Gene Expression in Fourth and Fifth Instar *Manduca Sexta*

Robert A. Long and Bradley A. Hartlaub, Department of Mathematics, Kenyon College, Gambier, OH 43022

Abstract

- The scaling exponent for metabolic rate in *Manduca sexta* is greater than expected based on isometric surface area scaling. We examined whether midgut genes involved in absorption and digestion are expressed differently in fifth versus fourth instar larvae. RNA was isolated from middle and posterior midgut and reverse transcribed to cDNA. Real-time PCR was used to quantify expression of two genes, aminopeptidase N (*msAPN3*) and potassium amino acid cotransporter (*KAAT*), by the relative quantification method using 18s ribosomal RNA as an internal control. A general linear model for ΔCt values from *APN* and *KAAT* identified significant instar and tissue effects.
- The best model for *KAAT1* and *KAAT2* provided a good fit, with an R-squared value of 0.888431. In posterior midgut, *KAAT* expression was 2.5-3.5 fold higher while *APN* was 1.3-1.4 fold lower in fifth compared to fourth instar larvae. Expression of *KAAT* was 3.5 fold higher in middle compared to posterior midgut.
- To evaluate the sensitivity of standard statistical methods to real time PCR data, a Monte Carlo simulation of the coverage probabilities for Ct, ΔCt and ΔΔCt was conducted for the normal, uniform, exponential, t and chi-square distributions using parameter estimates derived from our data. Coverage probabilities for Ct and ΔCt were close to the 95% confidence level, indicating that the interval estimates are relatively insensitive to distributional assumptions. A notable exception was the exponential distribution of Ct, where the coverage probability dropped to 0.88. Coverage probabilities for ΔΔCt were reasonable when using the mean of a comparison group, but not when arbitrarily selecting an individual measurement for comparison.

Data Collection

Data Collection:

- The data can be subdivided by: Cohort (a group of *Manduca Sexta* grown and analyzed together), Plate, Primer (*APN*, *KAAT* or *18s*), Tissue (MM or PM), and Instar (4^{th} or 5^{th}).
  - For each sample, six Ct values are measured (three with *18s* and three with *APN* or *KAAT*).
    - Replicates of three are a necessity in order to have a reasonable power (see Figure 1).
  - Ct is the number of cycles necessary for the amount of gene expressed to meet a predetermined threshold (see Figure 2).
    - Using these Ct values, we can compute ΔCt and ΔΔCt where $\Delta Ct = Ct_{KAAT} - Ct_{18s}$ and $\Delta\Delta Ct = \Delta Ct - \Delta Ct^*$.
      - For ΔΔCt there does not exist a clear treatment-control relationship within the data. Computer software will tend to use an arbitrary observation as the control (ΔCt^*); we have found that there are problems with this technique.

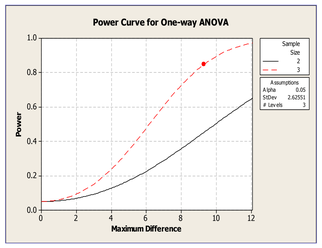

Figure 1: Power Curve for Replicates

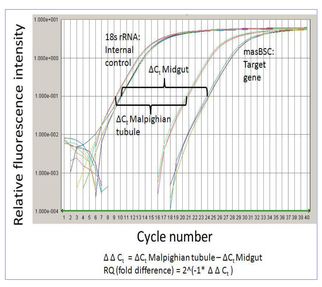

Figure 2: Ct and ΔCt Expression

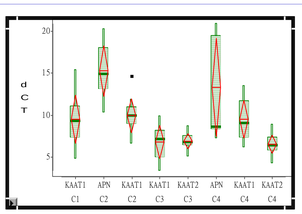

Figure 3: ΔCt vs. Cohort, Primer

Statistical Inference Results

- Anderson-Darling: H_0: Data (ΔCt) are normal, H_a: Data are not Normal
  - Test statistic: A-sq = 7.5525, P-value: 0.0050
  - However, when the data are separated by primer, tissue and cohort many of the subgroups confirm normality.
- T-Tests show significant differences between most cohorts (see Figures 3, 8).
- H_0: $\mu_i = \mu_j$ vs. H_a: $\mu_i \neq \mu_j$
  (where μ_i is the mean ΔCt for $cohort_i$)
- The F-test for equal variance shows significant differences between primers.
  - H_0: $\sigma_i^2 = \sigma_j^2$, H_a: $\sigma_i^2 \neq \sigma_j^2$
    where σ_i^2 is the variance of ΔCt for $primer_i$.

