



This is to certify that the  
dissertation entitled

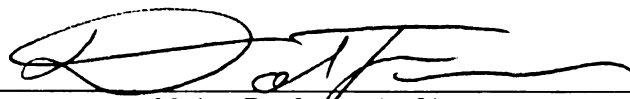
LUCILIA SERICATA DEVELOPMENT: PLASTICITY,  
POPULATION DIFFERENCES, AND GENE EXPRESSION

presented by

Aaron Michael Tarone

has been accepted towards fulfillment  
of the requirements for the

Ph.D. degree in Zoology



Major Professor's Signature

5/9/07

Date

LIBRARY  
Michigan State  
University

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.  
**MAY BE RECALLED** with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

**LUCILIA SERICATA DEVELOPMENT: PLASTICITY, POPULATION  
DIFFERENCES, AND GENE EXPRESSION**

By

Aaron Michael Tarone

**A DISSERTATION**

Submitted to  
Michigan State University  
In partial fulfillment of the requirements  
For the degree of

**DOCTOR OF PHILOSOPHY**

Department of Zoology

2007



## ABSTRACT

*Lucilia sericata* Development: Plasticity, Population Differences, and Gene Expression

By

Aaron Michael Tarone

Blow flies (Diptera: Calliphoridae) are parasitoids, parasites, and primary successional species on carrion. A detailed understanding of their development can impact medical, veterinary, and forensic sciences as they spread disease, debride wounds, spoil food, parasitize humans and livestock, and predictably develop on carrion. Comparisons of blow flies to *Drosophila melanogaster* and other higher flies can also reveal important evolutionary effects on a locus or trait of interest. The research detailed herein was based on the application of quantitative and molecular genetic methodology to questions regarding the development of the globally distributed blow fly *Lucilia sericata* (Diptera: Calliphoridae) (Meigen).

The main focus of this research was to improve the precision of postmortem interval (PMI) estimates produced with entomological evidence, based on the inclusion of developmentally variable gene expression data into the process of predicting blow fly age. The expression profiles of 9 genes, in conjunction with generalized additive models (GAMs), was useful in more precisely predicting the ages of immature *L. sericata*. Some genes were also useful in identifying individuals in a state of arrested development, which will further decrease error in entomology based PMI estimates.

The principles of quantitative genetics were also applied to the problem of understanding developmental variation in *L. sericata*. Complimentary studies were

undertaken to understand how the environment and genetics affect body size and development time. Components of developmental plasticity were identified, providing an improved understanding of how laboratory rearing conditions affect blow fly development time and enabling the identification of a rearing protocol that mimics growth on carrion. Likewise, the effects of genotype and temperature on variation in the development time and body sizes of *L. sericata* from California, Michigan, and West Virginia were determined, revealing significant temperature and strain effects on these phenotypes.

The results of these projects indicate that molecular and quantitative genetics are useful tools in understanding blow fly development, especially for predicting a PMI. The approach used herein demonstrated a quantifiably improved ability to predict blow fly age, addressing a key shortcoming in forensic entomology as interpreted by the Daubert standard of scientific evidence. However, the data obtained will also be useful in devising methodologies for controlling blow fly populations and in understanding the evolution of development. Several of the genes studied interact with insecticides; consequently, profiles of the expression levels of these genes can be useful for controlling blow fly pests. Similarly, knowledge about the factors affecting developmental plasticity may be used to better manage facilities affected by blow flies. Likewise, the high-resolution gene expression data produced in this research will aid in understanding the evolution of development in the Diptera. By studying gene expression, and other traits, throughout the development of the blow fly *Lucilia sericata*, valuable knowledge affecting a variety of biological studies has been acquired.

This dissertation is dedicated to my mother Cindy Tarone, who died without the opportunity to watch her son grow up, and to my father Lawrence Tarone, who made sure that I did.

## ACKNOWLEDGEMENTS

I greatly appreciated the advice received from all of my committee members: David Foran, Rich Merritt, David Arnosti, Will Kopachik, and Ned Walker. I would especially like to thank my advisor David Foran for the willingness to support a highly interdisciplinary project in his laboratory and Rich Merritt for his role in my professional development as a forensic entomologist. The proofreading skills of Missy Foran were also appreciated.

Thanks to the members of the Foran, Merritt, and Walker laboratories and to the Forensic Science graduate students. I am grateful for their friendship, companionship during coffee runs, assistance, and tolerance of my research organism (especially the unappealing odor associated with it). Thanks to Kim Jennings and Erin Lenz for help with the genetic identification of fly strains. Thanks to Jeff Dake, Dave and Anne Szymanski, Michiel Kramer, Lisa Ramos, Stephanie Kremer, Angela Donatelle, Mike Gehring, Cory Michaud, Ken Eilert, Christina Rauzi, and Shane “Sunshine” Hoffmann for minimizing the loss of sanity associated with the process of researching and writing a doctoral dissertation. Thanks to Ryan Kimbirauskas, Mollie McIntosh, and Eric Benbow for their assistance, advice, beers, and for their hospitality in the Merritt lab.

Various Michigan State University staff assisted the implementation of this research. Thanks to the staff of the Michigan State University Research Technology Support Facility, especially Annette Thelen, Jeff Landgraff, and Shari Tjigum-Holland, as they were instrumental in assisting with experimental design, the use of robotics, RNA isolation, and quantitative PCR. Randall Shoemaker and Constance Crew from Michigan State University Laboratory Animal Resources provided ethics advice and the rats for this

research. Brian McGill provided valuable insight into the world of multivariate statistical analyses. Conversations regarding variation in fly development with Alex Shingleton were appreciated, as was the microscope time. The validation experiment was possible due to the excellent effort of Trevor McLean during his summer internship at Michigan State University.

Many forensic entomologists provided useful insight and/or helpful assistance regarding experimental design, data analysis, and species identification. These include Ryan Kimbiraskaus at Michigan State University, Jeff Tomberlin at Texas A&M, Jeff Wells at West Virginia University, Bob Kimsey at the University of California at Davis, Neal Haskell at Saint Joseph's University, and Gail Anderson at Simon Fraser University.

To my father and grandparents: Your contributions to my childhood and education have always been greatly appreciated. Thanks to you, a child with an unsure future made it to adulthood intact and with a moderately good head on his shoulders. I love you all and want you to know that your efforts will never be forgotten. This degree would not have been possible without you.

I owe more to Paula Tarone than I could ever hope to list. Thank you for your love and support, demonstrated by your willingness to marry a first-year PhD student in December and then move from California wine country to East Lansing (in January!) to be with him. You were a source of strength through all of the stressful times. Thank you for sharing me with the flies.

Thanks to Dan Barbash and Sergey Nuzhdin for your mentorship at UC Davis.

You taught me many of the skills that are required to successfully design, implement, and complete a doctoral degree.

The funds provided by the MSU Graduate School, the MSU Department of Zoology, and the National Institute of Justice were integral to the completion of this research.

Thanks to all of my friends and family.

## TABLE OF CONTENTS

LIST OF TABLES.....	xi
LIST OF FIGURES.....	xii
LIST OF APPENDICES.....	xxv
CHAPTER 1: INTRODUCTION AND BACKGROUND.....	1
The importance of blow flies.....	1
Blow fly development.....	3
Evolution and ecology of blow flies.....	8
Forensic entomology.....	12
The use of developmental plasticity and gene expression for blow fly pest control.....	16
Research.....	17
CHAPTER 2: PLASTICITY IN BLOW FLY GROWTH.....	23
Introduction.....	23
Materials and methods.....	26
<i>Fly collection and rearing</i> .....	26
<i>Statistical analyses</i> .....	29
Results.....	30
<i>Species identification</i> .....	30
<i>Developmental plasticity</i> .....	31
<i>Development on carrion</i> .....	34
Discussion.....	37
<i>Environmental components of variation in the development of L. sericata</i> .....	37
<i>Other potential sources of variation</i> .....	40
<i>Optimal rearing condition using liver and growth on rat carrion</i> .....	41
<i>Applications to forensic entomology</i> .....	41
CHAPTER 3: THE USE OF GENERALIZED ADDITIVE MODELS IN ANALYZING FORENSIC ENTOMOLOGICAL DATA.....	46
Introduction.....	46
Materials and methods.....	48
<i>Species identification</i> .....	48
<i>Growth experiments</i> .....	48
<i>Statistical analyses</i> .....	51
Results.....	54
<i>Species identification</i> .....	54
<i>Immature development</i> .....	54
<i>Assessing statistical models</i> .....	59

<i>GAM validation with independent data</i> .....	63
Discussion.....	64
CHAPTER 4: GENE EXPRESSION IN BLOW FLY EGGS.....	72
Introduction.....	72
Materials and methods.....	74
<i>Species identification and egg collection</i> .....	74
<i>Gene sequencing and primer design</i> .....	74
<i>cDNA synthesis and quantitative PCR</i> .....	75
Results.....	77
Discussion.....	83
CHAPTER 5: PREDICTING BLOW FLY AGE WITH LARVAL AND PUPAL GENE EXPRESSION.....	87
Introduction.....	87
Materials and methods.....	92
<i>Species identification, fly rearing, and collections</i> .....	92
<i>Sequencing of L. sericata loci</i> .....	93
<i>RNA extraction</i> .....	93
<i>Reverse transcription and quantitative PCR</i> .....	94
<i>Statistical analyses</i> .....	95
Results.....	97
<i>Gene sequences</i> .....	97
<i>Quantitative PCR and statistical modeling</i> .....	97
Discussion.....	150
CHAPTER 6: VALIDATION OF GENE EXPRESSION BASED PREDICTIONS OF BLOW FLY AGE WITH BLIND PREDICTIONS.....	168
Introduction.....	168
Methods.....	168
<i>Strain collection and identification</i> .....	168
<i>Collection of unknowns</i> .....	168
<i>Converting RNA later size measurements to live size measurements</i> .....	169
<i>Molecular biology and statistics</i> .....	170
Results.....	171
Discussion.....	178
CHAPTER 7: GENE EXPRESSION ANALYSES OF NON-PUPATING LARVAE.....	182
Introduction.....	182
Methods.....	183
Results.....	184
Discussion.....	189
<i>Potential causes of the “Peter Pan” state</i> .....	189
<i>Interpreting gene expression changes in “Peter Pan” larvae</i> .....	191
<i>Forensic analyses of non-pupating larvae</i> .....	193



CHAPTER 8: POPULATION AND TEMPERATURE EFFECTS ON BODY SIZE AND DEVELOPMENT TIME.....	194
Introduction.....	194
Materials and Methods.....	199
Results.....	200
Discussion.....	210
<i>Temperature effects on body size</i> .....	210
<i>The evolution of age and size at metamorphosis</i> .....	212
<i>Year-to-year variation</i> .....	213
<i>Forensic entomology</i> .....	214
CHAPTER 9: DISCUSSION.....	217
Improved predictions of blow fly age.....	217
<i>Introduction</i> .....	217
<i>Improving accuracy and precision in forensic entomology</i> .....	218
<i>Addressing the Daubert standards in forensic entomology</i> .....	219
<i>Logistical improvements associated with molecular predictions of blow fly         Age</i> .....	221
<i>Future research in forensic entomology</i> .....	222
Controlling blow fly populations with gene expression and plasticity data.....	228
Conclusions.....	229
LITERATURE CITED.....	239

## List of Tables

Table 2.1. Treatment types for the 37 liver-fed cohorts.....	43
Table 2.2. Average development times $\pm$ standard deviations in hours and accumulated degree-days* for significant rearing treatments of <i>Lucilia sericata</i> .....	44
Table 3.1. The 18 generalized additive models for predicting development percent assessed in this experiment.....	69
Table 3.2. Summaries of the estimated significance of each variable in a model.....	70
Table 4.1. Primers used in this study.....	86
Table 5.1. Primers used for sequencing and qPCR for all loci in this experiment.....	160
Table 5.2. Gene sequencing results.....	161
Table 5.3. Generalized additive models assessed in this experiment.....	162
Table 5.4. Estimated significance and degrees of freedom for variables in all GAMs..	163
Table 6.1. Regressions of live length against preserved length for each stage.....	181
Table 6.2. Regressions of live weight against preserved weight for each stage.....	181

## List of Figures

Images in this dissertation are presented in color.

Figure 1.1. The lifecycle of a blow fly. Females lay a clutch of eggs on carrion (1), typically either at an orifice or on a wound. First larval instars hatch from the eggs (2) and feed until they molt into the second instars, which feed and grow until they molt into third instars. Feeding third instars (3) can be identified by three slits in their posterior spiracles, the visibility of blood in their crop, and by the fact that they are actively feeding. Postfeeding third instars (4), which lack visible tissue in the crop, cease feeding and leave the food source to form a puparium. In the puparium (5), the fly metamorphoses and eventually ecloses as an adult by breaking the lid off of the puparium with its ptilinum (top and middle of 5). After a few hours, the adult cuticle and wings achieve a normal appearance and the fly is fully capable of flight (6).....5

Figure 1.2. A phylogeny of the Diptera. Taken from Yeates and Wiegmann (2005). The Nematocera are the most primitive flies and include mosquitoes and crane flies. Cyclorrhapha are the “higher flies”, which have three larval instars and form a puparium. The Schizophora, which include the Calliphoridae and Drosophilidae, have a specialized structure, a ptilinum, which is used to break the lid off of the puparium during eclosion. The Drosophilidae belong to the Acalyptrata, as they lack calypters on their wings. The Calliphoridae belong to the superfamily Oestroidea, which is part of the Calyptrata. Evolutionarily, the Drosophilidae and Calliphoridae are closely related families within the Diptera, which means that they are expected to share many molecular, developmental, and morphological similarities including three larval instars, the formation of a puparium, and the presence of a ptilinum.....11

Figure 1.3. Ecoregions of the United States. Map from work by Hargrove and Hoffman (2004). The white dots indicate approximate origins of the CA, MI, and WV flies studied. Regions of similar color have comparable elevations, soil, and climates. Both the MI and WV strains were located in ecosystems that were mapped in shades of blue, while the CA strain came from a yellow region.....21

Figure 2.1. Developmental variation among *Lucilia sericata* cohorts by treatment type. Boxplots of total development time (hours) for each of the 37 liver-fed cohorts. The line within the box represents the median development hours, the box represents the development times between the 25<sup>th</sup> and 75<sup>th</sup> percentiles, and the ‘whiskers’ (outer-most lines) represent the 5<sup>th</sup> and 95<sup>th</sup> percentiles. 1a: fresh meat daily or no fresh meat daily (FMD vs. NFMD); 1b: paper towel (moist paper towel placed under meat); 1c: transfer: transfer of larvae to fresh substrate for pupation; 1d: destructive (removal of 12 individuals each day). Note: treatments were in combination with other treatment types (Table 1) that had significant effects on development time. For instance, the two outliers in the FMD boxplot (1a) are those that were also destructively sampled.....33

Figure 2.2. Growth curves of *Lucilia sericata* on liver versus growth on rat carrion. Non-linear quantile regression curves created from the lengths of maggots in daily collections of each treatment type. 2a: meat added every 3<sup>rd</sup> day, no moist paper towel, larvae were not transferred to fresh substrate to pupate; 2b: fresh meat daily, no moist paper towels, larvae transferred fresh substrate to pupate; 2c: fresh meat daily, moist paper towel used, larvae transferred to fresh substrate to pupate; 2d: the locally weighted sum of squares curve of data estimated from Grassberger and Reiter (2001) plotted against larval growth on rat carrion. Numbers of cohorts plotted for each treatment were 3, 4, 6, and 6 for Rat, NFMD, FMD, and FMDPT respectively. The solid line on each curve is the 50<sup>th</sup> percentile plot from cohorts raised on rats. Treatments are shown as dashed lines, with the thicker dashed line representing the 50<sup>th</sup> percentile and the thinner lines representing the 97.5<sup>th</sup> and 2.5<sup>th</sup> percentiles (95% confidence intervals). Confidence intervals for the rat cohorts are present in 2d.....35

Figure 2.3. Linear growth of *Lucilia sericata* on liver versus growth on rats. Regression lines of the same treatments displayed in Figure 2.2, for the first three days of growth—the linear phase of development. Line types used to indicate treatments are the same as in Figure 2.2.....36

Figure 2.4. Development times of *Lucilia sericata* cohorts raised on liver versus rat carrion. Comparison of total development hours produced by the 37 liver-fed cohorts in this study to the development of the three rat-fed cohorts. Development time on rats was much less variable than growth on liver, with a development time most similar to the fastest growing liver-fed cohorts. Boxplot design is as in Figure 2.1.....36

Figure 3.1. The lengths (mm) of 2559 immature *L. sericata* throughout immature development (percent of development values are on a 0–1 scale). Note the tight distribution of sizes during the earlier, linear growth phase compared to the more variable postfeeding third instar and pupal stages.....56

Figure 3.2. A plot of the distribution of development percents for individuals at each developmental stage. As development progressed, the proportion of the lifecycle spent in a stage increased. 3<sup>rd</sup> indicates the feeding portion of the third instar; 3<sup>rd</sup>PF indicates the postfeeding stage of the third instar.....56

Figure 3.3. The lengths and weights of individuals throughout development from the 6 cohorts. Growth is compared by strain and by temperature. Solid lines represent the average for all strains or both temperatures. a) Length (mm) plots for each strain. The largest strain, denoted by the line with short bars and spaces, was CA, and the smallest strain, designated by the line with short bars separated by dots, was WV. The MI strain was close to the average size and is represented by the spaced line with long bars and short spaces. Less size variation existed during the feeding portion of the lifecycle (when size was increasing) than in the postfeeding and pupal stages. b) Length plots comparing growth at 20°C versus 33.5°C. Growth at 20°C is represented by the spaced line with short bars separated by dots and 33.5°C is represented by the line with short bars and long spaces. The higher temperature resulted in a growth curve that had a steeper slope during

the linear growth phase of development; individuals from these treatments peaked in body size proportionally faster than cooler treatments, which resulted in smaller body sizes as pupae. c) Weight (mg) plots for each strain. Comparisons among strains were as in (a). d) Weight plots for the two temperature treatments, with similar results as in (b).....58

Figure 3.4. A comparison of diagnostic plots for model 10 (a, c, e, g) and model 18 (b, d, f, h). Model 10 predicted age using length and stage, while model 18 used all data to make age predictions, and explained the most deviance in the data. Panels a) and b) are quantile-quantile plots, assessing the validity of each model's assumptions. The line is relatively straight and increasing for both models, indicating that the distributional assumption of the model did not violate the other assumptions of the model, however model 18 generated a smoother line. Panels c) and d) show the distribution of residuals throughout the lifecycle; in both models residuals are of equivalent size for all ages, showing that there is no bias in residuals based on age. Panels e) and f) are the distributions of residuals, which are normally distributed thus the gamma distribution was acceptable to use with these data. Panels g) and h) are plots of true (response) versus predicted (fitted) values for data used to construct the models. Fitted values accurately predicted true age through linear (feeding) development (the first 25% of the lifecycle), after which the first gap in prediction appeared. This indicates that an overestimation of ages for postfeeding larvae between c. 19.1 and 30% developed was produced using model 10. Model 18 was less likely to overestimate age for individuals at this point in the lifecycle, but predictions were still biased toward overestimating age in young postfeeding third instars. Neither model represents a noticeable improvement over using stage alone for aging pupae. Predicting pupal age with just developmental stage resulted in an age estimate between 43.2 and 100% of immature development. The most predictive model (model 18) plots fitted values between 61.9 and 81.2 % (pupae). At either end of this predicted range, true values could be between 43.2 and 100%. As with all size/stage models, error increased with age.....61

Figure 3.5. A plot of predicted versus true ages of 252 individual flies raised on rats as estimated with a GAM for stage and length (model 10). A plot of predicted versus true development percents of 252 larvae raised on rats as estimated with a GAM for developmental stage and length (model 10). Development is displayed to 50% because length measurements ceased when pupation occurred. The solid line represents the  $Y$  (predicted age) =  $X$  (true age) line. Dashed lines represent 95% confidence intervals for the predictions based on the data in (a). Model 10 accurately (data span  $Y = X$ ) and precisely (ranging c. 5% above and below  $Y = X$ ) predicted age for the feeding portion of the lifecycle (<26% of development), when body size increased in a linear fashion. As expected, precision decreased greatly once the postfeeding third instar was reached, although overall error for the model was within predicted levels.....64

Figure 4.1. Binary gene expression profile for *cs* at 32°C in *L. sericata* eggs, from 0–1 through 8–9 hours of development. 0 indicates no detectable expression of the gene and 1 indicates detectable expression of *cs*. Expression was not detected from 0–2 hours.

