2 = 5.8$, $p=0.02$). The numbers in a) indicate how many genes had no differential expression. In b), (c) and d) the numbers in each quadrant indicate how many genes were significantly differentially expressed in the relevant comparison.

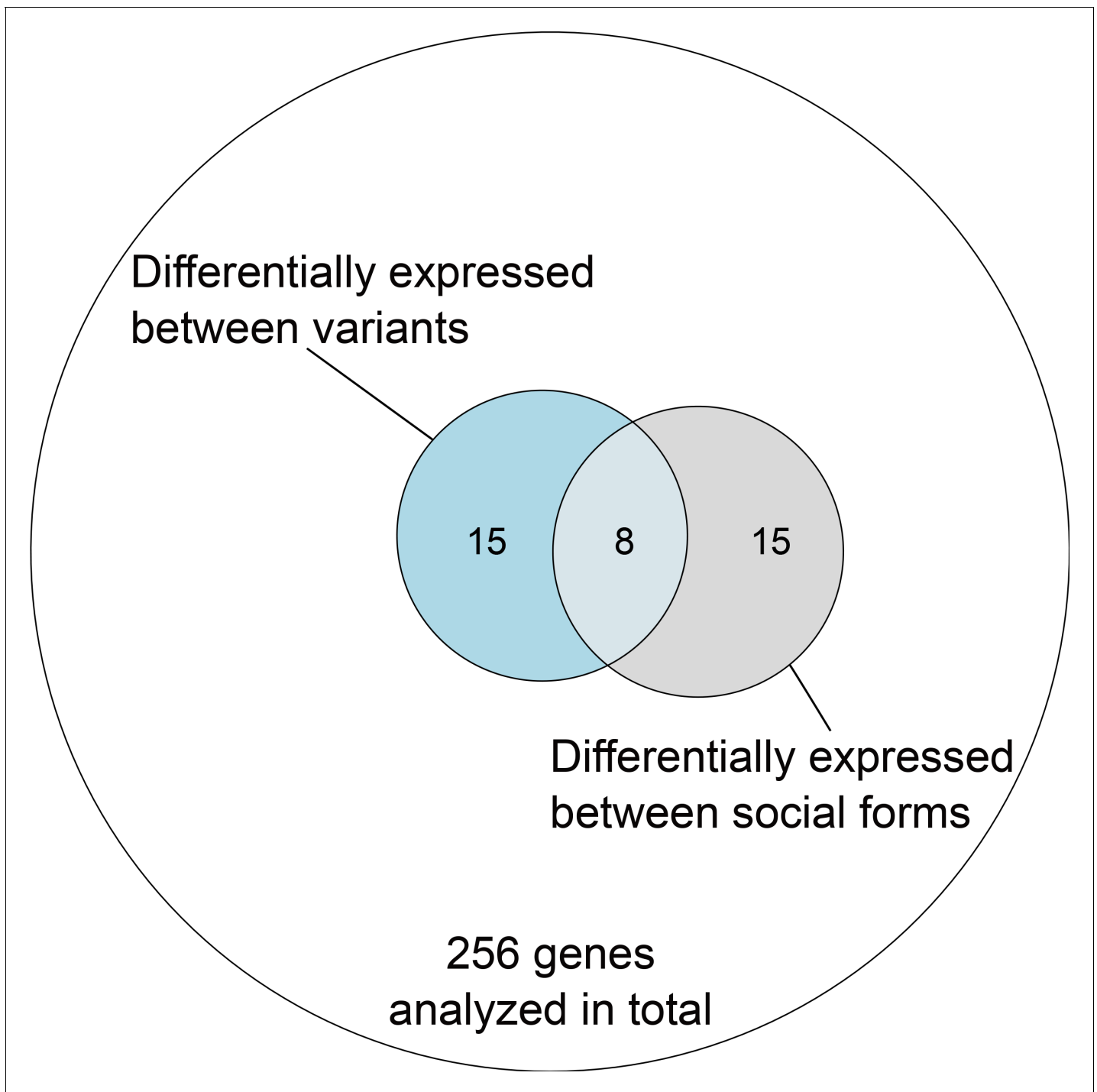


Figure 4—figure supplement 1. Overlap of genes with expression differences between variants and social forms out of all genes within the supergene region with sufficient data in both comparisons in the North American dataset.