Cohort	N	Mean	SD	Min	Median	Max
1	23	9.48	2.87	4.79	9.34	15.37
2	32	12.59	3.71	6.54	11.70	20.22
3	54	6.82	1.65	3.38	6.88	9.83
4	73	9.87	4.36	4.42	8.36	20.96

Primer	N	Mean	SD	Min	Median	Max
KAAT1	104	8.70	2.62	3.38	8.38	15.37
KAAT2	40	6.72	0.99	4.24	6.62	8.84
APN	38	14.12	4.90	7.25	14.17	20.96

$Cohort_i$	$Cohort_j$	DF	t-value	P-Value
Cohort 1	Cohort 2	52.64	31.88	0.0009
Cohort 1	Cohort 3	28.37	4.15	0.0003
Cohort 1	Cohort 4	56.38	-0.50	0.6167
Cohort 2	Cohort 3	38.40	8.33	<.0001
Cohort 2	Cohort 4	69.05	3.27	0.0017
Cohort 3	Cohort 4	97.62	-5.48	<.0001

$Primer_i$	$Primer_j$	Num. DF	Den. DF	F-Value	P-Value
KAAT1	KAAT2	103	39	7.03	<.0001
APN	KAAT1	37	103	3.50	<.0001
APN	KAAT2	37	39	24.58	<.0001

General Linear Model

- $\Delta Ct_i = \beta_0 + \beta_1 * Plate_i + \beta_2 * Tissue_i + \beta_3 * Primer_i + \beta_4 * Instar_i + \beta_5 * Instar_i * Plate_i + \varepsilon_i$
- Instar and the interaction between plate and instar were found to be significant (see figure 9). As expected, Plate, Tissue (MM, PM) and Primer (*KAAT1*, *KAAT2*) were also found to be significant.
- $R^2 = 0.888431$

Source	DF	Type III SS	Mean Square	F Statistic	P-Value
Plate	17	489.888	28.81694	31.88	<.0001
Tissue	1	21.454	21.48505	23.77	<.0001
Primer	1	15.065	15.06456	16.66	<.0001
Instar	1	20.018	20.01758	22.14	<.0001
Instar*Plate	17	95.289	5.60523	6.20	<.0001

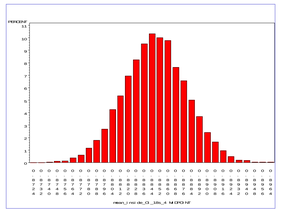

Figure 4: Ct for Exponential

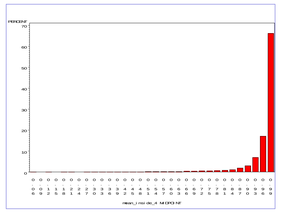

Figure 5: ΔΔCt for Exponential with Random

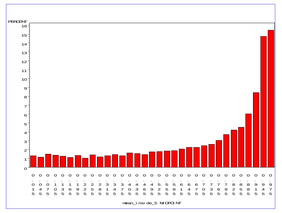

Figure 6: ΔΔCt for t with Mean

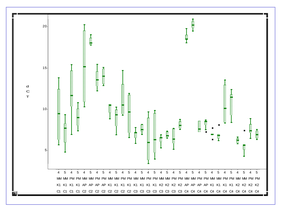

Figure 7: ΔCt vs. Cohort, Primer, Tissue and Instar

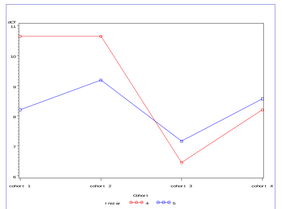

Figure 8: Profile Plot of ΔCt vs Cohort (KAAT1, KAAT2)