Between 2 and 6 hours some eggs clusters expressed *cs* and some did not. After 6 hours, all egg clusters expressed *cs*.....79

Figure 4.2. Standardized expression of *cs* in *L. sericata* eggs, from 0–1 through 8–9 hours of development. The CT was standardized against the average of *rp49* and  $\beta$  *tubulin* 56D CT values and plotted through time. The regression of *cs* CT over time was also included. High expression levels are indicated by low CT values. *cs* was not expressed from 0–2 hours and then its expression increased.....80

Figure 4.3. Standardized transcript abundance of *bcd* in *L. sericata* eggs, from 0–1 through 8–9 hours of development. Standardization was as in Figure 4.2. *bcd* gene expression was highest from 0–2 hours, then transcripts decreased in abundance.....81

Figure 4.4. Standardized transcript abundance of *sll* in *L. sericata* eggs, from 0–1 through 8–9 hours of development. Standardization was as in Figure 4.2. *sll* expression was highest from 0–2 hours, then tended to be lower as eggs developed, though variance in expression was high.....82

Figure 4.5. Predicted (fitted) versus true (response) ages for a generalized additive model that made age predictions for the 33 egg masses that expressed *bcd*, *sll*, and *cs*. Estimated ages were within 2 h of the true age in 30 of the 33 cases.....83

Figure 5.1. The positive correlation between housekeeping genes. The CT scores followed a line with a slope of 1 (after accounting for the average CT difference of 1.2) with a correlation of 0.84.....98

Figure 5.2. Variance in gene expression due to the standardization against housekeeping genes. Since the average of both housekeeping genes was used for standardization, one half of the difference between *rp49* and the *beta tubulin* gene was plotted throughout development. The variance in this plot was relatively constant and indicates that an approximate difference in expression of up to 4 fold (2 CT units) could be explained by variation in housekeeping gene expression.....99

Figure 5.3. Standardized gene expression of *scalloped wings* throughout the immature development of *L. sericata*. No apparent pattern of gene expression was observed throughout development.....100

Figure 5.4. Standardized gene expression of *wingless* throughout the immature development of *L. sericata*. No informative pattern of gene expression was observed throughout development.....101

Figure 5.5. Standardized gene expression of *rhodopsin 3* throughout the immature development of *L. sericata*. No apparent pattern of gene expression was observed throughout development.....102

Figure 5.6. Standardized *chitin synthase* CT scores plotted by stage. The postfeeding third instar stands out as expressing significantly less *cs* than the other stages.....103

Figure 5.7. Standardized *chitin synthase* CT scores plotted by development percent for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The shift in the red line between (a) and (b) indicates a cluster of individuals that did not express the gene at the time point where gene expression is at its lowest levels for this locus.....104

Figure 5.8. Standardized *chitin synthase* CT scores plotted by development percent at 33.5°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The shift in the red line between (a) and (b) indicates a cluster of individuals that did not express the gene at that time point.....105

Figure 5.9. Standardized *chitin synthase* CT scores plotted by development percent at 20°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. Few individuals did not express the gene at this temperature. The CA strain expressed more of *cs* than the MI strain, though in the same pattern.....106

Figure 5.10. Standardized *heat shock protein 60* CT scores plotted by stage. Gene expression was highest during the first two stages and lowest during the last two stages.....107

Figure 5.11. Standardized *heat shock protein 60* CT scores plotted by development percent for all individuals that expressed the gene. The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The expression levels of this gene decrease from hatching, until 60 percent development (early pupation), then increase until eclosion...108

Figure 5.12. Standardized *heat shock protein 60* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The strains expressed the gene in the same general pattern, though the CA strain had higher expression around 40 percent development.....109

Figure 5.13. Standardized *heat shock protein 60* CT scores plotted by development percent at 20°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA

strain expressed more of this gene than the MI strain, at this temperature. Few individuals did not express the gene at this temperature.....110

Figure 5.14. Standardized *heat shock protein 90* CT scores plotted by stage. Gene expression was lowest during the postfeeding third instar.....111

Figure 5.15. Standardized *heat shock protein 90* CT scores plotted by development percent for all individuals that expressed the gene. The black line is the lowest curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The expression levels of the gene decreased from hatching, until approximately 25 percent development (when postfeeding larval development is attained), then increased until eclosion when flies were raised at the high temperature. At the low temperature, expression was higher than at high temperatures, and expression was maintained at a steady level after 25 percent development.....112

Figure 5.16. Standardized *heat shock protein 90* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. Expression decreased until the postfeeding third instar (around 25 percent) then increased until eclosion. The strains expressed the gene in the same general pattern, though the time of minimum expression occurred later in the CA strain.....113

Figure 5.17. Standardized *heat shock protein 90* CT scores plotted by development percent at 20°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more of this gene than the MI strain, at this temperature.....114

Figure 5.18. Standardized *acetylcholine esterase* CT scores plotted by stage. Gene expression decreased through the third instar, then increased until eclosion.....115

Figure 5.19. Standardized *acetylcholine esterase* CT scores plotted by development percent for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The black line is the lowest curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The shift in the lines between (a) and (b) indicates a cluster of individuals that do not express the gene at the time point where gene expression is at its lowest levels for this locus. Many individuals did not express *ace* between approximately 20 and 50 percent development (late third instar through early pupation). This time period mirrors the period of least *ace* expression at the high temperature treatment. Expression increased through development in the low temperature treatments.....116

Figure 5.20. Standardized *acetylcholine esterase* CT scores plotted by development percent at 33.5°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green



line indicates gene expression in the WV strain. The shift in these lines between (a) and (b) indicates a cluster of individuals that do not express the gene at that time point.....117

Figure 5.21. Standardized *acetylcholine esterase* CT scores plotted by development percent at 20°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more *ace* than the MI strain (a). MI was also more likely to not express *ace*.....117

Figure 5.22. Standardized *ecdysone receptor* CT scores plotted by stage. Gene expression increased through the postfeeding third instar, then decreased until eclosion.....118

Figure 5.23. Standardized *ecdysone receptor* CT scores plotted by development percent for all individuals that expressed the gene. The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The expression levels of this gene increase from hatching, until approximately 35 percent development (during the postfeeding third instar), then decrease until eclosion.....119

Figure 5.24. Standardized *ecdysone receptor* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The strains expressed the gene in the same general pattern, though the MI strain expressed less *ecr* than the other strains....119

Figure 5.25. Standardized *ecdysone receptor* CT scores plotted by development percent at 20°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more of this gene than the MI strain, at this temperature, though they converge to the same maximum expression level during the postfeeding third instar.....120

Figure 5.26. Standardized *resistance to organophosphate 1* CT scores plotted by stage. Gene expression increased through the postfeeding third instar, then decreased until eclosion.....121

Figure 5.27. Standardized *resistance to organophosphate 1* CT scores plotted by development percent for all individuals that expressed the gene. The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The expression levels of this gene increase from hatching until approximately 25 percent development at high temperatures and until approximately 35 percent development at low temperatures, then gene transcript abundance decrease until eclosion.....122

Figure 5.28. Standardized *resistance to organophosphate 1* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The strains expressed the gene in the same general pattern.....123

Figure 5.29. Standardized *resistance to organophosphate 1* CT scores plotted by development percent at 20°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more of the gene than the MI strain, at this temperature, though both followed the same pattern.....124

Figure 5.30. Standardized *white* CT scores plotted by stage. Gene expression increased through the postfeeding third instar, then decreased until eclosion.....125

Figure 5.31. Standardized *white* CT scores plotted by development percent for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The shift in the lines between (a) and (b) indicates a cluster of individuals that did not express the gene at the time point where gene expression is at its lowest levels for this locus. The expression levels of the gene increased from hatching until approximately 25 percent development at high temperatures and until approximately 35 percent development at low temperatures, then gene transcript abundance decrease until eclosion. High temperature treatments were less likely to express the gene.....126

Figure 5.32. Standardized *white* CT scores plotted by development percent at 33.5°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The shift in the lines between (a) and (b) indicates that CA was most likely to not express *w* and MI was most likely to express the gene.....127

Figure 5.33. Standardized *white* CT scores plotted by development percent at 20°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more *w* than the MI strain (a). MI was also more likely to not express *w* during pupation (the last half of development).....128

Figure 5.34. Standardized *ultraspiracle* CT scores plotted by stage. Gene expression increased through the postfeeding third instar, then decreased until eclosion.....129

Figure 5.35. Standardized *ultraspiracle* CT scores plotted by development percent for all individuals that expressed the gene. The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for

expression at 33.5°C. The expression levels of the gene increased from hatching until approximately 35 percent development, then gene transcript abundance decreased until eclosion.....130

Figure 5.36. Standardized *ultraspiracle* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The strains expressed the gene in the same general pattern.....130

Figure 5.37. Standardized *ultraspiracle* CT scores plotted by development percent at 20°C for all individuals that expressed the gene. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more of the gene than the MI strain, though both followed a similar pattern.....131

Figure 5.38. Standardized *slalom* CT scores plotted by stage. Gene expression increased through the third instar, then decreased until eclosion.....132

Figure 5.39. Standardized *slalom* CT scores plotted by development percent for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The shift in the lines between (a) and (b) indicates a cluster of individuals that did not express the gene at the time point where gene expression was at its lowest levels. The expression levels of the gene increased from hatching until approximately 25 percent development, then gene transcript abundance decreased until approximately 50 percent, and then increase again until eclosion. High temperature treatments were less likely to express the gene.....133

Figure 5.40. Standardized *slalom* CT scores plotted by development percent at 33.5°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The shift in the lines between (a) and (b) indicates that CA was most likely to not express *sll* and MI was most likely to express the gene.....134

Figure 5.41. Standardized *slalom* CT scores plotted by development percent at 20°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The strains expressed the gene in a similar pattern (a). MI was more likely to not express *sll*.....134

Figure 5.42. The combined standardized expression (CT score) patterns for all nine genes throughout *L. sericata* immature development. The lowess curves for each gene represent expression for all individuals that expressed the gene. The combined expression patterns of the genes were used to predict blow fly age with GAMs.....136

Figure 5.43. Diagnostic plot for Model 1, which used developmental stage to predict *L. sericata* development percent (on a scale of 0–1). A gamma distribution was used with this model. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. Error increased with age, and gaps exist between predictions for each developmental stage.....137

Figure 5.44. Diagnostic plot for Model 15, which used developmental stage, strain, temperature, length, and weight to predict *L. sericata* development percent (on a scale of 0–1). A gamma distribution was used with this model. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. Error increased with age, and a gap exists between predictions for postfeeding third instar and pupal ages.....138

Figure 5.45. Diagnostic plot for Model 18, which used developmental stage, strain, temperature, length, weight, binary gene expression for four genes, and quantitative gene expression for nine genes, to predict *L. sericata* development percent (on a scale of 0–1). A gamma distribution was used with this model. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to other models and the gap between predictions for postfeeding third instar and pupal ages has shrunk.....139

Figure 5.46. Diagnostic plot for Model 19, which used the same parameters as Model 18, to predict *L. sericata* development percent (on a scale of 0–1), but a gaussian distribution was used. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to Model 15 and the gap between predictions for postfeeding third instar and pupal ages has shrunk. However, compared to Model 18, this model exhibits more error in predictions of younger individuals.....140

Figure 5.47. Diagnostic plot for Model 20, which used the same parameters as Model 18, to predict *L. sericata* development percent (on a scale of 0–1), but a gaussian distribution was used with this model and all genes expression scores were anchored against *hsp60* expression. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to Model 15 and the gap between predictions for postfeeding third instar and pupal ages has been eliminated. However, compared to Model 18, this model exhibits more error in predictions for younger individuals, though it is an improvement over predictions made with Model 19.....141

Figure 5.48. Diagnostic plot for Model 21, which used the same parameters as Model 18, to predict *L. sericata* development percent (on a scale of 0–1), but a gaussian distribution was used with this model and all genes expression scores were anchored against length measurements. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to Model 15 and the gap between predictions for

postfeeding third instar and pupal ages has been eliminated. However, compared to Model 18, this model exhibits more error in predictions for younger individuals, though it is an improvement over predictions made with Model 19.....143

Figure 5.49. Diagnostic plot for Model 22, which used the same parameters as Model 21, to predict *L. sericata* development percent (on a scale of 0–1), but parameters that were non-significant predictors of age in Model 21 (length, weight, *cs*, *w*, strain) were removed. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to Model 15 and the gap between predictions for postfeeding third instar and pupal ages has been eliminated. However, compared to Model 18 this model exhibits more error in predictions for younger individuals, though it is an improvement over predictions made with Model 19. Removing strain, *cs*, and *w*, had little effect on the predictions made with the model compared to predictions made with Model 21.....144

Figure 5.50. Diagnostic plot for Model 23, which used the same parameters as Model 22, to predict *L. sericata* development percent (on a scale of 0–1), except that temperature was removed from this model. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. Removing temperature had little effect on the predictions made with this model compared to predictions made with Model 22.....145

Figure 5.51. Diagnostic plot for postfeeding third instar incorporating stage/length/weight/ temp/strain. Descriptions for each panel type can be found in Figure 3.4. Note the extensive scatter in response vs. fitted values.....147

Figure 5.52. Diagnostic plot for pupae incorporating stage/length/weight/ temp/strain. Descriptions for each panel type can be found in Figure 3.4. Note the extensive scatter in response vs. fitted values.....148

Figure 5.53. Diagnostic plot for postfeeding third instar incorporating stage/length/weight/ temp/strain and expression data. Descriptions for each panel type can be found in Figure 3.4. Note the tightening in response vs. fitted values over Figure 5.51.....149

Figure 5.54. Diagnostic plot for postfeeding third instar incorporating stage/length/weight/ temp/strain and expression data. Descriptions for each panel type can be found in Figure 3.4. Note the tightening in response vs. fitted values over Figure 5.52.....150

Figure 6.1. Temperature plots for the ambient temperature experiment. Temperatures spiked in the afternoon as sunlight was directly on the terrarium.....172

Figure 6.2. The regression of ADH (base 10) percents versus percent development in hours. The values were strongly correlated, with an R-squared and slope almost 1.

However, the 1.2448 shift up the Y axis indicates that ADH is a better measure of developmental progress.....172

Figure 6.3. Predicted versus true development percents for all 90 individuals in the blind study. Predictions were made with GAMs that did not use gene expression information. The four points in the gray bar represent the four larvae that failed to pupate, and were eliminated from analyses in this section.....174

Figure 6.4. Predicted versus true development percents for the 86 normally developing individuals in the blind study. Predictions were made with GAMs that used unanchored gene expression terms.....176

Figure 6.5. Predicted versus true development percents for 49 normally developing individuals in the blind study. Predictions were made with GAMs that used unanchored gene expression terms and were Type 2 models.....176

Figure 6.6. Predicted versus true development percents for 71 normally developing individuals in the blind study. Predictions were made with GAMs that used length anchored gene expression terms.....177

Figure 6.7. Predicted versus true development percents. Predictions were made with GAMs that used length anchored gene expression terms and were Type 2 models.....177

Figure 7.1. Standardized gene CT scores for *hsp90*, plotted by stage, for all individuals from the previous section, for the 55 non-pupator individuals from this study, and the individuals in the blind study. The left panel represents expression from a previous section (Chapter 5), with results from the 55 non-pupators (NP) placed next to postfeeding third instars (3rdPF). The right panel represents gene expression for the individuals in the blind study (Chapter 6), including the 4 non-pupators. 1<sup>st</sup> = first instar. 2<sup>nd</sup> = second instar. 3<sup>rd</sup> = feeding third instar. 3rdPF = postfeeding third instar. NP = non-pupator or “Peter Pan” larvae. On average, *hsp90* was expressed at higher levels in non-pupators than in postfeeding third instars.....186

Figure 7.2. Standardized gene CT scores for *rop-1*, plotted by stage for all individuals from the previous section, for the 55 non-pupator individuals from this study, and the individuals in the blind study. The left panel represents expression from a previous section (Chapter 5), with results from the 55 non-pupators (NP) placed next to postfeeding third instars (3rdPF). The right panel represents gene expression for the individuals in the blind study (Chapter 6), including the 4 non-pupators. 1<sup>st</sup> = first instar. 2<sup>nd</sup> = second instar. 3<sup>rd</sup> = feeding third instar. 3rdPF = postfeeding third instar. NP = non-pupator. On average, *rop-1* was expressed at lower levels in non-pupators than in postfeeding third instars, though this pattern was not apparent in the blind study.....187

Figure 7.3. Standardized gene CT scores for *sll*, plotted by stage, for all individuals from the previous section, for the 55 non-pupator individuals from this study, and the individuals in the blind study. The left panel represents expression from a previous

section (Chapter 5), with results from the 55 non-pupators (NP) placed next to postfeeding third instars (3rdPF). The right panel represents gene expression for the individuals in the blind study (Chapter 6), including the 4 non-pupators. 1<sup>st</sup> = first instar. 2<sup>nd</sup> = second instar. 3<sup>rd</sup> = feeding third instar. 3rdPF = postfeeding third instar. NP = non-pupator. On average, *sll* was expressed at lower levels in non-pupators than in postfeeding third instars.....188

Figure 7.4. Standardized gene CT scores for *usp*, plotted by stage, for all individuals from the previous section, for the 55 non-pupator individuals from this study, and the individuals in the blind study. The left panel represents expression from a previous section, with results from the 55 non-pupators (NP) placed next to postfeeding third instars (3rdPF). The right panel represents gene expression for the individuals in the blind study, including the 4 non-pupators. 1<sup>st</sup> = first instar. 2<sup>nd</sup> = second instar. 3<sup>rd</sup> = feeding third instar. 3rdPF = postfeeding third instar. NP = non-pupator. On average, *usp* was expressed at lower levels in non-pupators than in postfeeding third instars.....189

Figure 8.1. Boxplots of minimum development time comparisons among strains at 33.5°C in 2006. The CA strain developed more slowly compared to the other strains.....202

Figure 8.2. Boxplots of minimum development times comparisons among strains at 20°C in 2006. The strains exhibited similar median and fastest development times, but CA had a faster mean development time as compared to the other strains, which had some minimum development times that were longer than 540 h.....203

Figure 8.3. A comparison of average minimum development times for each strain in 2005 (a) and in 2006 (b). CA is red, MI is blue, WV is green. The MI mean minimum development time was consistently faster than WV, though their development times were comparable. The CA mean was fastest at 20°C and slowest at 33.5°C in both years. Development was faster at 33.5°C.....204

Figure 8.4. Boxplots of average weights compared among the strains at 33.5°C in 2006. On average, CA was larger than MI, which was larger than WV.....206

Figure 8.5. Boxplots of average weights compared among the strains at 20°C in 2006. On average, CA was larger than MI, which was larger than WV.....207

Figure 8.6. A comparison of average lengths for each strain in 2005 (a) and in 2006 (b). CA is red, MI is blue, WV is green. The CA mean length was the largest in both years. WV was the smallest in 2006, but MI was smallest at 33.5°C in 2005. Lengths were greater for 20°C treatments.....208

Figure 8.7. A comparison of average weights for each strain in 2005 (a) and in 2006 (b). CA is red, MI is blue, WV is green. The CA mean weight was the largest and WV was the smallest. Weights were greater for 20°C treatments.....209

LIST OF APPENDICES

Appendix 2.1. Treatments and duration of the immature lifecycle and individual stages  
from all cohorts of *Lucilia sericata*.....232



## CHAPTER 1: INTRODUCTION AND BACKGROUND

### *The Importance of Blow Flies*

Flies from the family Calliphoridae (Diptera), commonly known as blow flies, are parasites, parasitoids, and primary successional species on carrion, which regularly interact with people in several key areas of human activity. Calliphorids are parasites of sheep and cattle, resulting in financial damage that was significant enough, in one instance, to warrant the development of a sterile male release program to eliminate the primary screwworm fly *Cochliomyia hominivorax* from Texas and Florida (Crystal and Ramirez 1975). Blow flies also influence human health. Maggot therapy has been employed at some hospitals to efficiently debride wounds (Sherman and My-Tien 1995). More importantly, they detrimentally affect human health, as they can transmit diseases (Faulde et al. 2001) and cause myiasis (parasitization) of humans (Sherman 2000). Calliphorid flies are also useful to forensic scientists and medical examiners. The immature forms of these flies are frequently found at death scenes in association with bodies, providing valuable information for estimating a postmortem interval (PMI) in death investigations (Haskell and Catts 1990; Greenberg and Kunich 2002). Due to these interactions, a more detailed knowledge of blow fly development has the potential to promote the beneficial interactions, and ameliorate negative interactions with humans, by enabling the production of tools that can be used to better describe and control blow fly development.

As genomic sciences progress, detailed knowledge of development in non-model organisms will help to answer questions in the emerging field of comparative genomics.

Evolutionary biologists compare genomes to understand how selection acts on them and to determine whether adaptations among similar organisms are due to gene sequence evolution or differences in the regulation of those genes (Yeates and Wiegmann 2005, Khaitovich et al. 2006). Medical researchers also benefit from comparative genomics, as they can learn how to apply results from model organism research to human medicine. Both fields will increasingly rely on genomic comparisons to answer questions about how a gene is regulated in similar species, and the consequences of differential regulation (Collins et al. 1998).

*Drosophila melanogaster* is a well-established model organism that is poised to be a central pillar of comparative genomic research. In the near future, twelve *Drosophila* genome sequences will be complete ([www.flybase.org](http://www.flybase.org)), allowing for multi-species comparisons (Kulathinal and Hartl 2005). Such comparisons have already revealed that the conservation of a seven-striped pattern (known as a pair-rule pattern) of gene expression in the embryos of this genus is due to different variants of the same regulatory mechanisms (Ludwig et al. 1998). Essentially, each species has been shown to use the same set of gene expression-inducing and inhibiting proteins, in subtly different ways, to achieve the same embryonic pattern.

Just as comparisons within the *Drosophila* genus are valuable, comparisons of gene regulation between *Drosophila* and other higher flies (including blow flies) have revealed insight into the nature of gene expression evolution, including *bicoid* (McGregor 2005), *wingless* (Mellenthin et al. 2006), and *slalom* (Ali et al. 2005). The sequences, timing, and locations of *wingless* and *slalom* expression are conserved between *Drosophila* and blow flies. Similarly, *bicoid* serves the same anterior determination

function in blow flies as it does in *Drosophila*, though the blow fly genes cannot fully rescue *Drosophila* mutations due to regulatory sequence evolution. Likewise, a comparison of gene expression within the Calliphoridae has shown that acrostichal bristle differences between *Calliphora vicina* and *Protophormia terraenovae* are due to heterochrony in the expression of proneural genes, with larger bristles resulting from earlier initiation of the genes (Skaer et al. 2003). Given the advances that have already arisen from comparisons of developmentally regulated genes in blow flies to homologous genes in other species, it is likely that any additional knowledge of blow fly development, at a molecular level, will further contribute to the understanding of gene expression and phenotype evolution among higher flies, and thus among animals in general.

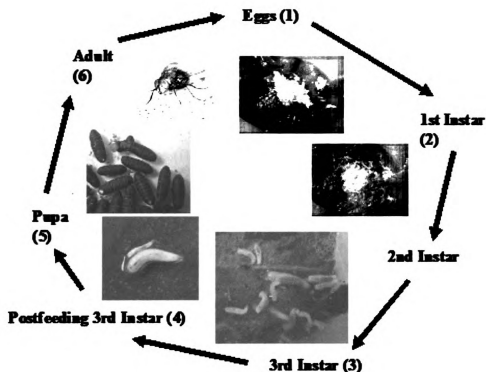
### *Blow Fly Development*

The interactions of blow flies with humans occur primarily with immature developmental stages (the one exception is the transmission of disease, which is due to adult flies feeding on filth and human food sources). Also, it is necessary to have knowledge of the development of an organism, before it is possible to understand how that process evolved. Given the potential importance of immature blow flies to forensic, medical, veterinary, and evolution sciences, a detailed understanding of blow fly development is required.

The lifecycle of a blow fly progresses predictably (Catts and Haskell 1990) as shown below (Figure 1.1). Female blow flies lay eggs around orifices or wounds. The eggs hatch into larvae, which feed on a body and grow through three larval instars. Each instar is separated by a molt of the cuticle that enables further larval growth. The

appearances of the anterior and posterior spiracles differ among instars, making them easily distinguishable from each other. During the third instar, larvae cease feeding and (usually) leave the body to form a hardened cuticular structure known as a puparium. Within the puparium, the fly experiences metamorphosis and eventually ecloses as an adult by breaking the lid off of the puparium with an extendable sac called a ptilinum.

An immature lifecycle with three instars is a hallmark of cyclorrhaphan or “higher flies”, like the Drosophilidae and Calliphoridae, which pupate within a puparium (McAlpine 1989, Yeates and Wiegmann 2005). Similarities in development among cyclorrhaphans can be used to infer molecular processes that drive developmental changes in blow flies. As all flies develop, they must incorporate information from physiological and environmental signals to determine when it is appropriate to molt or metamorphose. They are interpreted in the brain where neurosecretory cells respond to information provided by the insulin pathway (Shingleton 2005), temperature, and photoperiod (Flatt et al. 2005); these stimuli ultimately cause the cells to induce prothoracicotropic hormone (PTTH) secretions from the corpora allata (Kalthoff 2001). The release of PTTH directs the prothoracic glands to produce a steroid hormone, ecdysone, and release it into the hemolymph.



**Figure 1.1. The lifecycle of a blow fly.** Females lay a clutch of eggs on carrion (1), typically either at an orifice or on a wound. First larval instars hatch from the eggs (2) and feed until they molt into the second instars, which feed and grow until they molt into third instars. Feeding third instars (3) can be identified by three slits in their posterior spiracles, the visibility of blood in their crop, and by the fact that they are actively feeding. Postfeeding third instars (4), which lack visible tissue in the crop, cease feeding and leave the food source to form a puparium. In the puparium (5), the fly metamorphoses and eventually ecloses as an adult by breaking the lid off of the puparium with its pilinum (top and middle of 5). After a few hours, the adult cuticle and wings achieve a normal appearance and the fly is fully capable of flight (6).

