Variability in Measurements of *Manduca sexta* Midgut Gene Expression

Kyle E. Davis and Bradley A. Hartlaub, Department of Mathematics, Kenyon College, Gambier, OH 43022

Abstract

- Using metabolic data that has been collected over the past several years from cohorts of *Manduca sexta*, we analyzed the variability that arose between cohorts, fourth and fifth instars, tissues, PCR plates, and animal weights. RNA was isolated from the anterior, middle, and posterior midgut and reverse transcribed to cDNA. Real-time PCR was used to quantify expression of four genes by the relative quantification method using *18s* ribosomal RNA as an internal control. The genes we looked at were the potassium amino acid transporter *KAAT1*, the aminopeptidase msAPN3 (*APN*), the cation chloride cotransporter *masBSC* and the e subunit of the V-type H-ATPase (*VATPase*).
- General linear models were fit and evaluated to predict ΔCt and the standard deviation for *KAAT1, APN, masBSC*, and *VATPase.*
- Gene expression changes between tissues and instars were calculated at for the genes.
- The performances of statistical procedures were compared with duplicate and triplicate measurements to obtain the desired power.

Data Collection

- The data is subdivided by: Cohort (a group of *M. sexta* grown and analyzed together), Plate (by date), Primer (*KAAT1*, *APN*, *VATPase, masBSC,* or *18s control*), Tissue (anterior midgut, posterior midgut, or middle midgut), Instar (4th or 5th), and Weight (in grams).
- For each sample, six Ct values are measured (three with *18s control* and three with one of the genes).
- After the performances of statistical procedures were compared with duplicate and triplicate measurements to obtain the desired power, it was determined that triplicate measurements were necessary.
- Ct is the number of cycles necessary for the amount of gene expressed to meet a predetermined threshold.
- Using these Ct values, we can compute ΔCt and ΔΔCt where $\Delta Ct = Avg(Ct_{Gene}) - Avg(Ct_{18s})$ and $\Delta\Delta Ct = \Delta Ct - \Delta Ct^{*}$.
- For ΔCt^{*}, the average of the ΔCt values for that particular gene were used.

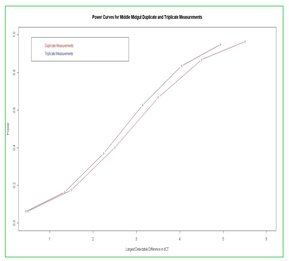

Figure 1: Power Curve for Middle Midgut Replicate Measurements

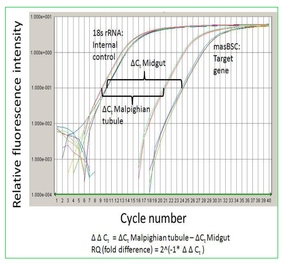

Figure 2: Ct and ΔCt Expression

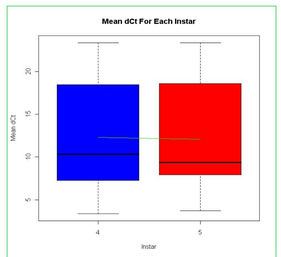

Figure 3: Mean ΔCT for Each Instar from Full Data Set

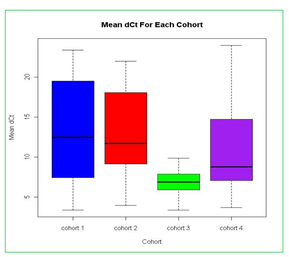

Figure 4: Mean ΔCT for Each Cohort from Full Data Set, experimental issues with Cohort 3