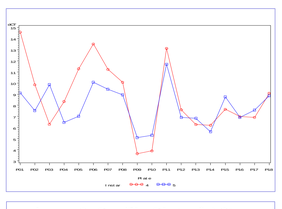

Figure 9: Profile Plot of ΔCt vs Plate by Instar (KAAT1, KAAT2)

Monte Carlo Simulation

Methods:

- Coverage probabilities were simulated for Ct, ΔCt and ΔΔCt for 4^{th} and 5^{th} instars.
- Ct values were randomly generated as either Normal, Uniform, Exponential, t or Chi-square using parameter estimates derived from our data.
- We then constructed a confidence interval and checked whether the mean was within the bounds.
  - This procedure was repeated 10,000 times to estimate the coverage probability.
  - Then 10,000 coverage probabilities were calculated and we created a histogram to show the distribution of the coverage probabilities.

Results:

- Ct was found to be relatively insensitive to changes in distribution.
  - Notable exceptions occurred with the uniform (coverage dropped to around 0.92) and the exponential (around 0.885)
- ΔCt saw a general increase in coverage (normal = 0.957)
  - Exponential improved to 0.955 and uniform to 0.954
- Three different controls were used for computing ΔΔCt:
  1. An arbitrary ΔCt value
     - The coverage, though often centered around 0.95, had an extremely large range (from 0.04 to 1).
  2. The mean of the 4^{th} instar ΔCt values
     - This method showed excellent coverage, though slightly below 0.95, the range was small and the coverage estimates were normally distributed. This method was unsuccessful for the t-distribution because of the extreme outliers.
  3. The median of the 4^{th} instar ΔCt values
     - Unaffected by the outliers in the t-distribution.

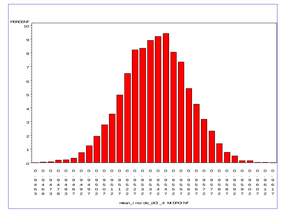

Figure 10: ΔCt for Exponential

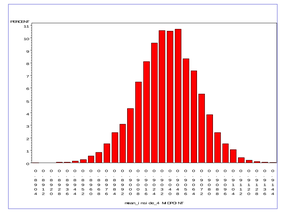

Figure 11: ΔΔCt for Exponential with Mean

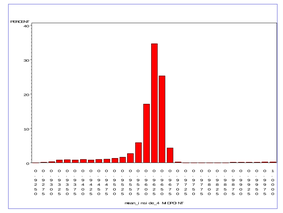

Figure 12: ΔΔCt for t with Median

Conclusions

- The general linear model showed significant differences in the expression of middle and posterior midgut genes in 4th and 5th instar *Manduca Sexta*.
- As expected there were also differences in the expression based on tissue, plate and primer.
- The Monte Carlo simulation confirmed ΔCt as a more reliable measurement than Ct.
  - Ct and ΔCt were both relatively insensitive to distributional changes though key exceptions were in the uniform and exponential distribution
  - Of the three ways to compute ΔΔCt, choosing a random observation as the control was shown to be unreliable.
  - The mean worked well as long as we were not dealing with the t-distribution where we needed to use the median.
  - Future research should consider nonparametric alternatives because of the non-normality of our data. The Monte Carlo simulation could also be extended to simulate fold changes (defined as $2^{-\Delta\Delta Ct}$).

References

1. Yuan, Joshua S., Ann Reed, Feng Chen and C. Neal Stewart Jr. "Statistical Analysis of Real-Time PCR Data." *BMC Bioinformatics* 7:85 (22 February 2006):
2. Gibson, Ursula E.M., Christian A. Heid, and P. Mickey Williams. "A Novel Method for Real Time Quantitative RT-PCR." *Genome Research* 6 (16 August 1996): 995-1001.
3. Karlen, Yann, Alan McNair, Sebastien Perseguers, Christian Mazza and Nicolas Mermod. "Statistical Significance of Quantitative PCR." *BMC Bioinformatics* 8:131 (20 April 2007).
4. Livak, Kenneth J. and Thomas D. Schmittgen. "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta Ct}$ Method." *Methods* 25 (2001): 402-408.

Acknowledgements

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