The release of ecdysone initiates larval molt or metamorphosis by activating a series of gene expression changes that were first observed in *Drosophila*. Experiments by Ashburner (1974) demonstrated that when polytene chromosomes were treated with ecdysone, they developed thickened regions, called “puffs”, which appeared in a temporally specific order. Early puffs were shown to respond directly to ecdysone treatments, while late puffs require the expression of genes (Kalthoff 2001). Since the

early *Drosophila* experiments, there have been significant discoveries as to how ecdysone affects development at a molecular level and many of the genes expressed within puffs are currently known (Kalthoff 2001). For instance, the *ecdysone receptor* (*ecr*) and *ultraspiracle* (*usp*) genes, which are both specific types of transcription factors known as nuclear hormone receptors, form a heterodimer that when bound to ecdysone, activates the expression of the early genes (Henrich and Brown 1995). In fact, multiple nuclear hormone receptors in *Drosophila* have been shown using northern blot analysis to respond to ecdysone pulses (Sullivan et al. 2002). Recent genomic studies in *D. melanogaster* indicate that thousands of genes are regulated throughout development (Arbeitman et al. 2002) and thousands of genes respond to changes in ecdysone concentrations (Beckstead et al. 2005). As the genomic age advances, most (possibly all) of the identities of the loci that are expressed within chromosomal puffs will be revealed, as well as the genes that interact with them, enabling an improved understanding of how flies develop.

The response of a developing fly to ecdysone pulses, however, is not uniform among stages. Juvenile hormone (JH) modulates the nature of the response to an ecdysone pulse, through a currently unknown mechanism (Henrich and Brown 1995, Flatt et al. 2005). Like PTTH, JH is released by the corpora allata in response to temperature, photoperiod, and hormonal signals that are relayed from the nervous system (Kalthoff 2001, Flatt et al. 2005). When JH is in high concentrations, an ecdysone pulse initiates the next larval molt. However, when an ecdysone pulse occurs in the absence of JH, metamorphosis is initiated.

Though development is predictable, environmental factors are known to affect the development of flies and this plasticity can occur via several mechanisms. Temperature affects blow fly development times, with lower temperatures causing slower development (e.g. Anderson 2000, Grassberger and Reiter 2001). Blow fly larvae also diapause (an environmentally induced delay of development) at low temperatures and under short photoperiods (Tachibana and Goto 2004, Tachibana et al. 2005), though different populations can exhibit different propensities for this behavior (Greenberg and Kunich 2002). In *Drosophila*, insufficient diet has also been shown to cause delays in development as well as small body size. Nutritional plasticity is thought to be influenced by the insulin receptor pathway (Shingleton et al. 2005). Temperature sensitive *insulin receptor* (*Inr*) mutations can affect development time if their products are inactivated before the acquisition of a critical size (when a larva has accumulated enough energy to pupate). If the larva has fed enough that it has sufficient energy to complete metamorphosis, however, the inactivation of *Inr* protein will cause a decrease in body size. Similar delays in blow fly development have been observed when *L. sericata* and *Calliphora vicina* (Diptera: Calliphoridae) were exposed to different diets (Kaneshraja and Turner 2004, Clark et al. 2006, Tarone and Foran 2006), indicating the potential for a similar response in these species.

Unfortunately, little is known about the molecular control of blow fly development. Though similarities in developmental biology among cyclorrhaphan flies can be inferred from research in *D. melanogaster* and other dipterans (reviewed by Henrich and Brown 1995), the specific molecular controls of development are not always the same across all species. For instance, differences in *ultraspiracle* expression between

*Chironomus tentans* and other insects have been observed, with the *C. tentans* gene located within a late puff, indicating a potentially novel regulatory mechanism in the ecdysone response of this fly (Henrich and Brown 1995). There is also a difference between *L. sericata* and *Sarcophaga bullata* for *hsp90* expression during diapause, with *S. bullata* down-regulating the gene during diapause while expression of the *L. sericata* gene is unchanged (Tachibana et al. 2005). However, many genes have comparable functions and regulation in Cyclorrhaphan flies. There are similarities between the Calliphoridae and Drosophilidae in *bicoid* protein functions (McGregor 2005), which determines the anterior region of embryos, though Calliphorid *bicoid* fails to fully rescue *Drosophila* mutants due to differences in regulatory sequences among species. Two other genes, *wingless* (*wg*) (Mellenthin et al. 2006) and *slalom* (*sll*) (Ali et al. 2005), have highly conserved gene expression patterns in *L. sericata* and *D. melanogaster*. Cell signaling, as directed by *wg*, occurs in the same tissues and in the same patterns on wing imaginal discs and *sll* is expressed in the salivary glands of larvae in both species. Observations of comparable and dissimilar gene expression patterns in Cyclorrhaphan flies highlight a need for specific information regarding the molecular genetics of blow fly development to identify when molecular results can be generalized across all flies and when they are specific to a lineage.

### *Evolution and Ecology of Blow Flies*

To fully understand blow fly development, it is necessary to interpret the results of developmental research in the context of blow fly evolution and ecology. The phylogenetic relationships among the Diptera (Figure 1.2) were recently and extensively

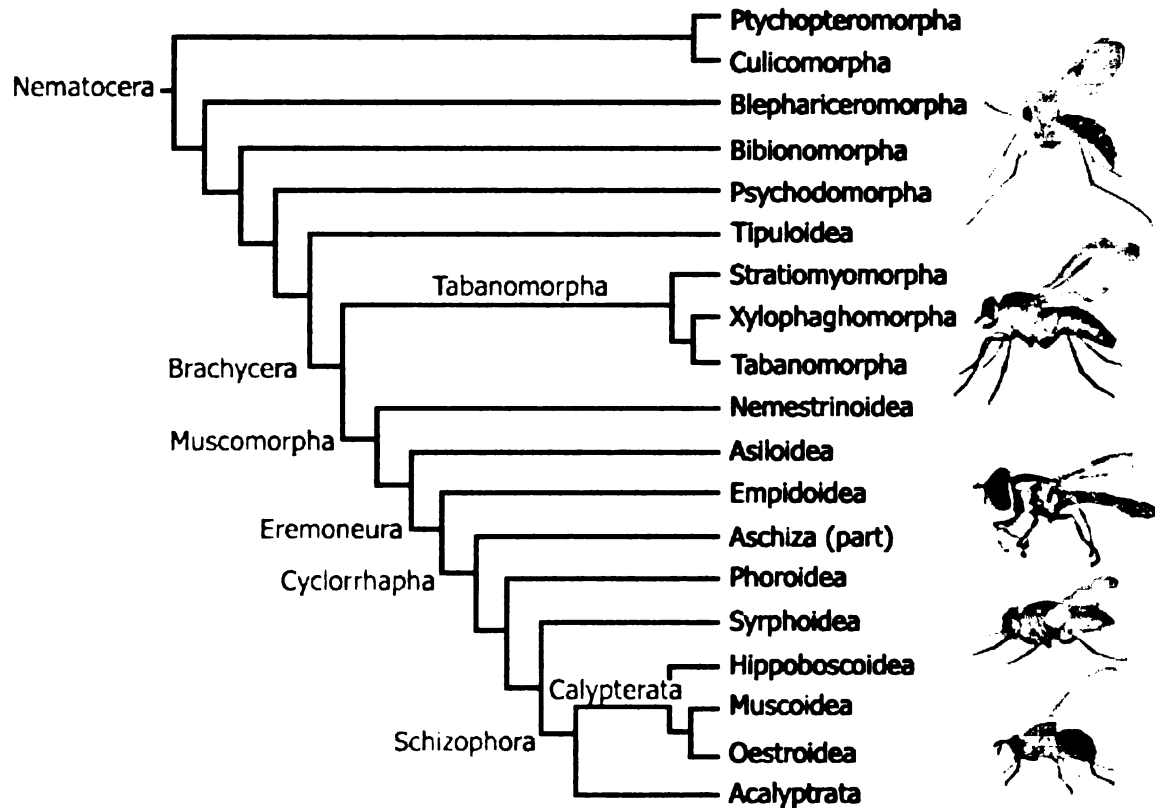


reviewed by Yeates and Wiegmann (2005). The earliest known Diptera are estimated to have diverged from other insects 233 Mya based on fossil evidence, and 248–283 Mya based on molecular evidence. The most primitive flies are the Nematocera, which includes mosquitoes and crane flies. They typically have 4–8 larval instars and pupae with visible appendages. They are generally aquatic and depend on water/moisture for reproduction and/or survival. The next infraorder of flies is the Brachycera, which are estimated to have diverged approximately 208 Mya (fossil) to 194–241 Mya (molecular). They are a monophyletic group as compared to the Nematocera, with many traits, such as male genital rotation and the number of ganglia in the nervous system, that are intermediate to the Cyclorrhapha and Nematocera.

The Cyclorrhapha are another well-supported monophyletic infraorder. They have reduced larval head structures and three larval instars that are both presumed to be adaptations that are related to pupation within a puparium. They diverged from other flies approximately 130 Mya (fossil) to 122–172 Mya (molecular). Within this group are the Schizophora, which eclose from the puparium with a ptilinum. The Schizophora include the Drosophilidae and Calliphoridae and are estimated to have diverged 80 Mya (fossil) to 71–113 Mya (molecular). However, the Drosophilidae belong to the Acalypterata, whereas the Calliphoridae belong to the Calypterata. The major distinction between the two clades is the presence of a calypter (a specialized wing structure attached to the basal portion of the wing) in the Calypterata. The *Drosophila* and *Musca* (the latter also belonging to the Calypterata) genera are estimated to have diverged approximately 70 Mya (fossil) to 29–76 Mya (molecular). Based on the similarities in highly derived dipteran characteristics, molecular and developmental data from the

*Drosophilidae* and other cyclorrhaphan flies are likely to be directly relevant to blow fly biology, and may be used as a guide when data do not exist for the *Calliphoridae*.

Many of the *Calliphoridae* have successfully occupied a niche as the primary successional species on carrion. Blow flies are capable of locating carrion within hours of death (Catts and Haskell 1990), allowing them to regularly find and dominate a rich but ephemeral food source. The sporadic occurrence of carrion results in a potential tradeoff between body size and age at metamorphosis. Such a tradeoff has been studied in spadefoot toads (Morey and Reznick 2000, Day and Rowe 2002, Gomez-Mestre and Bucholz 2006), as well as in mites (Plaistow 2004) and butterflies (Kingsolver 2000, Kingsolver et al. 2001) and it seems that blow flies are susceptible to similar ecological forces as these species. For example, spadefoot toads are under intense selection to complete development before their vernal ponds dry up, mirroring the need for blow flies to utilize carrion before it disappears. However, selection has also been shown to favor larger adult toads that require longer feeding periods, resulting in a tradeoff between the two phenotypes. Additionally, the tradeoff between development time and body size is associated with the evolution of plasticity in both traits that has also been observed in blow flies. The *Calliphoridae* exhibit a range of minimum development times and body sizes (Kamal 1958, Anderson 2000) and limited nutrition can both delay development and lead to smaller body sizes (Kaneshrajah and Turner 2004, Clark et al. 2006). Given the similarities between the toad and fly systems, information on blow fly development may also contribute to the study of the evolution of age and size at metamorphosis.



**Figure 1.2. A phylogeny of the Diptera.** Taken from Yeates and Wiegmann (2005). The Nematocera are the most primitive flies and include mosquitoes and crane flies. Cyclorrhapha are the “higher flies”, which have three larval instars and form a puparium. The Schizophora, which include the Calliphoridae and Drosophilidae, have a specialized structure, a ptilinum, which is used to break the lid off of the puparium during eclosion. The Drosophilidae belong to the Acalyptata, as they lack calypters on their wings. The Calliphoridae belong to the superfamily Oestroidea, which is part of the Calypttrata. Evolutionarily, the Drosophilidae and Calliphoridae are closely related families within the Diptera, which means that they are expected to share many molecular, developmental, and morphological similarities including three larval instars, the formation of a puparium, and the presence of a ptilinum.

An ecological factor that has been shown to affect the biology of body size and development time of many organisms is latitude. This is best demonstrated by Bergmann’s rule, which states that endothermic organisms are larger in cooler climates (Ridley 1996). However, the same concept applies to insects in many cases. Latitudinal clines for several traits and genotypes are well documented in the *Drosophila* (reviewed

by Powell 1997). Dobzhansky demonstrated that chromosomal inversions in populations of *D. pseudoobscura* and *D. persimilis* were geographically distributed along a latitudinal grade. Similarly, the *alcohol dehydrogenase* locus has been shown to vary in such a manner, as has body size in populations of *D. melanogaster* on multiple continents. Latitudinal variation occurs in other dipteran species as well. For example, *Rhagoletis pomonella* (Diptera: Tephritidae) exhibits latitudinal differences in development time (Feder et al. 2003), highlighting the potential for latitudinal effects on blow fly development.

Given the potential effects of latitude and other ecological factors on blow fly development, adaptation to the local environment must be considered when evaluating developmental traits for specific populations. Though differences in body size and development time within the Calliphoridae have been briefly mentioned in the literature (Greenberg 1991, Grassberger and Reiter 2001), the potential for genetic differences for these traits, due to local adaptations, have not, until recently, been discussed as potential lines of future research on blow fly development (Tarone and Foran 2006, Goff 2007). Differences among recorded development times for *L. sericata* (Kamal 1958, Greenberg 1991, Anderson 2000, Grassberger and Reiter 2001) could be explained by genetic differences among strains, though the interpretation of variation in development time is clouded by potential plasticity caused by variable rearing conditions among the studies. The observed variation in development times indicates a need for quantitative genetic comparisons among populations of blow flies to differentiate between genetic and environmental effects on body size and development time traits in the Calliphoridae.

## *Forensic Entomology*

The main purpose of the dissertation research presented in the following chapters was to improve the precision and accuracy of PMI estimates made with entomological evidence, through the application of molecular and quantitative genetic methods.

Forensic entomology is a powerful tool that can aid in estimating a minimum PMI during death investigations (Catts and Haskell 1990; Greenberg and Kunich 2002) and has long enabled investigators to use tables that outline the minimum development times of each immature stage (e.g. Kamal 1958) to estimate the ages of flies associated with remains. Blow fly development is dependent on temperature (e.g. Anderson 2000, Grassberger and Reiter 2001), so if investigators know the developmental stage of the oldest flies collected as evidence and have historical weather data for the scene, they can determine the window of time necessary for a species to reach that stage. The time since insect colonization of remains is generally assumed to be the minimum PMI.

The standard method for predicting a PMI from insect evidence has remained relatively unchanged for decades (Catts and Haskell 1990). The static nature of the approach is due, in part, to the general success of the method, which has been upheld numerous times in American and international courts (Greenberg and Kunich 2002). However, its continued acceptance has resulted in the persistence of a number of caveats associated with PMI predictions based on fly evidence. One of these is that, since each developmental stage gets progressively longer through fly development, a PMI prediction obtained from later stages will necessarily incorporate a much larger window of time. For example, Kamal (1958) found that, for the species *L. sericata* growing at 26.7°C, the second larval instar lasted 9–26 hours while the longest stage, pupation, was 5–11 days.

At lower temperatures pupation can be even longer (Anderson 2000; Grassberger and Reiter 2001), and age/PMI estimates must be correspondingly broad.

One method for generating a more precise age estimate within developmental stages is to include body size in the PMI prediction process. As blow fly larvae feed, they increase in size in a generally linear fashion, with relatively little variance (Wells and Kurahashi 1994, Grassberger and Reiter 2001; Greenberg and Kunich 2002). Thus, during the feeding stages linear regression can be used to refine age estimates. However, the approach highlights the second caveat: larvae shrink when feeding ceases during the third instar (Wells and Kurahashi 1994, Grassberger and Reiter 2001; Greenberg and Kunich 2002) and exhibit a larger variance in body size than previous stages. Additionally, pupae do not change in size as they age. These facts mean that the last two (and longest) developmental stages provide relatively imprecise (though generally accurate) PMI estimates, resulting from their long durations, variance in body size, or unchanging body size.

The last caveat associated with the status quo approach for predicting blow fly age is that it can sometimes be difficult to distinguish between feeding and postfeeding third instar larvae. The determination of a third instar state is made by observing the visibility of tissue in the crop (Figure 1.1) and the cessation of feeding, which can be difficult in some cases (reviewed by Anderson 2000). If investigators do not note the feeding behavior of larvae when collected from a body, the only physical evidence a forensic entomologist will have to work with is the visibility of tissue in the crop. Several factors, including starvation of larvae and the leakage of crop contents into storage solution over time can make distinguishing between the feeding and postfeeding

stages of the third instar difficult or impossible if crop contents alone are relied upon (Anderson 2000). The inability to identify the developmental state of a third instar larva will result in a larger window associated with an age estimate, as both stages must be included in the PMI estimate.

The shortcomings of predicting blow fly age based on developmental stage and body size may be addressed by including new information in the PMI prediction process. With the arrival of the genomic age biologists have developed new tools that enable them to assay gene expression levels at relatively low cost. From these, a detailed understanding of gene regulation has emerged (reviewed by Kalthoff 2001). As eukaryotes (including blow flies) develop, a variety of proteins are required at various times, thus the cellular transcriptional machinery initiates the expression of RNA from many genes in a temporally variable pattern. Given the highly specific temporal and spatial requirements of gene expression during development, the expression levels of these genes may be useful predictors of age as they are up- or down-regulated in a reliable fashion, and thus have the potential to aid forensic entomologists in more accurately estimating blow fly age.

The developmental similarities among Cyclorrhaphan flies mean that genes known to vary throughout the development of *D. melanogaster* are excellent *a priori* candidates for study in the context of forensic entomology. Two recent genomic studies in *Drosophila* (Arbeitman et al. 2002 and Beckstead et al. 2006) demonstrated that thousands of genes have predictable and temporally variable gene expression. Of these, there are myriad gene expression patterns that can be used to indicate different points in development. For example, in *D. melanogaster*, the gene *Amalgam* is expressed at its

highest levels during early pupation, while CG17814 is expressed at its highest levels during late pupation (Arbeitman et al. 2002). Knowing the expression levels of both genes could help distinguish among early, middle, and late pupal development.

Though gene regulation in *D. melanogaster* demonstrates a theoretical capacity to predict blow fly age, gene expression profiles must be produced in a forensically useful blow fly species to enable precise PMI predictions. The species studied in this research was *L. sericata*; chosen because it is a common fly encountered in forensic entomology that is globally distributed (Greenberg 1991). The *Lucilia* genus has also been included in multiple molecular studies, mostly due to the economic effect of *L. cuprina* infestations of Australian sheep (East and Eisemann 1993). This meant that gene sequence information could be obtained from the public domain, or easily sequenced in *L. sericata* by designing PCR primers from *L. cuprina* sequence, thereby limiting the effort spent on acquiring gene sequences. By focusing on *Lucilia* genes, which were known to vary throughout *Lucilia* and/or *Drosophila* development, it was possible to assess the development of *L. sericata* in terms of nine gene expression profiles.

#### *The Use of Developmental Plasticity and Gene Expression for Blow Fly Pest Control*

The expression profiles produced in this research also have agricultural and human health implications. Four of the genes studied (*chitin synthase*, *acetylcholine esterase*, *ecr*, and *usp*) encode proteins that are targets of insecticides (Ware and Whitacre 2004) and alleles of another (*resistance to organophosphate-1*, an alpha carboxylesterase) can affect the toxicity of organophosphates (Newcomb et al. 2005). Any variation in the expression of these genes during development or among strains may



be useful in identifying specific insecticides to use for targeted control of a specific developmental stage or strain. Additionally, certain temperature treatments will likely result in differential susceptibility to insecticides if temperature is shown to change expression profiles of target genes. Knowledge of temperature variation in insecticide target expression can identify an optimal season to apply specific insecticides and help insect control personnel maintain facilities (such as barns) at an optimal temperature to maximize the effectiveness of an insecticide treatment. Armed with data for variation in insecticide-targeted gene expression, it should be possible to develop insect control programs that maximize the effectiveness of insecticides, thus limiting the amount needed to control a pest species.

Knowledge of factors affecting variation in body size may also be useful in limiting the effects of myiasis events on sheep and cattle. *L. cuprina* extracts have been used to elicit an immune response to myiasis that decreases the amount of biomass lost to parasitization by limiting larval size (East and Eisemann 1993), thereby ameliorating economic damages. Other factors that affect body size of agriculturally important blow flies could be used in place of, or in conjunction with, fly extracts to further limit the amount of damage they caused. Additionally, such an approach could be incorporated into an integrated pest management scheme that combines the manipulation of developmental plasticity with insecticide applications to better control pest species.

### *Research*

The dissertation is divided into multiple chapters, each detailing experiments designed to better understand *L. sericata* development through the application of

molecular and quantitative genetic methods. The first set of experiments revealed conditions that affected plasticity in development stemming from variation in protocols used for rearing flies. Four earlier publications contain data on the laboratory growth of *L. sericata* that outline various growth rates for the species (Kamal 1958; Greenberg 1991; Anderson 2000; Grassberger and Reiter 2001). The authors used different rearing procedures and populations of flies, making observed developmental variation among the studies impossible to dissect. Likewise, no author compared growth under laboratory conditions to growth on carrion. Hence a series of experiments was undertaken to determine how the different rearing methods described in the literature affected development, and which combination of rearing conditions best mimicked growth on cadavers.

The second set of experiments was devoted to understanding growth differences among strains and between replicates raised at different temperatures. The approach also enabled the development of a statistical application, generalized additive models (GAMs), to make predictions with non-linear length and weight data, as well as gene expression profiles in future work. Various GAMs, which use likelihood statistics to incorporate multiple non-linear variables into a prediction, were constructed and compared as to their abilities to predict blow fly age (Hastie and Tibshirani 1990, Wood 2006). The models were first tested against each other using stage, length, and weight data from 2559 immature flies. In addition to non-linear body size data, three regional strains, originating from California, Michigan, and West Virginia, were included in the study to determine if regional strains develop similarly. The flies were also raised at two

different temperatures, to understand potential effects of temperature on variation in PMI predictions.

The third set of experiments dealt directly with gene expression during egg development, through the analysis of a modest set of three genes. Eggs represent the shortest developmental stage in flies (~10 hours at 32°C), so they were used in a feasibility study to determine if gene expression levels would be useful in identifying more specific periods of development within a stage. If estimates of age can be made more precise in eggs, it is likely that estimates of longer stages will be more precise as well. Much is known about Dipteran embryogenesis (e.g. McGregor 2005), which indicates that it should be possible to differentiate between young and old eggs, based on the expression levels of various patterning genes. Three genes were assessed during *L. sericata* egg development to demonstrate the principle that gene expression can be used to estimate blow fly ages.

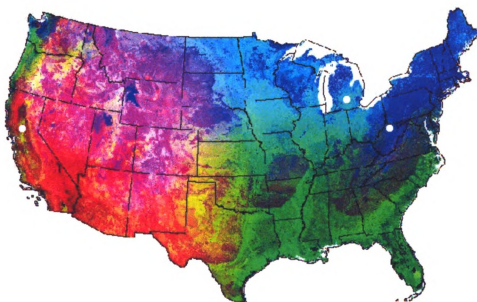
Since the egg expression profiles yielded promising results, the forth and most extensive line of research was conducted. A subset of 958 individuals from the 2559 staged, measured, and weighed larvae and pupae in the GAM experiment were profiled for the expression of 9 developmentally regulated genes (three other genes were removed as they were uninformative). All strains were assessed at 33.5°C, and gene expression in the CA and MI strains was also assessed at 20°C. GAMs were constructed using gene expression levels/size/stage, and compared to GAMs using size/stage to predict age. In the next chapter, a blind study is detailed in which the ages of larvae and pupae were predicted using GAMs developed from the gene expression data. The successful validation of the methodology was a critical part of the research, because the results of

any statistical modeling must be confirmed using independent data (Scheiner and Gurevitch 2001).