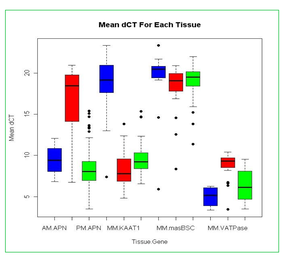

Figure 5: Mean ΔCT for Each Tissue from Full Data Set

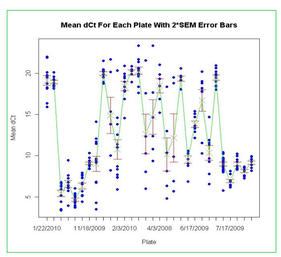

Figure 6: Mean ΔCT for Each Plate from Full Data Set

Descriptive Statistics

- Descriptive statistics from the entire data set show overall trends.
- For *KAAT1,* gene expression was more than 1500-fold higher in middle midgut and more than 300 fold higher in posterior midgut compared to anterior midgut. No significant differences among midgut regions were observed for *APN*, *masBSC*, or *VATPase*.
- Expression between Instars varied by gene.

ΔCt modeling

- ΔCt models were fit and evaluated using the overall predictor variables: Cohort, Plate, Gene, Tissue, Instar, and Weight. Interaction terms for all possible combinations of the predictor variables were also included. From the fully saturated model, stepwise analysis was used to eliminate variables that did not significantly affect the prediction of ΔCt. The best overall models for each gene contain the significant explanatory variables.
- The overall model for *KAAT1* is: ΔCt = Instar + Tissue + Weight + Plate + Instar:Plate; Adjusted R-Squared = 0.7270
- The overall model for *APN* is: ΔCt = Plate + Weight; Adjusted R-Squared = 0.8697
- The overall model for *masBSC* is: ΔCt = Instar + Tissue + Plate + Weight + Instar:Plate + Instar:Weight; Adjusted R-Squared = 0.7629
- The overall model for *VATPase* is: ΔCt = Plate; Adjusted R-Squared = 0.5075

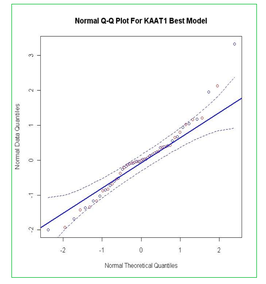

Figure 9: Normal Quantile-Quantile Plot for ΔCt of *KAAT1*

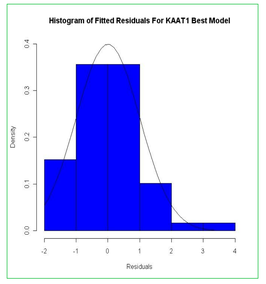

Figure 10: Histogram of Fitted Residuals for ΔCt of *KAAT1*

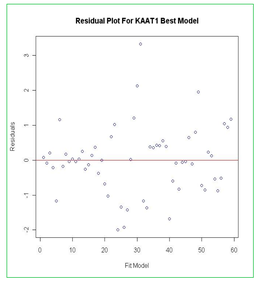

Figure 11: Residual Plot for ΔCt of *KAAT1*

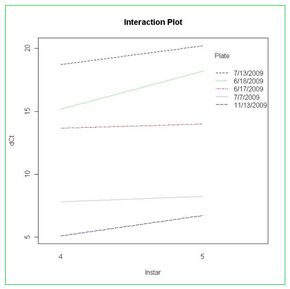

Figure 7: Interaction Plot for ΔCt of *KAAT1*

Source	DF	Type III SS	Mean Square	F Statistic	P-value
Instar	1	29.173	29.173	19.9565	< 0.001
Tissue	1	42.650	42.650	29.1760	< 0.001
Weight	1	19.231	19.231	13.1555	< 0.001
Plate	9	108.545	12.061	8.2504	< 0.001
Instar:Plate	9	56.853	6.317	4.3214	< 0.001
Residuals	37	54.087	1.462		

Table 1: ANOVA table for ΔCt of *KAAT1*

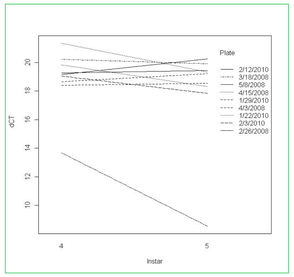

Figure 8: Interaction Plot for ΔCt of *masBSC*

Source	DF	Type III SS	Mean Square	F Statistic	P-value
Plate	3	773.25	257.51	82.2980	< 0.001
Weight	1	12.80	12.802	4.0876	0.0514
Residuals	33	103.35	3.132		