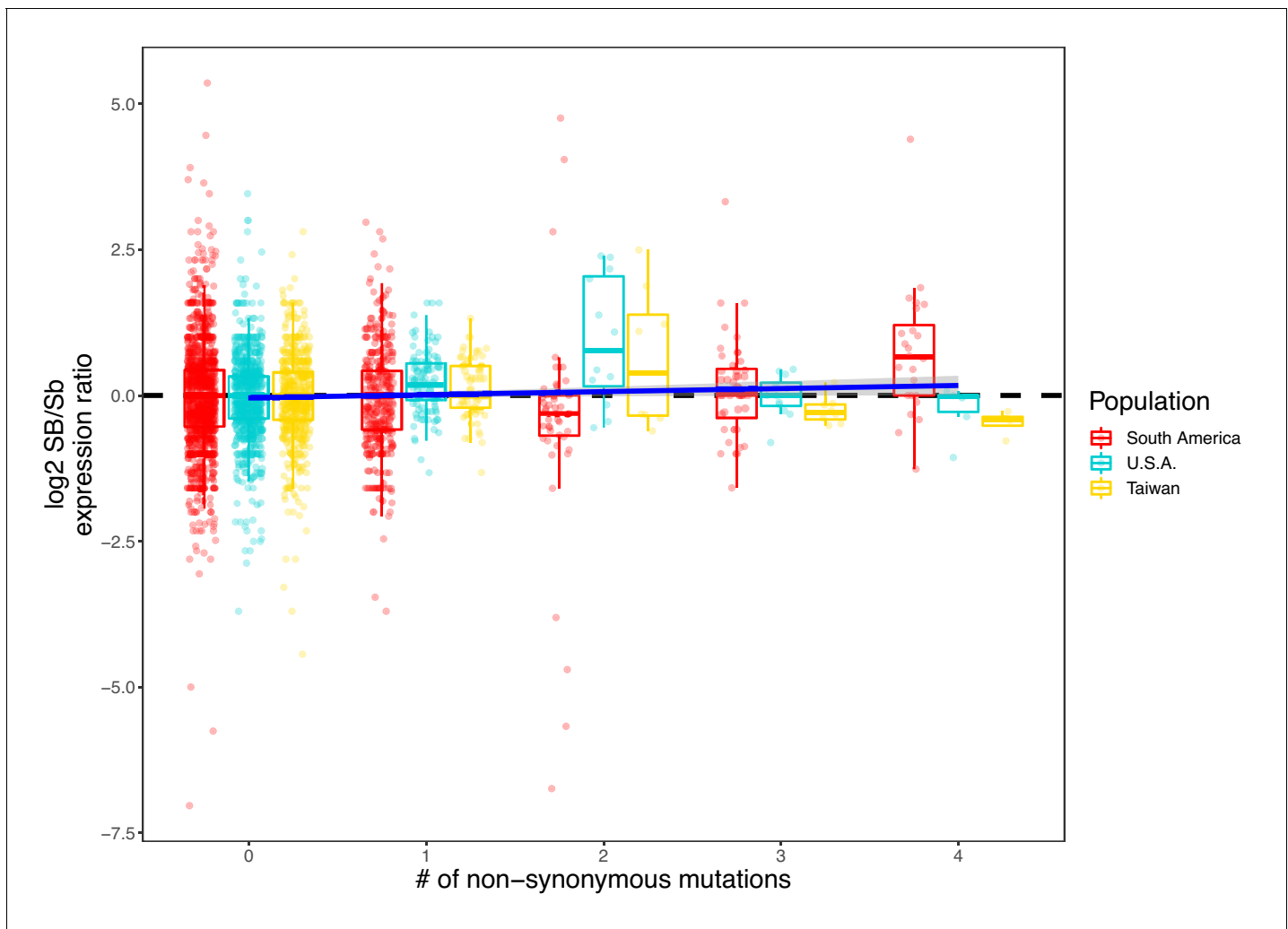


Figure 4—figure supplement 2. Allelic bias measured as the log2 ratio of expression between the SB and Sb alleles compared to the number of nonsynonymous mutations per gene. Positive values indicate higher expression in SB; negative values indicate higher expression in Sb. Allelic expression bias was measured in all populations analyzed in this study: South America, U.S.A. and Taiwan. Each dot represents a gene from a sample in each population. Because genes with low expression result in noisy log2 ratios of expression, we excluded genes with a median expression of 5 reads or fewer). A linear regression (dark blue line, ' $\log_2 \text{ratio} \sim \text{Population} + \text{Sample} + \text{Number of nonsynonymous mutations}$ ') shows that, overall, SB bias increases with the number of nonsynonymous mutations (coefficient = 0.052, $p=0.01$).