During the aforementioned GAM research it was noted that some individuals in all replicates failed to form a puparium, even as adults were eclosing from that replicate. Forensically, the collection of non-maturing (Peter Pan) individuals at a crime scene could be very misleading, resulting in inaccurate PMI estimates. However, the ability to recognize such individuals might be useful for identifying flies that should not be included in a PMI analysis, as they would skew results, and may also indicate that pupae were missed in sampling. Individuals in a state of arrested development are also an important focus of research as they could be diapausing (Tachibana and Goto 2004), estivating (Liu et al. 2006), or developmental mutants. Molecular information identifying any gene that may be involved in delaying development is useful for studying plasticity and may also be used to control pest species, as described above. Given the potential benefits of studying this state, ten (or as many as available) Peter Pan individuals were collected from each replicate. Of ~120 flies, 55 were profiled to determine if they exhibited differential expression patterns as compared to normal postfeeding third instars.

The last project undertaken was a comparison of newly collected strains of *L. sericata*, to confirm preliminary observations from the statistical modeling research that showed temperature and strain effects on development time and body size. From the California, Michigan, and West Virginia flies, 720 pupal length and weight measurements, and 60 minimum development times were assessed to determine how biological strain and temperature affected these traits. Genetic effects on development time or body size may be indicators of differential selection among the strains, possibly

driven by environmental factors in the ecoregions that each strain originated from (Figure 1.3).



**Figure 1.3. Ecoregions of the United States.** Map from work by Hargrove and Hoffman (2004). The white dots indicate approximate origins of the CA, MI, and WV flies studied. Regions of similar color have comparable elevations, soil, and climates. Both the MI and WV strains were located in ecosystems that were mapped in shades of blue, while the CA strain came from a yellow region.

Each section in this dissertation revealed information regarding one or more aspects of developmental variation in the blow fly *L. sericata*. Any improved understanding of development time (the main phenotype of interest in forensic entomology) and body size (a phenotype of secondary interest) will be useful in achieving more accurate and precise PMI predictions with entomological evidence. Assessing the environmental and genetic effects on body size and development time can also help to determine the conditions that should be followed in standard operating procedures for rearing flies in the laboratory. Likewise, genetic and ecological effects on development can be used to tailor a PMI estimate to the regional strain involved in an

investigation. Most importantly, gene expression data provide a new form of information, independent from body size and developmental stage, which can improve the precision of age estimates for later developmental stages. The use of GAMs to achieve this goal allows for improved descriptions of error that will satisfy the Daubert standard of scientific evidence, ensuring the place of forensic entomology in the courtroom. Through the application of quantitative and molecular genetic approaches, a more precise and accurate method of PMI prediction should emerge, while producing data of value to agricultural, medical, and evolution research.

## CHAPTER 2: PLASTICITY IN BLOW FLY GROWTH

### *Introduction*

Forensic entomologists rely on published data of blow fly development to estimate the time since initial colonization of remains, thus extrapolating a postmortem interval (PMI) (Catts and Haskell 1990). PMI estimates based on entomological evidence have been widely and successfully presented in legal proceedings, however the laboratory study of blow fly development, on which these estimates are founded, has never been standardized. Because of this, entomologists may utilize different blow fly developmental data sets, which can lead to variable PMI predictions. Further, a lack of scientific standardization has the potential to call into question the overall accuracy of entomological evidence (see Saks and Koehler 2005).

Prominent examples of differing laboratory rearing methods and resultant data sets can be found for the widely distributed green blow fly, *L. sericata* (Diptera: Calliphoridae) (Meigen) (Kamal 1958, Greenberg 1991, Anderson 2000, Grassberger and Reiter 2001). These data sets all present a developmental time scale from egg to adult. In his work, Kamal (1958) recorded only the duration of each developmental stage, while Grassberger and Reiter (2001) and Greenberg (1991) also measured the length of maggots until pupation, and Anderson (2000) measured crop length throughout development. Each of these studies utilized different fly-rearing techniques, varying in the quality and type of food, the quality of pupation substrate, and the destructiveness of sampling. Likewise, the authors measured fly development at different temperatures, and reported development data in assorted ways (minimum, average minimum, mode, and

maximum growth). The resulting picture of *L. sericata* development is clouded, with relatively small differences in minimum development time among all studies, while Anderson (2000) characterized a notably longer minimum development time at temperatures similar to the others. Unfortunately, direct comparison of these studies is impossible, as experimental conditions and genetic background of the flies varied among them. Further, even though the data sets were generated with a goal of relating larval development to PMI estimates on corpses, no attempt was made to tie laboratory-established growth rate data to ecologically relevant larval development on carrion.

Development time is a quantitative trait that is expected to vary due to both genetic and environmental factors (Mackay 2001; Conner and Hartl 2004). Understanding genetic and environmental effects on quantitative traits is best accomplished by altering one variable while keeping all others constant, and a limited number of such experiments have been conducted in a forensic entomological context. For instance, Kaneshrajah and Turner (2004) demonstrated that *C. vicina* reared under otherwise constant conditions showed variable growth when raised on different organs, and Wells and Kurahashi (1994) indicated that differences in rearing protocols were the likely source of discrepancies regarding development times of *Chrysomya megacephala* (Diptera: Calliphoridae). Likewise, high-density rearing conditions that increase maggot mass temperatures were shown to shorten development times of *C. megacephala* (Goodbrod and Goff 1990). Recently, *L. sericata* was found to exhibit variable growth patterns depending on the species and tissue type on which cohorts were raised (Clark et al. 2006). Certainly it appears that rearing conditions can have a major impact on the developmental timing of calliphorids.



Just as environmental factors influence calliphorid development, intra-specific differences have the potential to produce variation in fly developmental times. The field of ecological genetics is replete with cases demonstrating the effects of genetic background on quantitative traits (reviewed by Mackay 2001; Conner and Hartl 2004). Developmental variability has been documented in many fly species, including strains of *Drosophila* (Diptera: Drosophilidae) (Johnson and Schaffer 1973, Oudman et al. 1991, Hoffmann and Harshman 1999, Parsch et al. 2000), *Rhagoletis pomonella* (Diptera: Tephritidae) (Feder et al. 2003), and *Scathophaga stercoraria* (Diptera: Scathophagidae) (Blanckenhorn 2002). Since each *L. sericata* study referenced above was conducted on different populations, it is impossible to separate the effects of environment and genetics on fly development. Potentially, any (perhaps all) differences among *L. sericata* studies could be explained by genetic variation among strains, however this would only be demonstrated if each strain was raised using the same experimental protocol. Unless standard rearing conditions are adopted, such comparisons are impossible.

The potential influence of the environment and genetics on quantitative traits, and in particular development time, led to the hypotheses tested herein that *L. sericata* growth is plastic with respect to rearing conditions, and that fly development on carrion will best be predicted by a specific combination of laboratory conditions that affect this plasticity. Temperature and humidity are already known to affect development time (Greenberg 1991, Anderson 2000, Grassberger and Reiter 2001) and mortality (Wall et al. 2001) in this species, so these conditions were held constant to investigate the effects of other rearing variables. Likewise, the flies in these experiments originated from the same source population, allowing genetic differences to be largely ruled out as a source of

developmental variation. By changing the exposure of a single strain of *L. sericata* to specific environmental conditions, several questions related to the hypotheses were addressed. In particular: 1) Do laboratory rearing conditions affect the development time of *L. sericata*? 2) Are any developmental differences caused by laboratory rearing conditions large enough to explain the variation observed among published growth data on this species? And 3) Does growth generated under laboratory conditions accurately reflect larval development of *L. sericata* on a carcass?

### *Materials and Methods*

#### **Fly Collection and Rearing.**

*L. sericata* adults were collected from the Michigan State University campus in East Lansing, Michigan throughout the spring, summer, and fall of 2004, and were used to establish a general population cage of approximately 200 flies. Species identification was done using multiple keys, two independent identifications, and by comparing the DNA sequence of a 798 base pair region of the mitochondrial *cytochrome oxidase 1* gene to published sequences on the NCBI website using the BLAST link. Forward primers for DNA amplification were GATCAGTAGTAATTACAGCT, and TAATATTGCTCATGGAGGAG, while reverse primers were TTGACTTTTAAATATCTTAG, and CCTAAGAAATGTTGAGGGAAG. Polymerase chain reactions were run for 35 cycles by denaturing at 95°C for 30 seconds, annealing primers at 50°C for 30 seconds, and extending amplicons for one minute at 72°C. Sequences were generated on a CEQ 8000 capillary electrophoresis system, using a CEQ

DTCS Quick Start Kit and the manufacturer's suggested protocols (Beckman Coulter, Fullerton, CA).

Experimental rearings were carried out between January and March of 2005. To minimize the loss of genetic variation during this period, the population was expanded to three cages of more than 100 individuals, from which 20–50 migrants were transferred as pupae to the other cages each generation. Generations were allowed to overlap until the cage required cleaning, which was done monthly while the next generation was in the juvenile form.

Cages of adult flies were provided water and honey *ad libitum*. Beef liver was supplied as a protein source one day prior to oviposition. On days that eggs were collected, fresh liver was placed into a cage in the late morning to mid afternoon. Cages were checked every 15–30 min until oviposition was observed. Approximately 250–1000 eggs (1–4 egg masses) were removed one hour after the first observation of oviposition. The egg masses were immediately transferred to fresh liver and placed into a 1-liter glass canning jar (Alltrista, Muncie, IN), with a breathable cloth screwed on as a lid. Jars were placed into a temperature-controlled incubator at 25°C (+/-0.5°C) with a 12:12 hour light and dark cycle. A beaker filled with water was kept in the incubator, which provided a relative humidity of 25% (+/- 4%).

Several treatments were examined to assess the influence of rearing variables on the development time of specific immature stages, and on total immature development time (Table 2.1). These considered the freshness of food, moisture of food, type of pupation substrate used, orientation of the substrate with respect to food, transfer of larvae to fresh pupation substrate, and destructiveness of sampling. The influence of

meat freshness was tested by providing cohorts with 40g of liver every day (fresh meat daily or FMD) or 120g of liver every third day (no fresh meat daily or NFMD). Paper towel treatments received fresh meat daily, which was placed on a moist paper towel (FMDPT). The influence of pupation substrate was examined by providing either clean sand (Fairmount, Wedron, IL) or vermiculite (Therm-O-Rock West, Chandler, AZ) to jars containing postfeeding third instars. The influence of food orientation with respect to pupation substrate was tested by either placing meat on top of the substrate at the egg stage, or placing the substrate on top of meat when larvae reached the postfeeding third instar stage. Fresh pupation substrate was tested by removing 125 postfeeding third instars from individual cohorts and placing them into a jar with 500ml of fresh pupation substrate. The transfer treatments were taken from cohorts with far more than 300 individuals in the jar, meaning larval density was much greater in untransferred than transferred treatments. Destructive sampling was assessed by permanently removing or not removing 12 individuals from a cohort each day.

Experimental cohorts were checked approximately every twenty-four hours, except jars with eggs, which were checked every half hour until they hatched, and pupae, which were observed throughout the day until eclosion occurred. Length measurements were taken throughout larval development, incorporating the 12 most mature larvae in all treatment groups (either the largest maggots or postfeeding maggots lacking blood in their crops). Ruler-measured lengths of the maximum body extension (to the nearest 0.5 mm) were determined using a stereomicroscope for first instars (due to their small size) or by eye for all other stages. Advances in developmental stage were recorded to the closest 15 minutes, however given that most animals were observed once per day,

development time variation of less than one day was indistinguishable from sampling time variation. All experiments were conducted in the same temperature controlled incubator, with jars rotated within the incubator daily.

Development of larvae on mammal carcasses was performed using three Sprague-Dawley rats from breeding colonies at MSU, sacrificed by CO<sub>2</sub> asphyxiation within two days of egg placement on the body. The rats weighed approximately 500g and were in excess of the feeding needs of individual cohorts (larvae utilized approximately half of the carrion before the postfeeding stage). An egg mass collected in the manner detailed above was placed along the mouth of the rat. Rat carcasses were set in an open plastic bag, which was placed into a styrofoam container with an opening cut from the lid. A screen was fitted between the container and the lid to prevent escape of postfeeding larvae. Animals were reared at 25°C ( $\pm 0.5^\circ\text{C}$ ) and 25% ( $\pm 4\%$ ) relative humidity, with maggot length and the duration of developmental stages recorded as above. Larvae from rat treatments were transferred to sand substrates to pupate.

### **Statistical Analyses.**

Owing to unbalanced data (Table 2.1), MANOVA could not be used, thus analyses of variance were examined using Type III ANOVAs (Scheiner and Gurevitch 2001). This approach removes the variance from variables other than the one of interest, and compares the variance remaining to the dependent variable. ANOVA and regression statistics were performed with the R statistical package (R Development Core Team 2004) at  $\alpha < 0.05$  significance.

Development times in hours and accumulated degree-days (ADD), including standard deviations, were calculated for every significant treatment type and for rat cohorts. ADD was calculated using a base temperature of 10°C.

Graphs of larval growth were produced using the R statistical package. Curves were plotted by non-linear quantile regression using smoothing parameters that yielded curves comparable to published data from Greenberg (1991), Wells and Kurahashi (1994), and Grassberger and Reiter (2001). Treatments in the comparisons include FMD cohorts that were transferred to fresh pupation substrate, FMDPT cohorts that were transferred to fresh pupation substrate, and NFMD cohorts that were not transferred to new pupation substrate. The plots included average and 95% confidence intervals, from the day flies hatched until the first day pupae appeared, which were then compared to averages of larval growth on rats. Data from Grassberger and Reiter (2001) were also compared to larval development on rats, as that study included growth at 25°C. For these analyses a locally weighted sum of squares (lowess) curve was plotted through the estimates using R.

## *Results*

### **Species Identification.**

Morphological identification of flies indicated that all were *L. sericata*. To confirm identification, a 798 base pair mitochondrial *cytochrome oxidase 1* sequence (NCBI accession number DQ062660) was obtained from a collected adult fly. A BLAST search showed it was identical to a *cytochrome oxidase 1* sequence from a *L. sericata* population in Ontario, Canada (accession number L14947). The closest 13 NCBI gene

sequences were from *L. sericata*, with a maximum difference of 4 base pairs (<1%), confirming the species identification.

### **Developmental Plasticity.**

The pre-pupation period for this fly population (reared at 25°C) ranged from 145–264.5 hours (6–11 days), while the duration of egg to adult was 329–505.5 hours (14–21 days), with all data given in Appendix 2.1. Throughout the experiment replicate treatments followed synchronized growth trajectories during the feeding stages, with a small number of individuals lagging. In contrast, postfeeding larvae within a treatment advanced to pupation gradually over a week. Eclosion took place over a week also.

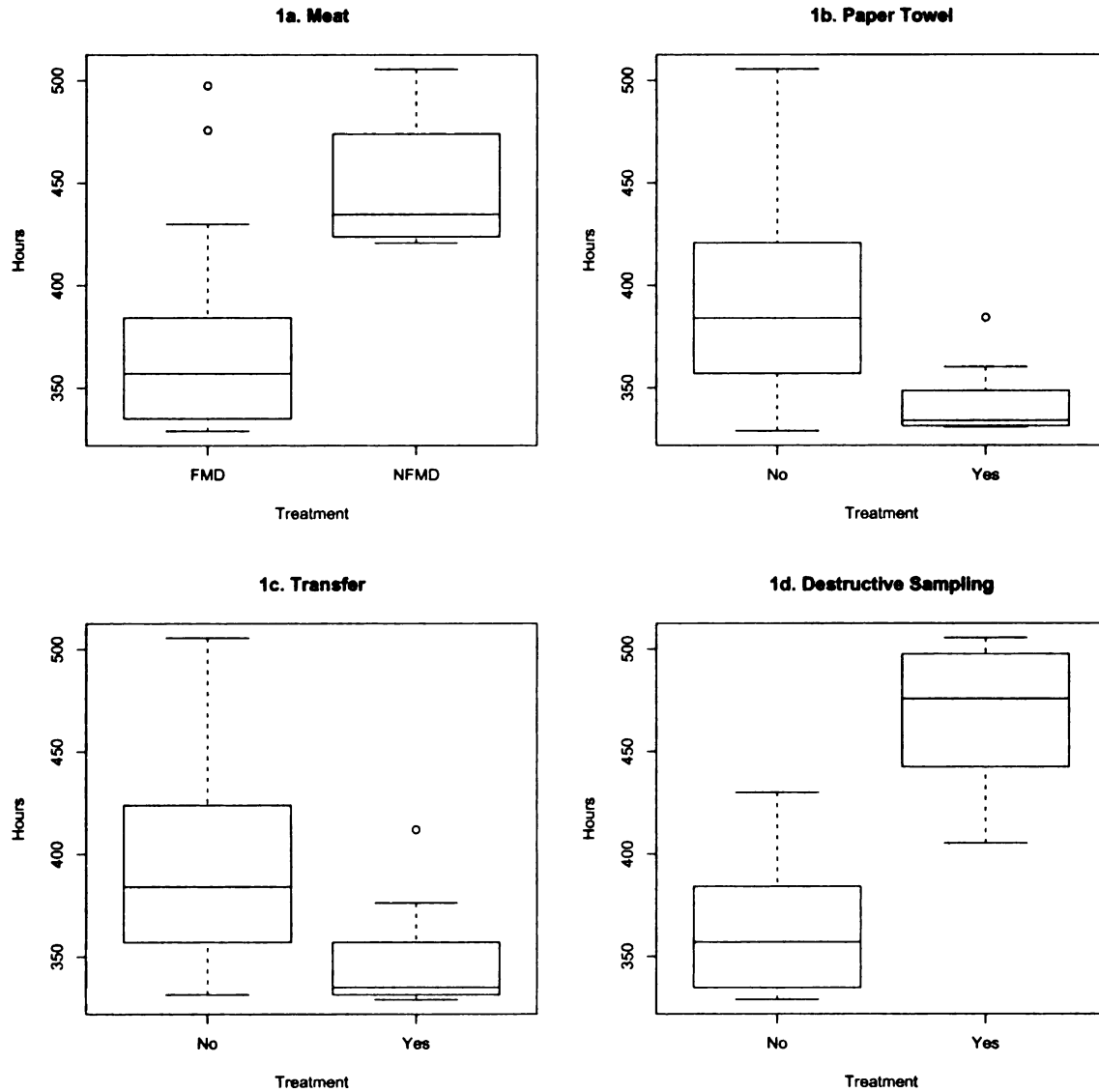
Development times for stages and treatments are given in the Appendix and are summarized in Table 2.2 (using both hours and ADD). Linear models showed that development among treatments did not exhibit statistical differences in the shortest stages—the egg or the first two instars (a single exception is detailed below)—nor did these stages significantly influence total development time (data not shown). In contrast, the feeding portion of the third instar ( $F=18.52$ ,  $df=1$ ,  $P=0.00013$ ,  $R^2=0.35$ ), the postfeeding stage of the third instar ( $F=27.67$ ,  $df=1$ ,  $P<0.0001$ ,  $R^2=0.44$ ), and pupation ( $F=53.59$ ,  $df=1$ ,  $P<0.0001$ ,  $R^2=0.62$ ) significantly affected overall development times.

Substrate type and its placement had no significant effect on development during any stage. Other treatments examined (Table 2.1) significantly impacted development time (Figure 2.1 and Table 2.2), while the stage at which that impact occurred differed. FMD accelerated development compared to treatments that received supplements every third day during the feeding portion of the third instar ( $F=12.19$ ,  $df=1$ ,  $P=0.0015$ ), although it was also a significant variable in the duration of the second instar ( $F=8.336$ ,

df=1,  $P=0.0072$ ). Accordingly, the two treatment types that developed in 14–16 days were FMD. FMDPT also resulted in faster growth during the feeding portion of the lifecycle compared to treatments without paper towels ( $F=206.8$ , df=1,  $P<0.0001$ ). Moist paper towels were not necessary for the most rapid overall development, given that the fastest recorded time from egg to eclosion was from a FMD transferred treatment [329 hours (cohort 14 in Appendix 2.1)], however they promoted consistently faster development (Figure 2.1). Once feeding ceased, the moisture of food did not contribute to developmental variation (postfeeding third instar  $F=0.8439$ , df=1,  $P=0.37$  for FMD and  $F=1.677$ , df=1,  $P=0.21$  for FMDPT), however transferring larvae to fresh substrate significantly shortened the amount of time spent as postfeeding third instar larvae ( $F=17.59$ , df=1,  $P=0.00022$ ). The results indicate that handling larvae during the study did not impede development.

Destructive sampling did not influence larval stages, but significantly increased the pupal stage ( $F=49.13$ , df=1,  $P<0.0001$ ). Finally, variables were assessed together to determine their relative influence on total immature development. Each had significant effects on total development time (FMD:  $F=4.644$ , df=1,  $P=0.039$ ; FMDPT:  $F=8.019$ , df=1,  $P=0.0079$ ; Transfer to fresh substrate:  $F=4.454$ , df=1,  $P=0.043$ ; Destructive sampling:  $F=26.14$ , df=1,  $P<0.0001$ ).