Table 2: ANOVA table for ΔCt of *APN*

SD modeling

- Models for the standard deviation (SD) of the Ct measurements were fit and evaluated using the same overall predictor variables as above. Interaction terms for all possible combinations of the predictor variables were also included. From the fully saturated model, stepwise analysis was used to eliminate variables that did not significantly affect the prediction of SD. The best overall models for each gene contain the significant explanatory variables.
- The overall model for *KAAT1* is: SD = Plate; Adjusted R-Squared = 0.1739
- The overall model for *APN* is: SD = Instar + Plate; Adjusted R-Squared = 0.0731
- The overall model for *masBSC* is: ΔCt = Plate + Instar + Weight + Instar:Weight; Adjusted R-Squared = 0.0506
- The overall model for *VATPase* is: SD = Instar + Plate; Adjusted R-Squared = 0.0869

Source	DF	Type III SS	Mean Square	F Statistic	P-value
Plate	9	0.2088	0.0232	2.3563	0.0268
Residuals	49	0.4824	0.0098		

Table 3: ANOVA table for SD of *KAAT1*

RQ Error Bars

- The PCR software calculates the standard deviation for the average ΔCt by using $\sqrt{s_1^2 + s_2^2}$, where s_1 estimates the standard deviation for the for the gene measurements and s_2 approximates the standard deviation for the *18s control* measurements. However, this does not take into account the triplicate measurements for the gene and control.
- Using standard error (se) instead incorporates the number of replicate measurements. Thus, it is better to use $\sqrt{\frac{s_1^2}{3} + \frac{s_2^2}{3}}$.
- So, for fold difference in expression (RQ), error bars were calculated based on the range of RQs calculated from the ΔΔCt values ± standard error.

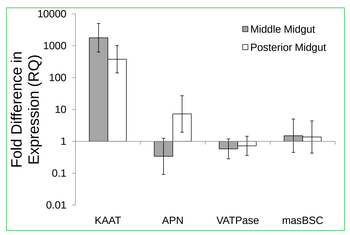

Figure 12: RQ in comparison to anterior midgut for each gene with SE error bars

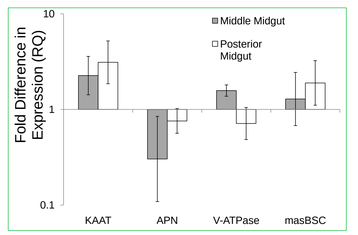

Figure 13: RQ of 5th instar larvae in comparison to 4th instars for each gene with SE error bars

Conclusions and Future Research

- Triplicate measurements were shown to have higher power than duplicate measurements, but the powers may be close enough to move to duplicate measurements. More research needs to be done with a focus on replicate measurements to see if the time, materials, and money saved by doing duplicate measurements are worth a drop in power.
- The best models for predicting ΔCt for *KAAT1*, *APN*, *masBSC*, and *VATPase* provided fits with Adjusted R-Squared values all above 0.5 and only one below 0.7.
- The best models for predicting SD for *KAAT1*, *APN*, *masBSC*, and *VATPase* provided fits with Adjusted R-Squared values all below 0.2. One of the major assumptions in fitting general linear models is that the variance in the number of cycles to hit a threshold is constant. Our models indicate that the variability in Ct measurements is heterogeneous (i.e. not all of the standard deviations are equal).
- Future research should look at minimizing the variability that was detected. Also, future work should look at more design changes to increase randomization of treatment assignments.

Acknowledgements

I would like to thank Brad Hartlaub for his insight and guidance throughout this project. I would also like to thank Chris Gillen for his assistance, the NSF (DMS - #0827208), and the Kenyon Summer Science Scholars Program.