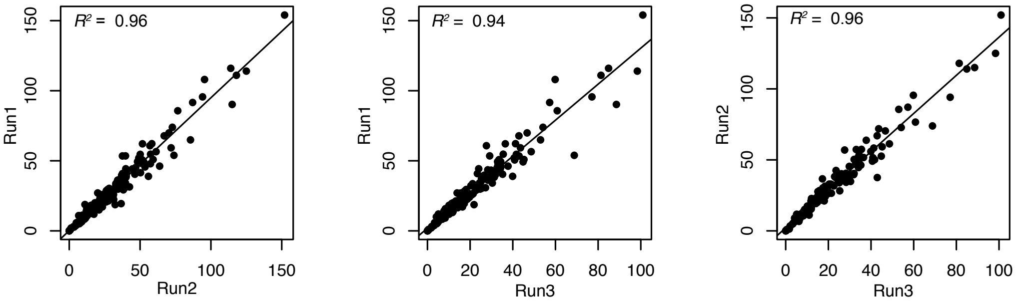

Figure S1: qPCR replicates were well-correlated. We plotted all pairwise comparisons for the three qPCR replicates and, following the data correction, found that replicates were well-correlated. R^2 is the coefficient of determination.

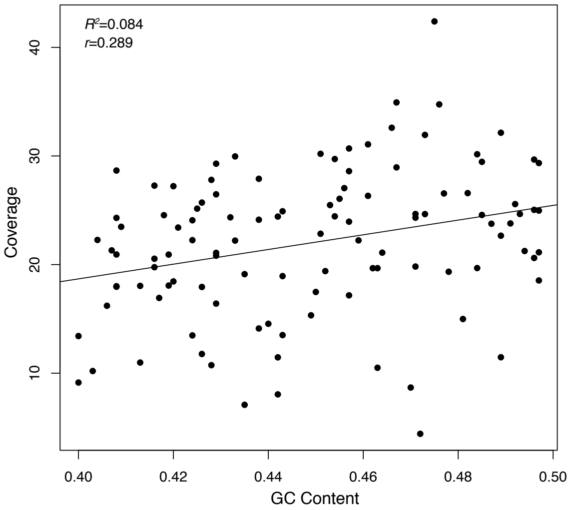

Figure S2: GC content did not significantly affect exon coverage. We plotted GC content versus coverage for the 109 nontoxin exons with a similar GC content to that of myotoxin exon 2 ($0.40 \leq \text{GC\%} \leq 0.50$) using the low-coverage, whole-genome reads and identified only a weak correlation. The solid line is the best-fit line. R^2 is the coefficient of determination, and r is Pearson's correlation coefficient.

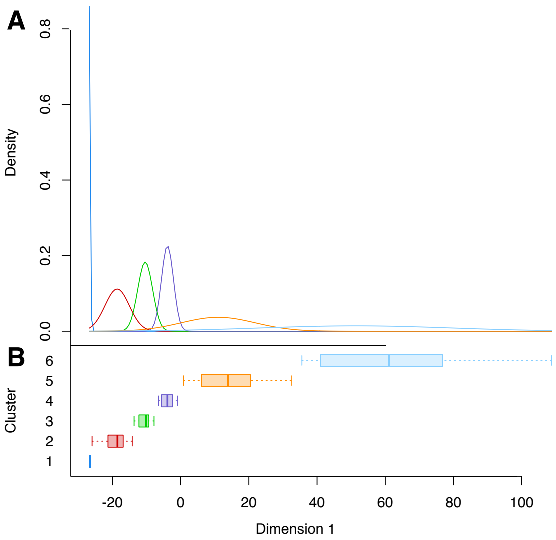

Figure S3: A dimension reduction plot visualizing the six clusters identified in the qPCR copy number data. We used the Mclust package in R and BIC criteria to fit a Gaussian mixture model to the copy number data and found that $k = 6$ was the best fit model. (A) A plot of mixture-based density estimates for the six clusters. (B) Box plots for each cluster. Whiskers extend to the most extreme data point.