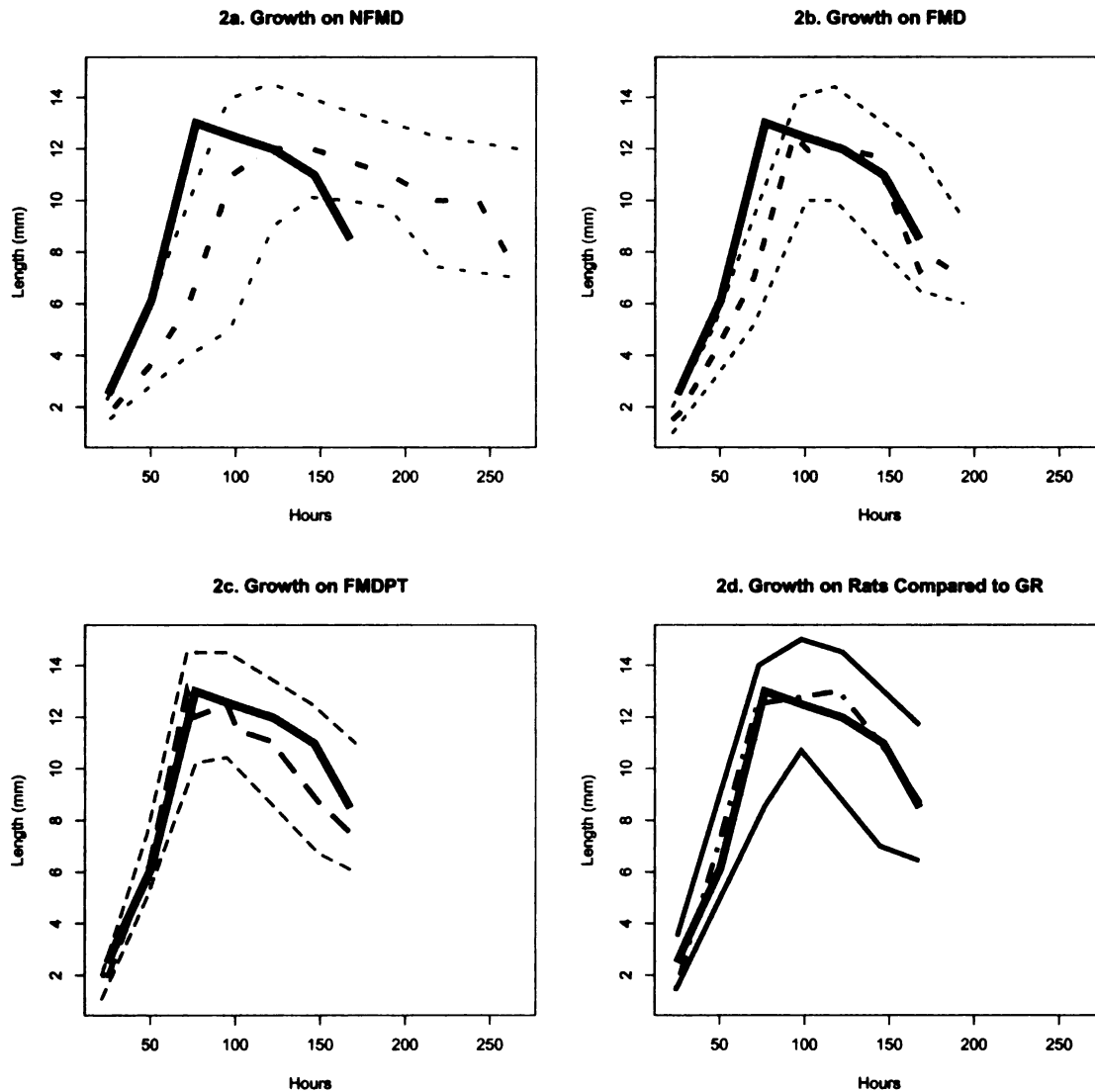
**Figure 2.1. Developmental variation among *Lucilia sericata* cohorts by treatment type.** Boxplots of total development time (hours) for each of the 37 liver-fed cohorts. The line within the box represents the median development hours, the box represents the development times between the 25<sup>th</sup> and 75<sup>th</sup> percentiles, and the 'whiskers' (outer-most lines) represent the 5<sup>th</sup> and 95<sup>th</sup> percentiles. 1a: fresh meat daily or no fresh meat daily (FMD vs. NFMD); 1b: paper towel (moist paper towel placed under meat); 1c: transfer: transfer of larvae to fresh substrate for pupation; 1d: destructive (removal of 12 individuals each day). Note: treatments were in combination with other treatment types (Table 1) that had significant effects on development time. For instance, the two outliers in the FMD boxplot (1a) are those that were also destructively sampled.

### **Development on Carrion.**

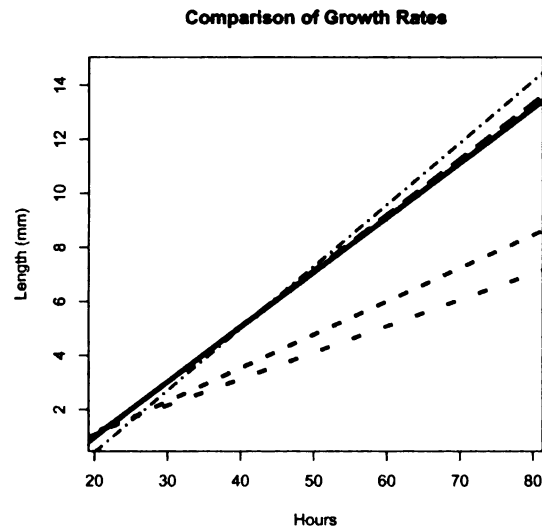
The pre-pupal growth of larvae on rats was compared to the statistically significant experimental treatments, as well as growth observed by Grassberger and Reiter (2001) (Figures 2.2 and 2.3). The results displayed in Figure 2.2 show that the shape and rate of larval growth curves for FMDPT treatments most closely matched the three cohorts reared on rats.

Figure 2.3 displays the growth of larvae during the first three days of development, when growth rate is relatively constant. A linear regression demonstrated different rates of growth among treatments, which were 0.20 mm/hr, 0.10 mm/hr, 0.12 mm/hr, 0.21 mm/hr, and 0.23 mm/hr, for Rat, NFMD, FMD, FMDPT, and Grassberger and Reiter (2001) respectively, with  $R^2$  values of 0.92, 0.77, 0.90, 0.95, and 0.99. The regression model showed that length varied significantly with age ( $F=7099$ ,  $df=1$ ,  $P<0.0001$ ), while the effect of treatment types on length was also statistically significant ( $F=281.8$ ,  $df=4$ ,  $P<0.0001$ ), as was the interaction between age and treatment type ( $F=155.0$ ,  $df=4$ ,  $P<0.0001$ ).

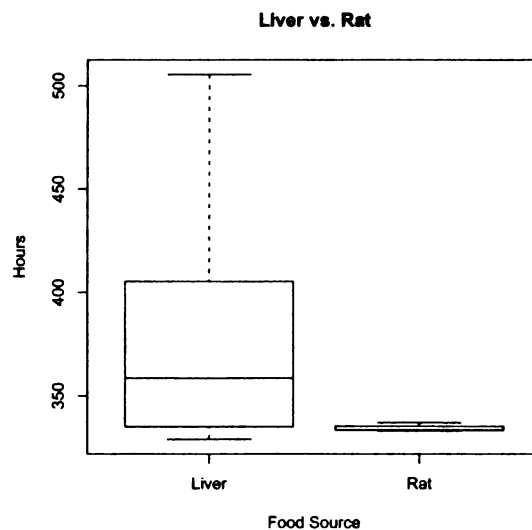
Figure 2.4 compares the development of the flies reared on rats to development of liver-fed treatments in this study. Cohorts on rats developed in a manner that was most similar to the observed maximal development of liver-reared flies (i.e., FMDPT and some FMD treatments), with development times between 333 and 337 hours (about 14 days). Further, growth on rat carcasses was much less variable than the growth of liver treatments.



**Figure 2.2. Growth curves of *Lucilia sericata* on liver versus growth on rat carrion.** Non-linear quantile regression curves created from the lengths of maggots in daily collections of each treatment type. 2a: meat added every 3<sup>rd</sup> day, no moist paper towel, larvae were not transferred to fresh substrate to pupate; 2b: fresh meat daily, no moist paper towels, larvae transferred fresh substrate to pupate; 2c: fresh meat daily, moist paper towel used, larvae transferred to fresh substrate to pupate; 2d: the locally weighted sum of squares curve of data estimated from Grassberger and Reiter (2001) plotted against larval growth on rat carrion. Numbers of cohorts plotted for each treatment were 3, 4, 6, and 6 for Rat, NFMD, FMD, and FMDPT respectively. The solid line on each curve is the 50<sup>th</sup> percentile plot from cohorts raised on rats. Treatments are shown as dashed lines, with the thicker dashed line representing the 50<sup>th</sup> percentile and the thinner lines representing the 97.5<sup>th</sup> and 2.5<sup>th</sup> percentiles (95% confidence intervals). Confidence intervals for the rat cohorts are present in 2d.



**Figure 2.3. Linear growth of *Lucilia sericata* on liver versus growth on rats.** Regression lines of the same treatments displayed in Figure 2.2, for the first three days of growth—the linear phase of development. Line types used to indicate treatments are the same as in Figure 2.2.



**Figure 2.4. Development times of *Lucilia sericata* cohorts raised on liver versus rat carrion.** Comparison of total development hours produced by the 37 liver-fed cohorts in this study to the development of the three rat-fed cohorts. Development time on rats was much less variable than growth on liver, with a development time most similar to the fastest growing liver-fed cohorts. Boxplot design is as in Figure 2.1.

## *Discussion*

### **Environmental Components of Variation in the Development of *L. sericata*.**

The green blow fly is a widely distributed species of great forensic importance. Several authors have examined different fly populations reared under various environmental conditions, and perhaps not surprisingly, the developmental times differ from one another, with Kamal (1958), Greenberg (1991), and Grassberger and Reiter (2001) estimating faster minimum development times than Anderson (2000). This variability could result from genetic differences among populations, but could equally result from dissimilarity in the conditions under which the animals were reared. Further, none of the authors compared the laboratory growth of flies to that on actual carcasses. In the current study, designed to estimate variation in developmental rates resulting from environmental differences, a single population of *L. sericata* was grown under laboratory conditions that mimicked those used in the earlier studies, and these treatment were compared to larval development on carrion.

Given the minimum development times of the treatments detailed here, the fastest fell within the standard errors for *L. sericata* reared at 22°C by Greenberg (1991) and Grassberger and Reiter (2001), and is close to the mode reported by Kamal (1958), which is a common forensic entomology resource. Likewise, the slowest minimum development time for flies in this study was longer than the developmental minimum at 23.3° C found by Anderson (2000). This indicates that environmental variation alone can potentially explain all differences in developmental rates detailed in earlier studies.

Results of these experiments demonstrate that variation in food moisture and pupation substrate have a significant influence on the growth of *L. sericata*; variation in rearing conditions generated a developmental difference of up to 7.4 days. Most notably, treatments designed to maintain meat moisture during feeding shortened the development of larvae. FMDPT treatments significantly shortened the feeding portion of third instar larvae, and produced a much smaller developmental range (Figures 2.1 and 2.2). These results accentuate the importance of considering food moisture when rearing fly larvae. Grassberger and Reiter (2001) provided larvae with fresh liver daily, resulting in a similar growth rate at 25°C. Other studies have included moist sawdust, paper towels, or wood chips underneath meat (Kamal 1958, Goodbrod and Goff 1990, Anderson 2000), which would be expected to hold moisture. Interestingly, moist paper towels changed the lifehistory table of FMD treatments toward the Greenberg (1991) estimate of third instar duration, which is approximately one day shorter than that of Grassberger and Reiter (2001). Unfortunately, Greenberg's (1991) report was vague about how flies were raised so it is unclear what other factors could be involved, but food moisture may play a role in the differences in third instar development time observations between these authors.

Transferring postfeeding larvae to a fresh substrate for pupation significantly shortened the time spent at this larval stage. The postfeeding portion of blow fly larval development is generally variable (Wells and Kurahashi 1994) and *L. sericata* is exceptional among blow flies for wandering far from its food to pupate (Anderson 2000). This may mean that *L. sericata* searches for a specific set of environmental cues for pupation, making the postfeeding stage susceptible to disturbance. The conditions that produced the fastest growth in this study yielded a postfeeding stage duration of two to

three days. Kamal (1958) provided sawdust with food, and observed a mode postfeeding duration of 90 hours at 26.7°C, with a minimum of 48 hours and a maximum of 192 hours. His mode observation is similar to untransferred treatments in this study, which lasted a day longer than transferred cohorts. Greenberg (1991) reported an average postfeeding time of 108 hours at 22°C while Grassberger and Reiter (2001) reported 94 hours at 20°C and 87 hours at 25°C (the temperature at which this research was conducted). With little information on rearing conditions described by Greenberg (1991), the shorter times reported in the latter study are hard to explain, but Grassberger and Reiter (2001) reared their flies with dry sawdust in jars, which may have resulted in the shorter average duration, given that the treatment seems similar to the transfer treatments in the research presented here.

There is little information in the literature that helps explain the developmental variability between transferred and untransferred postfeeding larvae found in the current study. Three plausible explanations for this phenomenon are density of individuals in each cohort, moisture differences between old and fresh substrates, and difference in odor between the treatments. Larval density seems unlikely to have had much influence on development time. Several treatments that were transferred to sand had larvae that had congregated on the substrate surface, and these densely packed cohorts still pupated in a timely manner. On the other hand, a lack of moisture and odor are both plausible agents behind the accelerated onset of pupation in transferred larvae. The sensitivity of larvae to moisture during feeding (outlined above) indicates that moisture is a potential cue for the cessation of feeding, with maggots actively searching out wet areas (tissues) while feeding, and reversing this behavior when heading towards pupation. Likewise, blow

flies are attracted to odors associated with decay (Catts and Haskell 1990, Chaudhury et al. 2002, Hall et al. 2003), thus it might be advantageous to be attracted to putrefying odors during feeding, followed by a pre-pupation move away from such odors.

Destructive sampling was found to be unimportant in larval development, yet was the only significant variable affecting the duration of pupation. The delay in pupation most likely resulted from the elimination of the earliest individuals to form a puparium, which were destructively sampled (removed) by necessity. Given these findings, studies of pupal development rates that require destructive sampling should consider its effects.

#### **Other Potential Sources of Variation.**

The data presented demonstrate the effects of differing rearing treatments on this population of *L. sericata*. It should be noted however, while most variation in growth existed among treatments, within-treatment variation was also observed. A portion of this could be explained by unmeasured environmental factors, as only a small number of rearing modifications were tested. Certainly, factors not considered in this study are likely to impact developmental differences.

Likewise, though environmental conditions were found to be highly significant in the development of *L. sericata*, genetic variation among fly populations used in different studies could potentially be just as important in understanding developmental variability. It is necessary to remember that each publication mentioned above outlined the development of flies that originated from a different ecogeographical region. There is precedence for population effects on the development of blow flies and several related species (Johnson and Schaffer 1973, Greenberg 1991, Oudman et al. 1991, Hoffmann and Harshman 1999, Parsch et al. 2000, Blanckenhorn 2002, Ames and Turner 2003, Feder et



al. 2003). Genetic makeup is likely to affect other populations of blow flies, although these have been largely untested. Genetic differences, including potential interactions between genotype and environment, may be important sources of developmental variation when comparing populations of *L. sericata*.

### **Optimal Rearing Condition Using Liver and Growth on Rat Carrion.**

One might expect that blow flies have evolved to develop fastest under natural conditions of carrion decomposition. If this is the case, the fastest growth rate obtained in laboratory rearings would be expected to mimic the growth of flies living on carrion at the same temperature. In the current study, *L. sericata* development on rat carcasses was most similar to flies reared under high moisture conditions (Figure 2.4). This finding helps address concerns raised by Kaneshraiah and Turner (2004) and Clark et al. (2006) who observed a significant effect of tissue type on the growth of *C. vicina* and *L. sericata*, respectively. Kaneshraiah and Turner (2004) were critical of rearing flies on liver, as it seemed to delay development. This delay was similar to slower developing treatments observed on desiccated liver in the current study, suggesting that larval rearing should take place on non-desiccated substrates to best mimic growth on a corpse.

### **Applications to Forensic Entomology.**

*L. sericata* development is plastic, at a level that alone could explain differences in the species' published developmental times. This finding highlights two important factors that need to be considered when estimating a PMI based on blow fly development. First, the discrepancies among development data sets can potentially be explained, *in toto*, by differences in laboratory rearing protocols used to develop such timetables. Accordingly, establishment of a common set of rearing conditions, which

best relate to growth on carrion, is critical if direct comparisons are to be made among datasets, and if these data sets are to be used in legal proceedings. Second, because forensic entomologists use a quantitative trait (development rate) and decomposition ecology to make PMI estimates, researchers conducting studies on development time must aim to address the effects of both genetics and environment on their findings. By doing so, the forensic community can achieve a greater understanding of how important each of these factors is to forensic entomology.

A final consideration regarding entomological evidence involves its legal use in general. In the wake of judicial decisions that place a far greater emphasis on systematic analyses, known error rates, and statistical probabilities (see *Daubert v. Merrell Dow Pharmaceuticals*, 509 US 579 (1993), and *KumhoTire Co. v. Carmichael*, 526 US 137 (1999)), forensic scientists are under increasing pressure to conduct research, present legal analyses, and draw conclusions in a methodical and scientifically replicable way, while relying less on generalized knowledge and personal experience. The field of forensic entomology, although based on sound scientific principles, can currently be included among an assemblage of forensic disciplines that may be called into question with regards to repeatability and standardized techniques (see Saks and Koehler, 2005). Efforts to establish calliphorid laboratory rearing protocols that best portray fly development on cadavers, and to standardize those techniques for future research, are central to meeting the demands of *Daubert* and *Kumho*. Such endeavors are necessary if forensic entomological evidence is to be routinely accepted in courts of law.

**Table 2.1 Treatment types for the 37 liver-fed cohorts**

<u>Meat</u>	<u>Destructive</u>	<u>Transfer</u>	<u>Substrate Under</u>	<u>Substrate</u>	<u>Paper Towel</u>
FMD: 33	No: 32	No: 24	No: 33	Sand: 14	No: 25
NFMD: 4	Yes: 5	Yes: 13	Yes: 4	Vermiculite: 23	Yes: 12

Meat: Fresh meat daily (FMD) or not (NFMD). Destructive (sampling): 12 individuals removed from the cohort at each sampling time. Transfer: 125 postfeeding third instars were transferred to 500 mL of fresh pupation substrate. Substrate Under: food was placed on top of a substrate or at the bottom of an empty jar. Substrate: pupation substrate used. Paper Towel: received FMD placed on a moist paper towel.

**Table 2.1 Treatment types for the 37 liver-fed cohorts**

Meat	Destructive		Transfer		Substrate Under		Substrate		Paper Towel	
FMD: 33	No: 32	Yes: 5	No: 24	Yes: 13	No: 33	Yes: 4	Sand: 14	Vermiculite: 23	No: 25	Yes: 12
NFMD: 4										
Meat: Fresh meat daily (FMD) or not (NFMD). Destructive (sampling): 12 individuals removed from the cohort at each sampling time. Transfer: 125 postfeeding third instars were transferred to 500 mL of fresh pupation substrate. Substrate Under: food was placed on top of a substrate or at the bottom of an empty jar. Substrate: pupation substrate used. Paper Towel: received FMD placed on a moist paper towel.										

**Table 2.2. Average development times  $\pm$  standard deviations in hours and accumulated degree-days\* for significant rearing treatments of *Lucilia sericata*.**

Treatment	Egg	1 <sup>st</sup> Instar	2 <sup>nd</sup> Instar	3 <sup>rd</sup> Instar	Postfeeding	Pupa	Total
FMD	21.74 $\pm$ 2.72	26.56 $\pm$ 2.18	26.05 $\pm$ 4.59	40.02 $\pm$ 12.55	76.04 $\pm$ 19.85	177.8 $\pm$ 26	368.2 $\pm$ 41.63
	(13.6 $\pm$ 1.7)	(16.6 $\pm$ 1.36)	(16.28 $\pm$ 2.87)	(25.01 $\pm$ 7.85)	(47.52 $\pm$ 12.41)	(111.1 $\pm$ 16.25)	(230.1 $\pm$ 26.02)
NFMD	21.75 $\pm$ 2.1	25.25 $\pm$ 0.87	36.44 $\pm$ 12.49	58.75 $\pm$ 12.7	98.5 $\pm$ 17.71	208.3 $\pm$ 58.93	448.9 $\pm$ 38.8
	(13.59 $\pm$ 1.31)	(15.78 $\pm$ 0.54)	(22.77 $\pm$ 7.81)	(36.72 $\pm$ 7.94)	(61.56 $\pm$ 11.07)	(130.2 $\pm$ 36.83)	(280.6 $\pm$ 24.25)
Paper Towel	22.5 $\pm$ 2.24	25.75 $\pm$ 1.98	26.13 $\pm$ 3.37	23.88 $\pm$ 1.4	68.06 $\pm$ 13.31	175.6 $\pm$ 15.91	341.9 $\pm$ 16.85
	(14.06 $\pm$ 1.4)	(16.09 $\pm$ 1.24)	(16.33 $\pm$ 2.11)	(14.92 $\pm$ 0.87)	(42.54 $\pm$ 8.32)	(109.7 $\pm$ 9.95)	(213.7 $\pm$ 10.53)
No Paper Towel	21.38 $\pm$ 2.77	26.74 $\pm$ 2.13	27.68 $\pm$ 7.59	50.76 $\pm$ 6.12	83.46 $\pm$ 21.87	183.8 $\pm$ 36.53	393.8 $\pm$ 49.25
	(13.36 $\pm$ 1.73)	(16.71 $\pm$ 1.33)	(17.3 $\pm$ 4.74)	(31.73 $\pm$ 3.82)	(52.16 $\pm$ 13.67)	(114.9 $\pm$ 22.83)	(246.1 $\pm$ 30.78)
Transfer	22.02 $\pm$ 3	26.25 $\pm$ 2.25	25.87 $\pm$ 2.84	37.6 $\pm$ 13.37	59.88 $\pm$ 15.33	175 $\pm$ 12.2	346.6 $\pm$ 24.35
	(13.76 $\pm$ 1.88)	(16.41 $\pm$ 1.4)	(16.17 $\pm$ 1.78)	(23.5 $\pm$ 8.36)	(37.43 $\pm$ 9.58)	(109.4 $\pm$ 7.62)	(216.6 $\pm$ 15.22)
No Transfer	21.59 $\pm$ 2.46	26.51 $\pm$ 2.07	27.89 $\pm$ 7.79	44.45 $\pm$ 13.58	88.53 $\pm$ 15.62	184.4 $\pm$ 37.78	393.4 $\pm$ 50.04
	(13.5 $\pm$ 1.54)	(16.57 $\pm$ 1.3)	(17.43 $\pm$ 4.87)	(27.78 $\pm$ 8.49)	(55.33 $\pm$ 9.76)	(115.3 $\pm$ 23.61)	(245.9 $\pm$ 31.2)
Destructive	21.7 $\pm$ 1.89	26.65 $\pm$ 2.33	28.7 $\pm$ 10.22	52.85 $\pm$ 9.8	92.2 $\pm$ 9.72	243.2 $\pm$ 41.23	465.3 $\pm$ 41.5
	(13.56 $\pm$ 1.18)	(16.66 $\pm$ 1.46)	(17.94 $\pm$ 6.39)	(33.03 $\pm$ 6.12)	(57.63 $\pm$ 6.08)	(152 $\pm$ 25.77)	(290.8 $\pm$ 25.94)
Not Destructive	21.75 $\pm$ 2.75	26.38 $\pm$ 2.11	26.94 $\pm$ 5.94	40.35 $\pm$ 13.58	76.32 $\pm$ 21.16	171.4 $\pm$ 14.22	363.2 $\pm$ 31.72
	(13.59 $\pm$ 1.72)	(16.49 $\pm$ 1.32)	(16.84 $\pm$ 3.71)	(25.22 $\pm$ 8.49)	(47.7 $\pm$ 13.22)	(107.1 $\pm$ 8.89)	(227 $\pm$ 19.82)

## **CHAPTER 3: THE USE OF GENERALIZED ADDITIVE MODELS IN ANALYZING FORENSIC ENTOMOLOGICAL DATA**

### *Introduction*

Daubert, et al. v. Merrell Dow Pharmaceuticals (509 US 579 (1993)) was a pivotal ruling for forensic scientists, in which the US Supreme Court declared that the Federal Rules of Evidence (particularly Rule 702), and not Frye (Frye v. United States (293 F. 1013, 1014, D.C. Cir. (1923))), were the standard for scientific evidence and expert testimony. In doing so, the High Court placed the burden of assessing the validity—and thus admissibility—of scientific evidence on the trial judge, based on five main criteria: Has the technique in question been tested; Do standard operating procedures (SOPs) exist for the technique; Has the technique been subjected to peer review and publication in the appropriate literature; Is the technique widely accepted by the relevant scientific community; and finally, What is the known or potential error rate of the technique? DNA-based evidence has set the ‘gold standard’ for meeting Daubert requirements, largely satisfying all of them. In contrast, many of the forensic sciences and resultant expert testimony are based on practitioners’ training and experience, often with little consideration for SOPs, method testing, potential error rates, or publication, even when the technique is generally accepted. As an example, the National Institute of Justice recently posted a solicitation for the study of fingerprints/friction ridges, though certainly this method of identification is extremely well-established. Other areas of forensic science fare far worse (Saks and Koehler 2005).

Forensic entomology falls between these extremes. The predictable growth of carrion-feeding flies has long been used to estimate the time a body has been exposed to

insects, and thus to estimate a post mortem interval (PMI). Using larval size and developmental stage to approximate age is well supported by research and observations in developmental biology, and this forensic technique is widely described in the scientific literature (e.g., Greenberg 1991, Grassberger and Reiter 2001). Likewise, countless legal rulings have assured its admissibility, just as countless juries have been guided by entomological testimony. However, scientists have reported different growth rates for immature flies (Kamal 1958, Greenberg 1991, Wells and Kurahashi 1994, Anderson 2000, Grassberger and Reiter 2001) and court qualified experts have come to incongruent conclusions about a PMI based on the same entomological evidence, depending on which growth data were utilized (e.g., *California v. Westerfield*, CD 165805 (2002)). This problem stems, at least in part, from a general failure to develop SOPs, and also from not fully considering the amount of variation present in larval growth (or more precisely, to account for error rates inherent in estimates of larval age), two of the major tenets of Daubert. The difficulty in estimating error is exacerbated by the fact that blow flies grow in a non-linear fashion and have variable size distributions at different ages, unequally affecting age estimates of developmental stages (Wells and Lamotte 1995).

The research presented here was designed to investigate the variability that occurs in larval and pupal growth of blow flies in order to discern which of a suite of variables have the largest influence on estimating age, and to explore the possibility of placing confidence intervals around juvenile age estimates. Using three regional strains of the blow fly *Lucilia sericata* (Diptera: Calliphoridae) (Meigen), collected in California, Michigan, and West Virginia, a data set containing linear (developmental stage, strain, rearing temperature) and non-linear (length and weight) measures was established.

Generalized additive models (GAMs) were developed taking these variables into account, examining the level to which each influenced/predicted the percent of immature fly development (Hastie and Tibshirani 1990, Wood 2006). Similar GAMs have already been used to assess the effects of cadmium on the growth of *L. sericata* cohorts (Moe et al. 2005), and were assessed here for their potential use as tools in predicting blow fly development percent. The utility of a model was then tested on an independent data set (larvae reared on rat carcasses), focusing on developmental stage and length. GAM predictions of larval development percent were plotted against true age to assess the error of the predictions and to define confidence intervals for these estimates.

### *Materials and Methods*

#### **Species Identification.**

Wild *L. sericata* were collected in California (CA), Michigan (MI), and West Virginia (WV), from the UC Davis campus in June of 2005, the Michigan State University campus starting in May 2005 (which were provisioned with new flies occasionally throughout the summer), and from the West Virginia University campus in August of 2005. Adult individuals from each strain were identified by key (Hall 1948 and Gorham 1987), with independent confirmations, and through mitochondrial cytochrome oxidase 1 gene sequencing (Tarone and Foran 2006).

#### **Growth Experiments.**

Cohorts of flies were raised in a round robin design, in which CA and MI were reared in one block, followed by CA and WV, and WV and MI, between 9/1/05 and 10/24/05. Flies ranged from two to five generations removed from their natural



population. Cohorts were initiated by placing fresh liver into the cages of adult flies, which was checked regularly for eggs. When oviposition occurred the time was recorded and meat and eggs were removed 1 h later. Cohorts were placed in either  $20\pm 0.5^{\circ}\text{C}$  or  $33.5\pm 1.8^{\circ}\text{C}$  incubators under a 12:12 h light cycle at  $25\pm 5\%$  relative humidity. Incubator temperature fluctuation was noted using a HOBO data logger (Onset Computer, Bourne, MA). Eggs were transferred to fresh liver, which was placed on a moist paper towel in 1 L jars, covered with a breathable fabric lid, based on rearing conditions previously found to best mimic those on carrion (Tarone and Foran 2006). Cohorts were given fresh liver daily until postfeeding third instars were observed, at which point 250 individuals ( $33.5^{\circ}\text{C}$  treatments) and 375 individuals ( $20^{\circ}\text{C}$  treatments) were transferred in batches of 125 to 1 L jars containing 500 mL of fresh sand as a pupation substrate.

Length and weight of 2559 larvae/pupae were recorded, starting approximately 24 h after eggs were laid. Length was measured with a ruler based on the furthest extension of a larva to the nearest  $\frac{1}{2}$  mm. Wet weight of live individuals was measured on a Cahn 27 Automatic Electrobalance (Cahn Instruments, Cerritos, CA) to the closest 1/100 mg. Developmental stage was assessed by observing feeding larvae microscopically, by visible crop length and migrating behavior for postfeeding larvae, and puparium formation for pupae. Ten larvae were removed from a cohort and measured/weighed, twice daily (in the morning and late afternoon). Ten pupae were collected once daily and measured/weighed; 5 individuals were collected if less than 10 were available.

Earlier research showed that the destructive sampling of pupae delayed the appearance of adults (Tarone and Foran 2006). To account for this, pupal age was calibrated to the day of pupation. This means that pupal samples were assessed in groups

that pupated within 24 h of each other (i.e. 0–1, 1–2, 2–3, etc. day old puparia) with the minimum development time for pupation being the minimum development time for any individual within a collective group of pupae.

Forensic entomologists generally assess fly growth progression using a measure of relative age, allowing them to take into account the substantial influence of temperature on development. Given that multiple variables had the potential to affect immature fly growth rates in the current research, including understood (e.g., temperature) and questioned (e.g., fly strain) factors, a method that would allow growth progression to be compared directly among all flies was required. Development percent, or the relative (developmental) age of an individual, was used to assess the extent to which a fly had progressed towards maturation (eclosion). This measure, often used for relative developmental comparisons (e.g., Rogina and Helfand 1995, Rogina et al. 1997, Anderson 2000), permitted individuals at all points in development to be compared, which would be impossible if, for instance, temperature and fly strain varied in their influence on growth. Development percent was calculated by determining the age in hours of an individual, then dividing the age by the minimum total development time of that experimental replicate. As an example, if an individual was sampled 100 h after oviposition and the minimum development time for the replicate was 285 h, then the individual was considered 35% developed.

The laboratory growth of larvae on rats have been described previously (Tarone and Foran 2006) and differed from the measured cohorts primarily in food source and temperature (25°C). Three cohorts of Michigan *L. sericata* larvae were reared on rat carcasses and the developmental stage and length of twelve individuals were recorded

daily from each cohort through the first day that puparia were observed. These data were used to predict age. The ethical guidelines of the Michigan State University Laboratory Animal Resources unit were followed, adhering to IACUC requirements.

### **Statistical Analyses.**

GAMs were developed using the mgcv library in the R statistical package (R core development team 2004). The models use likelihood statistics to predict a value (e.g., age) based on various input data. GAMs relate non-linear data such as fly length and weight to the predicted value (e.g., development percent) using smoothed, non-linear mathematical functions (Hastie and Tibshirani 1990, Moe et al. 2005, Wood 2006). In this manner, the relationship of two non-linear variables to each other can also be included in GAMs (Wood 2006), so a length-by-weight term was also evaluated. Distributions must be applied to the functions used to make predictions in a GAM, which is done through a link function. Based on the results of residual plots produced for the models, a gamma distribution (instead of a normal distribution) with a log link function was most appropriate for the models evaluated. Diagnostic plots were compared among models in order to confirm the validity of distributional assumptions in a model and to compare the predicted versus true age. The first plot was a quantile-quantile graph of modeled data versus data from samples. If the assumptions of a model are correct, this line is straight. The next plot was a graph of residuals against predictions. The data should be evenly distributed above and below zero, with no difference in residuals along the linear predictor axis; unevenly dispersed residuals indicate that the assumed data distribution in the model is inaccurate. The third plot was a comparison of the distribution of residuals, which should appear as a bell curve (most error is small, with

rare instances of larger error). The final plot was a graph of true (response) versus predicted (fitted) values for all data used to construct the model. For simplicity's sake this will be referred to as the  $Y = X$  line, or  $Y$  (predicted age) =  $X$  (true age). The most precise models have all predicted age values clustered close to the  $Y = X$  line, with no gaps in the line. A gap in predictions results in an aging inaccuracy because an individual of an age found in a gap will necessarily be predicted as either older or younger than it actually is. More detailed information on GAM can be found in (Hastie and Tibshirani 1990, Moe et al. 2005, Mansson et al. 2005, Wood 2006).

Models generated several statistics. For linear models the statistic used to explain how closely data match a model is  $R^2$ ; as length and weight data are non-linear the apposite statistic for GAMs is the percent deviance explained (Wood 2006). Degrees of freedom or estimated degrees of freedom (a non-linear equivalent) were determined, as was a P-value, which was based on the likelihood of a variable being predictive of age. P-values in GAMs are considered estimates because likelihood statistics do not yield actual P-values, but do provide values that are similar and can be used to estimate the more familiar statistic. These estimates can vary by up to two times the actual P-value (Wood 2006), thus terms were not considered significant unless P-values were less than 0.025. Additionally, multiple variables were included in some models, requiring a Bonferroni correction that resulted in a significance threshold of  $P < 0.0042$ . Given the inherent inaccuracy of estimated P-values, they were only used to identify informative terms or terms that were candidates for removal from a model owing to intermediate or non-significant P-values. The inclusion or removal of a term, however, was ultimately

decided by the statistic used to compare models: the generalized cross validation (GCV) score, which is an information criterion that is lower for better models (Wood 2006).

Six terms were used to develop models: fly developmental stage, length, weight, length-by-weight, strain, and temperature. Stage, strain, and temperature were considered linear variables, and length, weight, and the two plotted against each other were non-linear. This resulted in 63 possible models, hence only a subset is presented here. The first six models examined each variable by itself, while the remaining 12 combined variables to assess improvements gained (as measured by a decrease in GCV) from including specific terms. Developmental stage was considered the primary variable, as all forensic entomologists include this in PMI predictions. Body size is also often incorporated into PMI estimates, thus length and weight were added to several models, as well as being examined in combination. Next, the influences of strain and temperature were tested through inclusion with the more familiar variables (stage, length, weight). Similarly, since length-by-weight is a somewhat novel measure it was evaluated in combination with the three standard variables, and then with all variables.

Finally, a GAM incorporating the standard variables used to age flies in forensic entomological enquiries, developmental stage and length (Kamal 1958, Greenberg 1991, Anderson 2000, Grassberger and Reiter 2001, Tarone and Foran 2006), was tested against an independently derived data set. The model-based predictions of larval development percent for three previously collected fly cohorts raised on rats were plotted against their true development, comparing them to the predicted 95% confidence intervals for the model (precision) and the  $Y = X$  line (accuracy). Confidence intervals were superimposed over the predictions made for rat cohorts (using the quantreg library

in R) by plotting locally weighted sum of squares curves through the 97.5<sup>th</sup> and 2.5<sup>th</sup> percentiles.

## *Results*

### **Species Identification.**

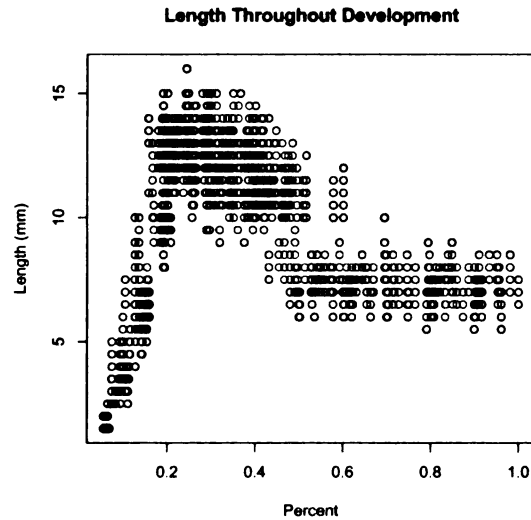
Flies collected from the three states were identified as *L. sericata* based on both visual verification, visual confirmation by an independent entomologist, and cytochrome oxidase 1 sequence data (accession numbers DQ868503, DQ868523, and DQ868524 for CA, MI, and WV respectively). Sequences obtained from the CA strain, the MI strain, and the WV strain were 428 and 227 non-overlapping base pairs, 774 continuous base pairs, and 776 continuous base pairs in length, respectively. BLAST results for the sequences showed the closest match for all was to *L. sericata*, with 100 % similarity to at least one other *L. sericata* sequence. The next closest species match was *L. cuprina* with a 98% to 99% similarity (5–8 base pairs difference).

### **Immature Development.**

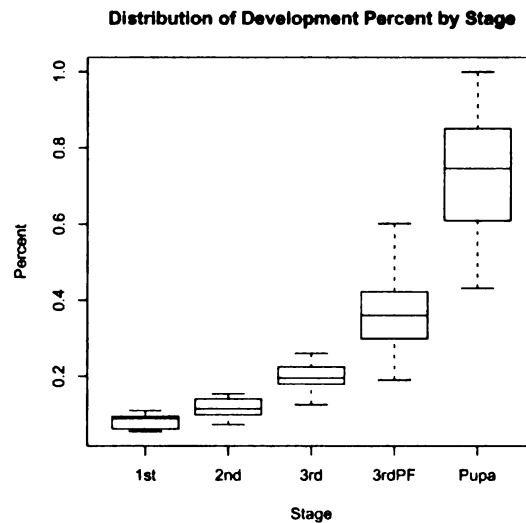
Figure 3.1 depicts a plot of fly length against percent juvenile development. The feeding portion of the lifecycle makes up the initial 25%, and shows a linear increase in length. The postfeeding third instar, where body size decreases and variation in size increases, is found from approximately 25–50%. The relatively unchanged second half of the plots is the pupal stage. Weight results displayed the same pattern (data not shown), and both demonstrated that the distribution of sizes in the feeding stages was much smaller than it was in postfeeding third instar larvae and pupae. Minimum and maximum development percents for each stage of development were: First instar = 5.5–

11.0%; Second instar = 7.4–15.4%; Feeding third instar = 12.6–26.0%; Postfeeding third instar = 19.1–60.1%; and Pupa = 43.2–100% (Figure 3.2).

Size was influenced slightly, but significantly, by temperature and strain. CA individuals tended to be larger than MI, which were larger than WV (Figure 3.3). Differences in size among strains were not observed during feeding stages, but were observed once feeding ceased (Figure 3.3) as each strain initiated the postfeeding third instar at different points in development, resulting in variation in average pupal sizes. Also, growth at 20°C yielded larger individuals on average than did growth at 33.5°C, presumably due to a change in the relative rate of development for feeding larvae (Figure 3.3). Size differences caused by both strain and temperature were repeatable, though average differences were well within the variation observed for size traits (e.g., Figure 3.1), resulting in an overlap of body sizes among all strains and both temperature treatments.



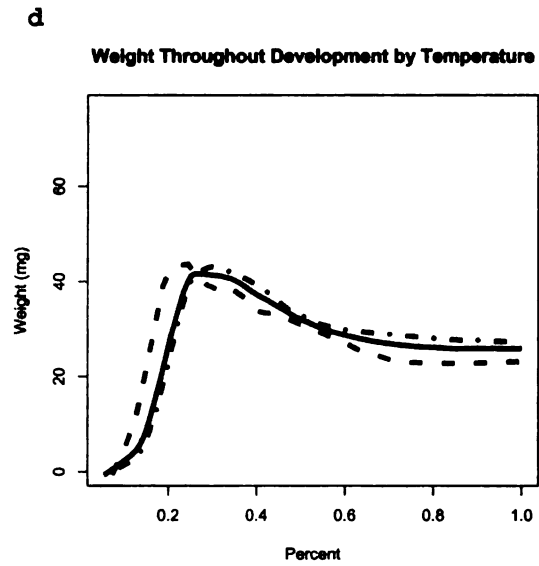
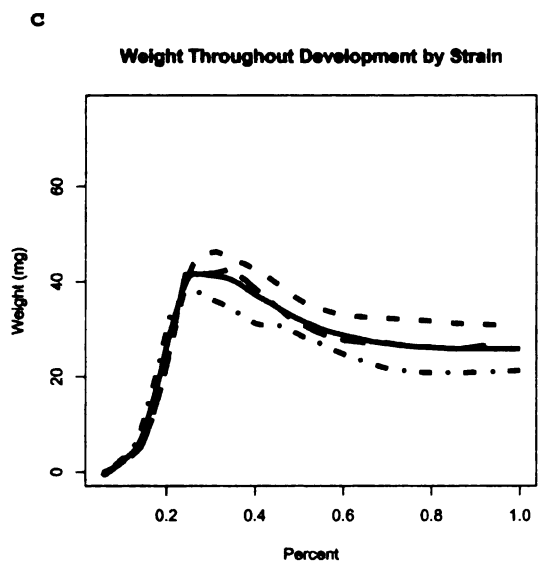
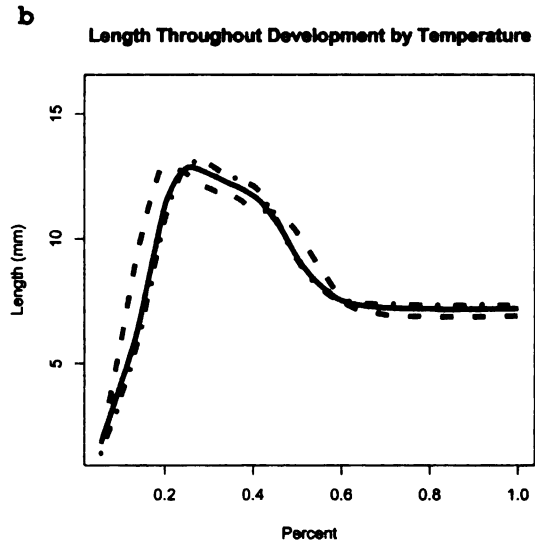
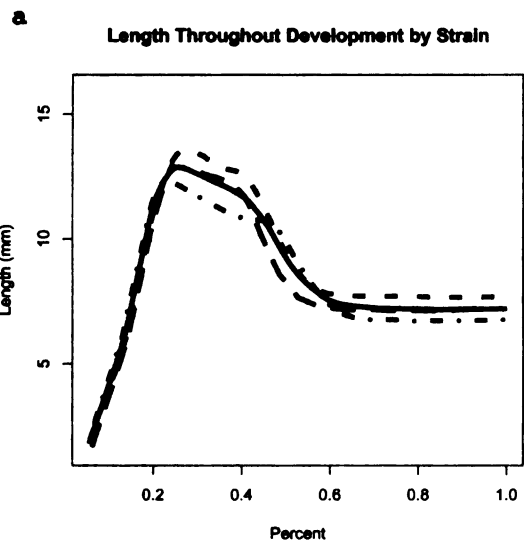
**Figure 3.1. The lengths (mm) of 2559 immature *L. sericata* throughout immature development (percent of development values are on a 0–1 scale).** Note the tight distribution of sizes during the earlier, linear growth phase compared to the more variable postfeeding third instar and pupal stages.



**Figure 3.2. A plot of the distribution of development percents for individuals at each developmental stage.** As development progressed, the proportion of the lifecycle spent in a stage increased. 3<sup>rd</sup> indicates the feeding portion of the third instar; 3<sup>rd</sup>PF indicates the postfeeding stage of the third instar.



**Figure 3.3. The lengths and weights of individuals throughout development from the 6 cohorts.** Growth is compared by strain and by temperature. Solid lines represent the average for all strains or both temperatures. a) Length (mm) plots for each strain. The largest strain, denoted by the line with short bars and spaces, was CA, and the smallest strain, designated by the line with short bars separated by dots, was WV. The MI strain was close to the average size and is represented by the spaced line with long bars and short spaces. Less size variation existed during the feeding portion of the lifecycle (when size was increasing) than in the postfeeding and pupal stages. b) Length plots comparing growth at 20°C versus 33.5°C. Growth at 20°C is represented by the spaced line with short bars separated by dots and 33.5°C is represented by the line with short bars and long spaces. The higher temperature resulted in a growth curve that had a steeper slope during the linear growth phase of development; individuals from these treatments peaked in body size proportionally faster than cooler treatments, which resulted in smaller body sizes as pupae. c) Weight (mg) plots for each strain. Comparisons among strains were as in (a). d) Weight plots for the two temperature treatments, with similar results as in (b).



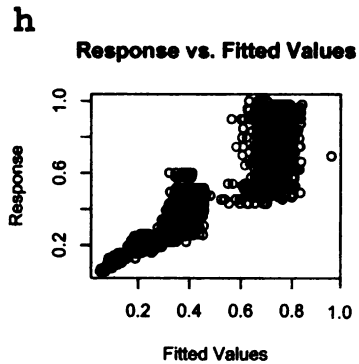
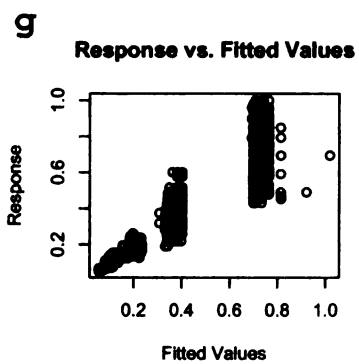
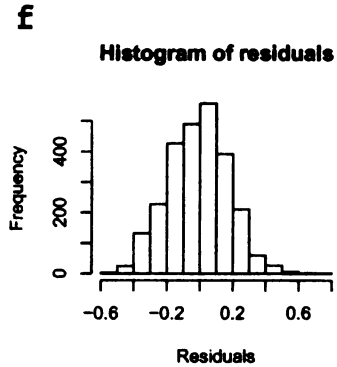
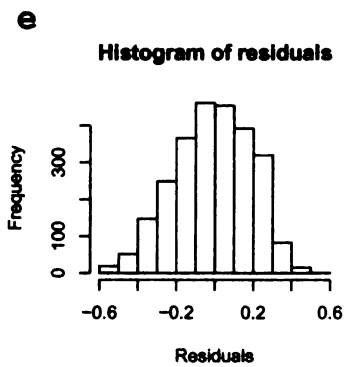
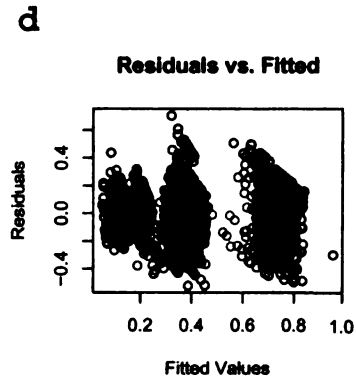
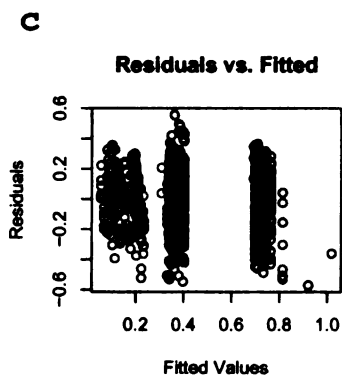
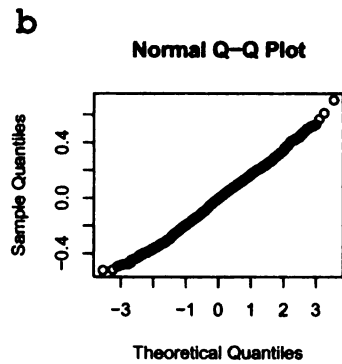
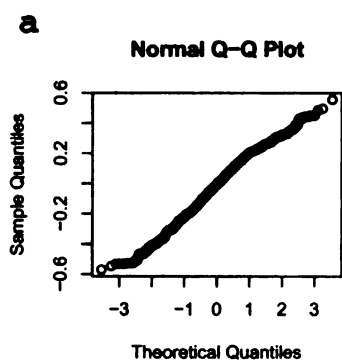
### **Assessing Statistical Models.**

All models demonstrated acceptable levels of error in the diagnostic plots (Figure 3.4a–f), indicating that the use of a gamma distribution with a log link function was appropriate. A comparison of all models examined (Table 3.1) displayed the utility of GAMs to predict development percent when different variables were included. Stage was the single most informative variable (GCV = 0.045), while length and weight garnered less information (GCV = 0.126 and 0.144 respectively); all were statistically significant ( $P < 0.0001$ ; Table 3.2). The length-by-weight term (model 4) provided an intermediate level of information in assessing development (GCV = 0.059). Temperature and strain were not significant predictors of age by themselves (models 5 and 6; GCV = 0.358 for both) and only provided useful information ( $P < 0.0001$  and a decreased GCV) when combined with other variables (e.g., model 18). Predictions with length and weight yielded similar results to model 4, approaching, but not improving upon, the explanatory power of stage alone (model 12; GCV = 0.064). Any model that included stage and at least one body size measure explained 90.8–92.6% of the deviance in the data and GCV scores of 0.04–0.034, with the model that included all variables garnering the highest percent deviance explained and the lowest GCV.

All models were limited in predicting the ages of postfeeding third instars and pupae, generating artificially narrow age ranges. Gaps between stages were most dramatic in model 1 (developmental stage alone), wherein individuals were simply predicted to be the average age of that stage, although true ages were continuous. Inclusion of body size improved predictive precision in the early stages, but not for postfeeding third instars and pupae. As an example, in model 10, which included

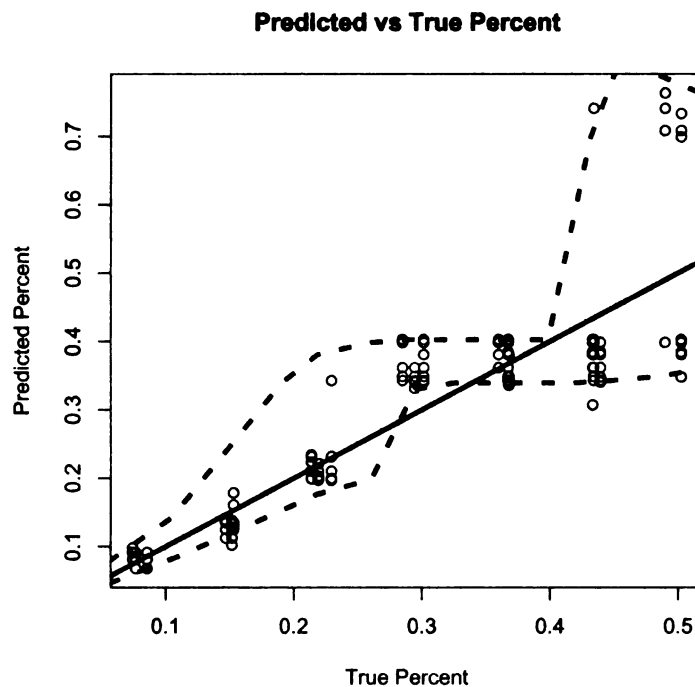
developmental stage and length, postfeeding third instars that were 19.1–60.1% developed (Figure 3.2) were given a restricted age range of 30.7–40.3% (95% confidence intervals in Figure 3.4g). The gap between feeding and postfeeding third instars closed somewhat in more complex models; model 18, which utilized all available data, showed no gap between these stages (Figure 3.4h), although the data still did not cluster along the  $Y = X$  line at the level seen during feeding. The inaccuracy of predictions remained for pupae in all models, where true pupal ages were between 43.2–100% of immature development, but 95% of predictions for pupae using model 18 had fitted values between 61.9–81.2%. Interestingly, predicted ages throughout this range were made for pupae of any true age; that is, there was no slope to the pupal data as there was for the other stages.

**Figure 3.4. A comparison of diagnostic plots for model 10 (a, c, e, g) and model 18 (b, d, f, h).** Model 10 predicted age using length and stage, while model 18 used all data to make age predictions, and explained the most deviance in the data. Panels a) and b) are quantile-quantile plots, assessing the validity of each model's assumptions. The line is relatively straight and increasing for both models, indicating that the distributional assumption of the model did not violate the other assumptions of the model, however model 18 generated a smoother line. Panels c) and d) show the distribution of residuals throughout the lifecycle; in both models residuals are of equivalent size for all ages, showing that there is no bias in residuals based on age. Panels e) and f) are the distributions of residuals, which are normally distributed thus the gamma distribution was acceptable to use with these data. Panels g) and h) are plots of true (response) versus predicted (fitted) values for data used to construct the models. Fitted values accurately predicted true age through linear (feeding) development (the first 25% of the lifecycle), after which the first gap in prediction appeared. This indicates that an overestimation of ages for postfeeding larvae between c. 19.1 and 30% developed was produced using model 10. Model 18 was less likely to overestimate age for individuals at this point in the lifecycle, but predictions were still biased toward overestimating age in young postfeeding third instars. Neither model represents a noticeable improvement over using stage alone for aging pupae. Predicting pupal age with just developmental stage resulted in an age estimate between 43.2 and 100% of immature development. The most predictive model (model 18) plots fitted values between 61.9 and 81.2 % (pupae). At either end of this predicted range, true values could be between 43.2 and 100%. As with all size/stage models, error increased with age.



### **GAM Validation with Independent Data.**

The utility of model 10 (developmental stage and length) was examined through analysis of the previously collected and independently produced rat carcass data set. Consistent with the finding above, error in larval age estimations increased with age (pupae were not considered here as length does not change during the stage). A plot of true versus predicted age (Figure 3.5) shows that age predictions generated for the rat data, when compared to known ages, spanned the  $Y = X$  line and were generally consistent with (inside) the 95% confidence interval provided by the diagnostic plot for model 10. In the feeding stages (<26% of total development) the predictions were approximately  $\pm 5\%$  (or less) of the true age. However, in postfeeding third instars, ages were initially overestimated, then clustered close to the line, and eventually disbursed well below  $Y = X$ , consistent with the expectation that postfeeding individuals could not be precisely aged using length. The model also continued to predict a narrower range of ages for postfeeding larvae (32.5%–40.1%) as compared to their true ages (22.9%–50.2%).



**Figure 3.5. A plot of predicted versus true ages of 252 individual flies raised on rats as estimated with a GAM for stage and length (model 10).** A plot of predicted versus true development percents of 252 larvae raised on rats as estimated with a GAM for developmental stage and length (model 10). Development is displayed to 50% because length measurements ceased when pupation occurred. The solid line represents the  $Y$  (predicted age) =  $X$  (true age) line. Dashed lines represent 95% confidence intervals for the predictions based on the data in (a). Model 10 accurately (data span  $Y = X$ ) and precisely (ranging c. 5% above and below  $Y = X$ ) predicted age for the feeding portion of the lifecycle (<26% of development), when body size increased in a linear fashion. As expected, precision decreased greatly once the postfeeding third instar was reached, although overall error for the model was within predicted levels.

### *Discussion*

The requirements of Daubert necessitate standardized methodologies and knowledge of potential error, two criteria where several forensic sciences, including entomology, may be found lacking. Previously we examined how variation in published rearing protocols, which are not standardized among laboratories, affect growth rates of



juvenile blow flies (Tarone and Foran 2006). In the current research the ability to conduct statistical analysis of blow fly growth and aging, including confidence intervals, error rates, and model comparisons, was tested. From a practical standpoint, the methodology allows for direct estimates of error that should satisfy both scientific and Daubert considerations. For instance, if stage and length are used to estimate that a larva is 15% developed, model 10 generates a 95% confidence interval of approximately 10–22% (Figure 3.5). Using a published minimum development time for *L. sericata* of 288 h at 26.7°C (Kamal 1958), an age estimate of 29–52 h is produced, with the requisite error described. For postfeeding third instars, an estimate of 40% developed (115 h) corresponds to a 95% confidence interval of approximately 22–60%, or 63–173 h. Though the precision necessarily decreases in the latter stages, the window of time placed around that prediction is now mathematically defined. The methodology also has the flexibility to incorporate other data that may be collected.

Through the modeling used in this study, several key points became apparent. First, developmental stage was the single most predictive factor in the models assessed, explaining 89.5% of the deviance in the data. Logically this makes sense, as stage is a direct measurement of developmental progress. In contrast, the non-linear measurements—weight and length—although still significant, proved far less effective in predicting development, while strain and temperature (genetic and environmental factors) were by themselves not significant predictors. The ability of stage, length, and weight to estimate age was greatest during the earliest phases of development, but for different reasons. Egg, first instar, and second instar are by far the shortest developmental stages in flies (Figure 3.2), thus identification of any of these simply described development

more accurately than did the much longer third instar and pupation. Weight and length on the other hand are related to feeding, and changed in a relatively linear fashion during the early larval stages, including the first portion of the third instar (e.g., Figure 3.1), but once feeding ceased their utility dropped dramatically due to the reversal in body size and larger overall variance in length and weight. Likewise, pupal size was of little utility as it is static throughout the stage. As a result, model 18, which used all available information, predicted a restricted pupal development of 61.9% to 81.2% (Figure 3.4h, 95% confidence intervals) and showed no specificity (that is, the youngest or oldest individuals were equally likely to be placed anywhere within that range). This means a pupal development prediction with the best GAM was essentially the same as using stage alone. Given these effects on predictive ability, it is not surprising that adding weight and/or length to stage resulted in minimally improved models.

Second, error increased as development progressed for all models, indicated by the gaps in predictive ability and the widening confidence intervals for successive developmental stages (Figure 3.4g,h), which were most pronounced in postfeeding third instars and pupae. The increasing inaccuracy of age approximation as fly development progresses has been noted in the literature (Wells and Kurahashi 1994, Wells and Lamotte 1995), and forensic entomologists account for it in PMI estimates by giving large age ranges to postfeeding flies, although these rarely include an objective estimate of error. The latter study (Wells and Lamotte 1995) used linear models to estimate blow fly age based on length data, and yielded an increase in error for older larvae. The similar findings indicate that there is a limit to the precision in blow fly age predictions that can be achieved when only developmental stage and body size are evaluated. Owing

to this, alternative developmental data independent of basic growth are likely required to increase the accuracy of PMI predictions, and in the future, traits that change regularly during fly development, such as hormone titers or gene expression, may be useful in generating a more specific PMI.

Third, the limited influence of fly strain and rearing temperature on development is an important consideration as it indicates the models have value regardless of where flies are collected or at what temperature they develop. This is not to say that temperature is unimportant when making PMI estimates—it is critical, and is always considered when estimating PMI (usually as accumulated degree days). However, temperature did not alter developmental patterns to any large extent, although lower rearing temperatures did result in slightly larger individuals overall for all strains, a finding we continue to investigate. Similarly, the strains of flies examined had different average sizes during development (Figure 3.3). For both criteria the distributions of body size throughout development overlapped, so these data modified age predictions minimally. Also, there was little difference in size among strains during the feeding stages of the lifecycle, where size best predicts age, thus size variability resulting from strain adds no confounding information during those stages.

Fourth, any given forensic case may present the entomologist with different data from which to estimate fly age. While developmental stage was the most useful datum for the development estimates in this study, other data, such as weight and length, increased their accuracy. Using a model that incorporates all available data can help ensure that investigators make the best possible prediction with the information they receive, while maintaining an understanding of the limitations inherent in that model. An

estimate of the relative reliability of a PMI prediction (based on the GCV and percent deviance explained) provides an understanding of its value in interpreting evidence. Overall, generalized additive models offer a useful means of incorporating information from multiple linear and non-linear variables to predict blow fly age, variables that can be accommodated even if they change from one case to the next.

Finally and most importantly, a comparison of modeled development predictions to the independently derived rat data made it possible to assess error rates and produce confidence intervals in these estimates. Individuals less than 26% developed (feeding larvae) generated the most accurate predictions; when 12 individuals of the same age were sampled from a cohort, the predictions clustered around the known age in all instances (Figure 3.5). In contrast, postfeeding stages had a much larger error rate. It is worth noting that even when model 10 yielded its best estimates, there was still about 10% total variance in predicted ages of larvae (note again, at a true age of 15%, the 95% confidence interval for rat data predictions was between *c.* 10–22%, thus stage and size were not ‘perfect’ in estimating development even in the youngest individuals. The utility of the methodology presented here is that it establishes a defined way to produce confidence intervals around entomologically based PMI predictions, regardless of fly age. Until new and independent variables that change in a predictable manner during development are incorporated into age estimates, this error will necessarily exist, and increase with age, however through these models that error can objectively be determined. Equipped with such knowledge the forensic entomologist can relay to the court the level of error found in a PMI prediction. Through this feat, one of the major requirements of Daubert is more fully addressed.

**Table 3.1—The 18 generalized additive models for predicting development percent assessed in this experiment.**

<b><u>Model</u></b>	<b><u>Development Percent=</u></b>	<b><u>Percent</u></b>	<b><u>GCV</u></b>
1	Stage	89.5	0.045
2	s(Length)	63.3	0.126
3	s(Weight)	65.7	0.144
4	s(Length,Weight)	86.8	0.059
5	Temperature	0.022	0.358
6	Strain	0.0041	0.358
7	Stage+Strain	89.5	0.044
8	Stage+Temperature	89.7	0.044
9	Stage+Strain+Temperature	89.7	0.044
10	Stage+s(Length)	91.2	0.038
11	Stage+s(Weight)	90.8	0.04
12	s(Length)+s(Weight)	85.9	0.064
13	Stage+s(Length)+s(Weight)	91.6	0.036
14	Stage+s(Length)+s(Weight)+s(Length,Weight)	92	0.035
15	Stage+s(Length)+s(Weight)+Temperature	91.8	0.036
16	Stage+s(Length)+s(Weight)+Strain	91.6	0.036
17	Stage+s(Length)+s(Weight)+Temperature+Strain	92	0.036
18	Stage+s(Length)+s(Weight)+s(Length,Weight)+Temperature+Strain	92.6	0.034

s(variable)—Indicates that a smoothing curve was used in the GAM for this variable.

s(Length,Weight)—Indicates that a smoothed contour surface of length plotted against weight was used in the GAM.

Development percent—Indicates the variables used in each model to predict development percent.

Percent—Indicates the percent deviance explained.

GCV—Generalized cross validation score; lower scores are better models for predicting development percent.

**Table 3.2—Summaries of the estimated significance of each variable in a model.**

<b><u>Model</u></b>	<b><u>Parameter</u></b>	<b><u>df or edf</u></b>	<b><u>Chi-sq</u></b>	<b><u>P-value</u></b>
1	Stage	4	13658	<0.0001
2	Length	8.734	3275.5	<0.0001
3	Weight	8.026	2629.3	<0.0001
4	Length,Weight	26.11	11420	<0.0001
5	Temperature	1	0.68	0.41
6	Strain	2	0.13	0.94
7	Stage	4	13757	<0.0001
	Strain	2	17.408	0.0002
8	Stage	4	13669	<0.0001
	Temperature	1	3.26	0.071
9	Stage	4	13767	<0.0001
	Strain	2	17.195	0.0002
	Temperature	1	3.0566	0.081
10	Stage	4	4921.9	<0.0001
	Length	8.318	120.38	<0.0001
11	Stage	4	6209.5	<0.0001
	Weight	4.738	143.62	<0.0001
12	Length	6.345	4109.9	<0.0001
	Weight	8.277	3668.9	<0.0001
13	Stage	4	912.31	<0.0001
	Length	7.33	30.746	<0.0001
	Weight	3.902	58.757	<0.0001
14	Stage	4	630.48	<0.0001
	Length	2.381	2.8875	0.3
	Weight	4.765	17.601	0.003
	Length,Weight	26.99	114.66	<0.0001
15	Stage	4	901.98	<0.0001
	Length	7.356	33.427	<0.0001
	Weight	3.953	59.039	<0.0001
	Temperature	1	12.533	0.0004
16	Stage	4	865.21	<0.0001
	Length	7.272	33.759	<0.0001
	Weight	4.236	64.411	<0.0001
	Strain	2	10.695	0.0048

**Table 3.2 Continued**

<b><u>Model</u></b>	<b><u>Parameter</u></b>	<b><u>df or edf</u></b>	<b><u>Chi-sq</u></b>	<b><u>P-value</u></b>
17	Stage	4	871.25	<0.0001
	Length	6.612	34.945	<0.0001
	Weight	4.182	71.738	<0.0001
	Temperature	1	15.726	<0.0001
	Strain	2	13.646	0.0011
18	Stage	4	606.21	<0.0001
	Length	3.842	9.5722	0.044
	Weight	8.498	5.3936	0.76
	Length,Weight	16.03	115.11	<0.0001
	Temperature	1	32.872	<0.0001
	Strain	2	34.101	<0.0001

df—Degree of freedom.

edf—Estimated degree of freedom (non-parametric variables).

Chi-sq—Chi-squared score.

P-value—As estimated from likelihood scores for each parameter in a model.

Figure 1

## CHAPTER 4: GENE EXPRESSION IN BLOW FLY EGGS

### *Introduction*

Insects found on human remains can be useful in estimating a postmortem interval (PMI) during death investigations (Catts and Haskell 1990). Primary among these are the blow flies (Diptera: Calliphoridae), whose state of development when collected from a corpse can be compared to published tables of juvenile fly growth, in order to approximate when the eggs were deposited. As development continues, the larvae pass through three instars and then move away from their food source in order to pupate. For many necrophagous fly species, including the widely distributed blow fly *Lucilia sericata*, growth rates are well defined (e.g., Kamal 1958, Greenberg 1990, Anderson 2000, Grassberger and Reiter 2001). However, developmental stages necessarily exist over a period of time, in some cases several days, making precise PMI estimates difficult. Given this, any information that could be added to fly development stage data has the potential to generate a more precise PMI.

While outward characteristics such as body size or instar have generally been used to estimate fly age, other traits that are developmentally regulated, including the differential expression of genes, offer great potential as an independent source of data for estimating blow fly age. Developmental biology research has uncovered numerous instances of gene expression changes throughout maturation (see Kalthoff 2001, and references therein). Flies have been particularly well studied in this regard (Henrich and Brown 1995, Arbeitman et al. 2002, Skaer et al. 2002, Sullivan and Thummel 2003, Luders et al. 2003, Ali et al. 2005, McGregor 2005), including the Calliphoridae.



Predictable changes in gene expression during development led to the hypothesis tested here, that differential gene expression could be used to make more precise PMI predictions, by effectively breaking a developmental stage into smaller developmental units. Towards this goal, mRNA levels of three genes differentially expressed in *Drosophila melanogaster* eggs (Arbeitman et al. 2002), *bicoid* (*bcd*), *slalom* (*sll*), and *chitin synthase* (*cs*), were assayed in *L. sericata*. Eggs were chosen because there is no quantitative means of assessing their degree of maturity, and if egg aging is attempted at all, investigators must rely on a qualitative evaluation of embryos, making it difficult to objectively divide the stage into developmental subgroups. *bcd* is required early in egg development, defining the anterior end of the egg during the formation of the anterior-posterior axis in Cyclorrhaphan flies (McGregor 2005), including the Drosophilidae and Calliphoridae. *sll* affects dorsal-ventral patterning (Luders et al. 2003), and is also highly expressed in the salivary glands of *D. melanogaster* and *L. sericata* larvae (Ali et al. 2005). *cs* was profiled as chitin formation is required for the proper assembly of the larval cuticle (Tellam et al. 2000). Transcript abundances were assessed to directly test the hypothesis that developmental stages of forensically important flies can be better defined by combining expression information from specific genes, resulting in more precise age estimates, as well as a more precise prediction of PMI.

## *Materials and Methods*

### **Species Identification and Egg Collection.**

*L. sericata* was collected in East Lansing, Michigan, and was identified visually and genetically as previously described (Tarone and Foran 2006). A fly cage at room temperature was presented with beef liver and examined every 15 minutes until females were observed laying eggs, which was allowed to continue for one hour. Egg masses (comprised of approximately 250 eggs) were placed on a moist paper towel in a petri dish at 32°C, and whole masses were collected hourly until hatching of the remaining eggs was observed. Sampled masses were immediately fixed in RNA Later (Applied Biosystems, Foster City, CA) and stored at -80°C. Two replicates were collected for each hourly age period. Prior to RNA extraction, egg masses were thawed and sliced with a razor blade into fifths, resulting in the analysis of 10 groups of eggs for each one-hour collection span. The first eggs hatched between 9 and 10 hours, thus the 8–9 hour time span was the oldest analyzed.

### **Gene Sequencing and Primer Design.**

Expression levels of *bcd*, *sll*, and *cs* were compared to the steady-state expression of two housekeeping genes (*rp49* and *β tubulin 56D*). *L. sericata* gene sequences were available for *bcd*, *sll*, and *rp49* on the National Center for Biotechnology Information website ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), thus quantitative PCR primers were designed directly from them using Applied Biosystems Primer Express software. *β tubulin 56D* and *cs* sequences were obtained using primers for the *D. melanogaster* and *L. cuprina* genes respectively (Table 1), taken from NCBI. PCR consisted of 35 cycles of denaturing at 95°C for 30 seconds, annealing primers at 55°C for 30 seconds, and extending amplicons

at 72°C. Extension times were 4 min for *cs* and 2 min for *β tubulin 56D*. Sequences were generated on a CEQ 8000 capillary electrophoresis system using a CEQ DTCS Quick Start Kit and the same primers, following the manufacturer's protocols (Beckman Coulter, Fullerton, CA). PCR products were analyzed via agarose gel electrophoresis; a single peak in dissociation curves of quantitative PCR (see below) confirmed the electrophoretic evaluation.

### **cDNA Synthesis and Quantitative PCR.**

Ninety RNA samples were isolated from egg masses in a 96-well format using an ABI PRISM 6100 Nucleic Acid PrepStation and the manufacturer's solutions and protocols (Applied Biosystems). Eggs were placed in 300μL of lysis solution without use of a pre-filtration plate. RNA was eluted from plates with 100μL elution solution and incubated at 37°C for 1 hour with 70 units of DNase-I (RNase-Free, Roche, Basel, Switzerland) and Ambion DNaseI buffer (Applied Biosystems). The enzyme was inactivated at 75°C for 10 minutes and the RNA was precipitated using 110μL of isopropanol followed by centrifugation at maximum speed for ½ hr at 4°C. Two 70% ethanol washes followed, using the same centrifuge settings. RNA samples were allowed to air dry for 15 minutes, at which point 32μL Ambion RNase-Free water (Applied Biosystems) was added to the pellet prior to freezing at -80°C.

cDNA was synthesized using a TaqMan Reverse Transcriptase kit (Applied Biosystems) primed by oligo(dT) 16-mers according to the manufacturer's instructions, including 30μL of RNA in a final volume of 120μL. Gene expression levels were assessed by quantitative PCR on an Applied Biosystems 7900HT using SYBRgreen PCR mastermix (Applied Biosystems) in 15μL reactions on a 384-well plate. Each reaction

received 1.5  $\mu$ L cDNA, 7.5  $\mu$ L SYBRgreen PCR mastermix, and 1  $\mu$ L each of forward and reverse primers. The Applied Biosystems recommended thermal cycling parameters were used with the exception that PCR cycles were increased to 50 and a dissociation curve was produced for every reaction. Results were considered valid if a single peak was present in the dissociation curve, indicative of a single amplicon being produced. Optimized primer concentrations, based on trial runs designed to ascertain concentration combinations that provided the largest signal to noise ratio in dissociation curves, are found in Table 4.1.

Reactions without reverse transcriptase acted as controls to confirm that amplification in quantitative PCR was not due to residual DNA. A known (positive) cDNA sample was analyzed in triplicate during every run, allowing for comparisons among 96-well plates, resulting in ninety cDNA samples (10 per time point, in duplicate), six negative controls (PCR mix with no DNA), and six positive controls being assessed for each locus.

Statistical analyses and the construction of plots were performed in the R statistical program (R core development team 2004). Linear regression models were analyzed via type III ANOVA. Standardized gene expression through time was plotted for samples that yielded detectable levels of a transcript. The use of gene expression to assess age was examined via generalized additive models (Hastie and Tibshirani 1990, Wood 2006), which produce a statistic, percent deviance explained (similar to  $R^2$ ), assessing the extent to which a variable influences the data. Predictions (fitted values) for the data were plotted against true ages (response), allowing for evaluation of the model's ability to predict the egg ages.

Final CT values for all loci were generated using the average of duplicate PCRs. CT values of *rp49* and *β tubulin 56D* were averaged and subtracted from those of the developmentally regulated genes to obtain a standardized CT. Regression curves were drawn through standardized plots. Binary gene expression values (1 = present, 0 = not present) were also assessed to determine if the presence or absence of gene expression corresponded to a particular age. A locally weighted sum of squares curve was drawn through the resultant plot. Generalized additive models then allowed prediction of egg age with CT scores and binary values. One model used binary *cs* expression and CT information from the other loci to predict age (N = 55). The other estimated age with CT data for all loci (N = 33). Sample sizes were smaller than the total as only egg masses that provided data for all loci were included in analyses.

## *Results*

*L. sericata* sequences for *β tubulin 56D* and *cs* are listed under the National Center for Biotechnology Information accession numbers EF056211 and EF056212 respectively. *cs* best matched its *L. cuprina* homolog (98% with no gaps) and *β tubulin 56D* exhibited the closest similarity to the fly *Glossina morsitans morsitans tubulin beta-1* gene (86% identity with no gaps); no *Lucilia* sequence was available for the latter comparison.

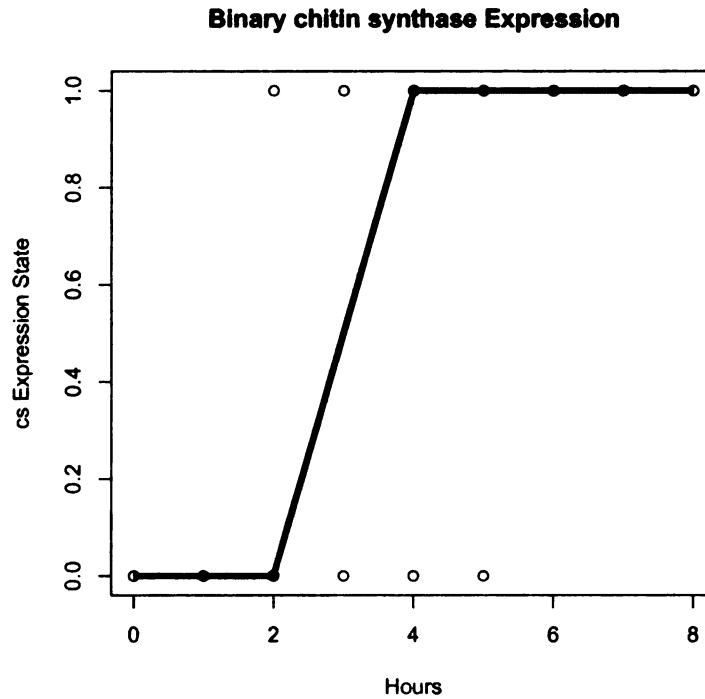
Eighty-four of the 90 samples yielded *rp49* and *β tubulin 56D* profiles, which demonstrated consistent expression levels throughout egg development. There was a significant positive relationship in expression of the two housekeeping genes ( $P < 0.0001$ ,

$R^2=0.63$ ), confirming their utility as internal standards. Of the 84 samples, *bcd*, *cs*, and *sll* had an undetectable transcript level in 23, 31, and 20 samples respectively.

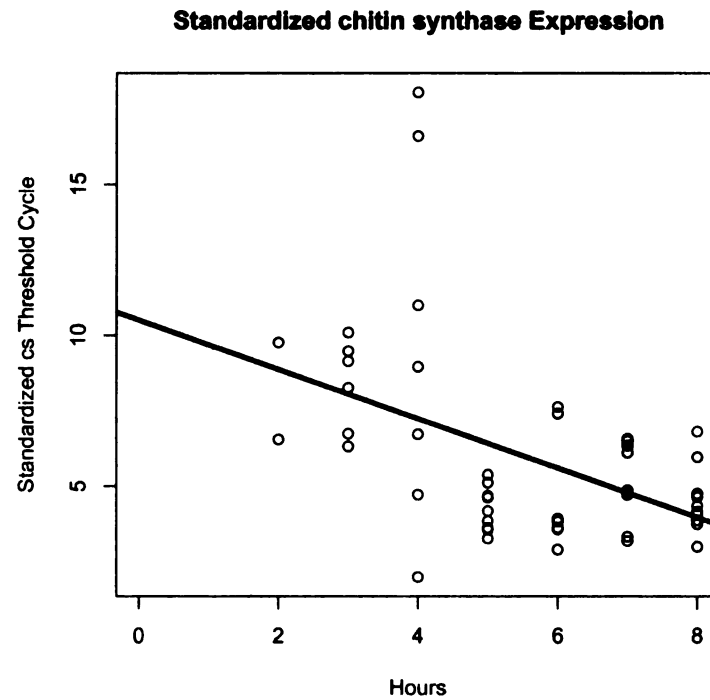
The developmentally regulated genes demonstrated qualitative and quantitative differences in expression throughout egg development. *cs* was the only gene that showed a consistent binary (on/off) pattern, with egg masses less than two hours old never producing the transcript, while those 6 hours and older always expressed the gene, therefore *cs* expression state could be plotted during egg development (Figure 4.1). The presence of the transcript was a statistically significant predictor of age ( $P<0.0001$ ,  $R^2=0.59$ ).

Each of the genes had a different quantitative expression pattern (Figures. 4.2–4; note that the displayed CT values are inversely related to gene expression levels). Though only expressed after hour 2, *cs* transcripts significantly increased during egg development ( $P=0.0004$ ,  $R^2=0.21$ ). Conversely, *bcd* and *sll* transcripts were at their highest levels and lowest variation at 0–2 h, and significantly decreased as development proceeded ( $P=0.0003$ ,  $R^2=0.19$  and  $P=0.0023$ ,  $R^2=0.13$  respectively).

Finally, generalized additive models were used to predict egg ages based on the gene expression data. The first model used the binary expression data for *cs* and CT scores for *bcd* and *sll* to predict egg age. It explained 72.1% of the deviance in the data and accurately enabled the identification of egg masses as either 0–4 or 2–9 h old (not shown). Next, CT scores for all three genes were used to predict age, which explained 76.7% of the deviance in the data. When predicted versus true ages were plotted (Figure 4.5), estimated ages followed the True = Predicted line, with 30 of 33 predictions within 2 h of the true age.

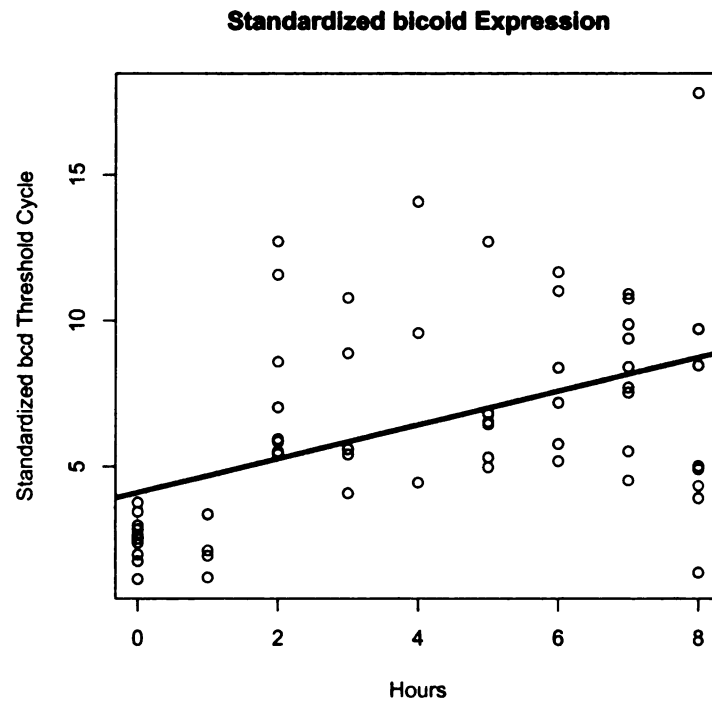


**Figure 4.1. Binary gene expression profile for *cs* at 32°C in *L. sericata* eggs, from 0–1 through 8–9 hours of development.** 0 indicates no detectable expression of the gene and 1 indicates detectable expression of *cs*. Expression was not detected from 0–2 hours. Between 2 and 6 hours some eggs clusters expressed *cs* and some did not. After 6 hours, all egg clusters expressed *cs*.

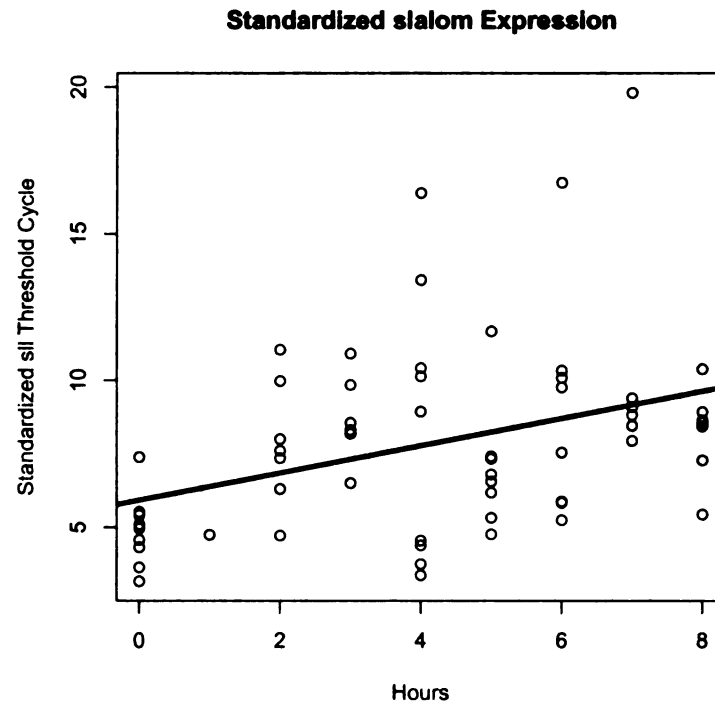


**Figure 4.2. Standardized expression of *cs* in *L. sericata* eggs, from 0–1 through 8–9 hours of development.** The CT was standardized against the average of *rp49* and  $\beta$  *tubulin 56D* CT values and plotted through time. The regression of *cs* CT over time was also included. High expression levels are indicated by low CT values. *cs* was not expressed from 0–2 hours and then its expression increased.





**Figure 4.3. Standardized transcript abundance of *bcd* in *L. sericata* eggs, from 0–1 through 8–9 hours of development.** Standardization was as in Figure 4.2. *bcd* gene expression was highest from 0–2 hours, then transcripts decreased in abundance.



**Figure 4.4. Standardized transcript abundance of *sll* in *L. sericata* eggs, from 0–1 through 8–9 hours of development.** Standardization was as in Figure 4.2. *sll* expression was highest from 0–2 hours, then tended to be lower as eggs developed, though variance in expression was high.

## Discussion

83

were within 2 h of the true age, while the binary *cs* data combined with *bcd* and *sll* CT scores separated the egg masses into two distinct groups.

A key factor in aging flies using gene expression is, of course, examining loci likely to vary in expression levels during the developmental period being examined. In eggs, genes that are important for developmental patterning (e.g., dorsal/ventral) are crucial for successful growth of the individual, thus their expression is under strict biological control. During very early fly embryogenesis high levels of maternally derived *bcd* and *sll* products are necessary to properly establish biological axes (Kalthoff 2001, Arbeitman 2002, Luders et al. 2003, McGregor 2005). In the current study, transcript levels of both genes dropped steadily through embryonic development, although neither became undetectable. In *D. melanogaster*, *bcd* is detectable only during early embryonic development (Arbeitman et al. 2002), thus its expression throughout embryogenesis in *Lucilia*, albeit at decreasing levels, is somewhat puzzling. Developmental heterogeneity among eggs may have resulted in this phenomenon, while it is also possible that *bcd* serves some unknown secondary function in blow flies, or that the transcript is stable but untranscribed in older eggs. In contrast, the existence of *sll* transcripts at the end of the egg stage can be accounted for, likely resulting from endogenous *sll* expression commencing in the developing salivary glands (Luders et al. 2003). Finally, *cs*, which is required only for production of the larval cuticle (Tellam et al. 2000), followed a different and predicted transcriptional course, wherein the earliest portion of egg development was defined by an absence of transcripts, the middle portion of development, as larval cuticle begins to form, was represented by low to intermediate levels of *cs* expression, and the highest levels were found late in egg development. Most

importantly, all three loci showed significant trends in egg transcript expression over time, and taken together increased the precision of egg age estimates.

Given that gene expression has the potential to more accurately age flies of forensic interest, other factors need to be considered, including both the feasibility, and legal acceptance, of the methods. The molecular techniques employed in this study have been widely utilized in developmental molecular biology, and as important, could readily be implemented in most laboratories equipped for forensic DNA investigation. They also provide information that can be used to generate predictable error rates/confidence intervals, meeting one of the major tenets of Daubert, an important consideration of any new forensic protocol. The methods are amenable to microarrays (e.g. Arbeitman et al. 2002) and robotics, potentially producing simplified and high throughput blow fly aging analyses. Finally, DNA-based methodologies overall have been widely accepted in courts of law, thus new but related methods should have less difficulty overcoming Daubert challenges.

The data presented here demonstrate that even the briefest phase of fly development, the egg stage that lasts only several hours, can be divided into smaller periods using gene expression data. Naturally, other stages of fly development, particularly those that last longer and therefore are more forensically challenging (e.g., the third instar and pupation) can be examined using these methods as well. Addition of more developmentally regulated genes into the analysis should further increase the precision of age estimates by providing more age-informative data. The final outcome of this is a more precise age given to developing blow flies, resulting in more precise estimates of PMI.

**Table 4.1. Primers used in this study.**

<b>Primer</b>	<b>Function</b>	<b>Sequence</b>	<b>Accession Number</b>	<b>Concentration (nM)</b>
qBcd F	Quantitative PCR	CCTCGCTCCGGGCCCCATAAC	AJ297856	400
qBcd R	Quantitative PCR	CAACTACTGGCACTTCCATTTGAA	AJ297856	400
qTub F	Quantitative PCR	TCCGTAAATTGGCCGTC AAC	EF056211	400
qTub R	Quantitative PCR	ACCAGGCATGAAAAAGTGAAGAC	EF056211	400
qCs F	Quantitative PCR	GCCGACGGAGAACCTATACCA	EF056212	66.67
qCs R	Quantitative PCR	GATGGCTGTCATTGTGGGTACA	EF056212	400
qSII F	Quantitative PCR	TCCAACGGCCACAACTCTTAAGTA	AY926574	66.67
qSII R	Quantitative PCR	CGTTTAGGTGTTGCCGCAAT	AY926574	400
qRp49 F	Quantitative PCR	ACAA TGTTAAGGA ACTCGAAGTTTGTG	AB118976	400
qRp49 R	Quantitative PCR	GGAGACACCGTGAGCGATT	AB118976	400
Tubulin F1	Sequencing	CGAGACCTACTGCATCGACA	NM_166357	1000
Tubulin R1	Sequencing	CACCAGATCGTTTCATGTTGC	NM_166357	1000
Cs F2	Sequencing	GAACTGCCCTATACCCGTGGA	AF221067	1000
Cs R2	Sequencing	GGATGTAAACACAGCCCGCTAT	AF221067	1000

Accession number denotes the sequence from which the primers were derived. Concentration denotes the final concentration (nM) of primers used in each PCR reaction.

## **CHAPTER 5: PREDICTING BLOW FLY AGE WITH LARVAL AND PUPAL GENE EXPRESSION**

### *Introduction*

Blow flies (Diptera: Calliphoridae) are important parasites, parasitoids, and primary successional species on carrion. Given their parasitic and necrophagous life histories, they affect human activities regularly, as they can spread disease (Faulde et al. 1995) and are parasites of sheep (East and Eisemann 1993), cattle (Crystal and Ramirez 1975), and humans (Sherman 2000), which can cause economic damage and human health hazards. Also, the carrion-feeding behavior of many blow flies makes them valuable evidence in death investigations. Forensic entomology is a forensic science discipline that can aid investigators in the determination of a postmortem interval (PMI) (Catts and Haskell 1990, Greenberg and Kunich 2002). The estimate of a PMI is possible because necrophagous flies, especially blow flies, are capable of colonizing remains and progress reliably through development. Blow flies can lay eggs on wounds or orifices within hours of death. The eggs hatch into larvae, which feed on carrion and progress through three instars (periods of larval development that are separated by molts of the cuticle). Eventually, third instar larvae leave the food source to form a puparium, wherein metamorphosis occurs, leading to the eventual eclosion of the adult. Each developmental stage has a defined, temperature dependent, duration that can be used to estimate the age of an insect collected as evidence. Reported datasets outlining the minimum development times of each stage (Kamal 1958, Greenberg 1990, Wells and Kurahashi 1994, Anderson 2000, Byrd and Allen 2001, Grassberger and Reiter 2001, Grassberger and Reiter 2002) can be used to backtrack from the developmental stage of

an insect associated with a decedent to the time that insects colonized the body. The period of colonization is generally assumed to be the minimum PMI.

The fundamental methods for estimating blow fly age have remained unchanged for decades (Greenberg and Kunich 2002). An investigator collects or receives insect evidence that was associated with a body. The species and developmental stage are identified. Since development is temperature dependent, historical temperature data, usually from weather stations (Catts and Haskell 1990), can be used to estimate the temperature at the death scene. Using a development table that outlines the durations of each stage for the identified species, in conjunction with temperature data, it is possible to determine the minimum and maximum time necessary for that species to achieve the stage that was collected as evidence, resulting in a calculation of the time of colonization of the remains (Catts and Haskell 1990). However, there are several problems associated with the method. As blow fly development progresses, stage durations increase resulting in accurate, but imprecise, estimates of fly age. The imprecision is due to the progressively larger windows of time that must necessarily be placed around an age estimate. For example, according to Grassberger and Reiter (2001), the first instar of *Lucilia sericata* (Diptera: Calliphoridae) (Meigen) lasts a minimum of 24 hours at 20°C, while the pupal stage is at least 209 hours. The windows of time around an age estimates are longest during the postfeeding third instar and pupal stages. The postfeeding third instar has a minimum duration that ranges from 82 hours at 34°C to 200 hours at 17°C, while the pupal stage can last between a minimum of 120 hours at 34°C to 442 hours at 17°C (Grassberger and Reiter 2001).



Body size measurements, such as length and weight, can improve the precision of age estimates, but only help to refine predictions for first, second, and early third instars, when the animals are actively growing. Body size does little to refine predictions of postfeeding third instars, as larvae cease feeding and actually begin to shrink during this period (Greenberg 1990, Anderson 2000, Grassberger and Reiter 2001). They also exhibit increased variance in body size, which adds to the difficulty of using size to estimate age (Wells and Kurahashi 1993, Wells and Lamotte 1995, Chapter 3).

Measuring length and weight does not help in pupal age estimates as pupae do not change in size during development. The failure of body size to refine age estimates for the last two stages of the immature lifecycle, means that predictions of age during those times remain imprecise. In addition, difficulties in distinguishing between feeding and postfeeding third instar larvae (reviewed by Anderson 2000) can compound the problem as predictions of undetermined third instars must include estimates for the entire stage, instead of just the feeding or postfeeding portion.

The difficulties associated with the status quo approach to estimating a PMI based on blow fly evidence must also be considered in the context of the legal requirements of forensic sciences. In the United States, the Daubert ruling on scientific evidence (Daubert v. Merrell Dow Pharmaceuticals, 509 US 579 (1993)) has raised the expectations for all forensic sciences, as it requires scientific methods to be tested and published, using standard operating procedures. The procedures applied must be commonly accepted in the field of study. Likewise, Daubert requires an understanding of error associated with the methods employed. Forensic entomology is based on solid principles in developmental biology and is supported by a body of literature that is

accepted by the scientific community; however, different growth rates for immature blow flies have been reported, in part due to a lack of standardized rearing protocols (Greenberg 1991, Grassberger and Reiter 2001, Kamal 1958, Wells and Kurahashi 1994, Anderson 2000). Further, forensic entomologists have estimated different ages for the same insect evidence, depending on which growth data they use (California v. Westerfield, CD165805). Similarly, there is a dearth of research on estimating error rates of PMI estimates derived from blow fly development tables (Wells and Kurahashi 1993, Wells and Lamotte 1995). Clearly, by addressing the lack of standard operating procedures and statistical understanding in the field, the acceptability of forensic entomology in US courts will be bolstered.

A new approach to predicting blow fly age may be warranted to improve shortcomings of the status quo method of PMI prediction. Any data that can be used as a reliable indicator of specific developmental time periods, especially within the postfeeding third instar and pupal stages when growth is non-linear, will be useful in making more precise age estimates. Gene expression levels are a potential source of new information that could help improve the precision of PMI predictions. As organisms develop they require the expression of proteins in various tissues at specific times in development (Kalthoff 2001). The temporal requirements of cells for proteins means that the genes encoding them will be up- and down-regulated in accordance with the developmental needs of an organism, and may be useful in predicting the age of an individual. The theoretical application of gene expression analyses to blow fly age prediction is well supported by genomic research in the model organism *Drosophila melanogaster* (Diptera: Drosophilidae). Time-series microarray analyses in *Drosophila*,

which have assessed the expression levels of thousands of genes at once, indicate that there are myriad temporally predictable gene expression patterns during fly development (Arbeitman et al. 2002, Beckstead et al. 2005). By choosing the right suite of genes that are up- or down-regulated at different times in development, it may be possible to identify very specific portions of development. For example, during pupation the genes *Amalgam* and CG17184 are expressed at their highest levels during early and late pupation, respectively (Arbeitman et al. 2002). With an evaluation of the expression levels of just these two genes, it should be possible to distinguish between early, middle, and late pupal development. Given the molecular, genetic, and physiological similarities between the Drosophilidae and the Calliphoridae (McAlpine 1989, McGregor 2005, Ali et al. 2005, Yeates and Weigmann 2005, Mellenthin et al. 2006), the *Drosophila* model can help to target blow fly genes with an *a priori* expectation of informative regulation.

Ultimately, predictable temporal variation in gene expression must be demonstrated in blow flies to make estimates of blow fly age, and thus PMI, feasible. The blow fly species studied herein, *L. sericata*, was chosen because it is globally distributed and forensically informative (Kamal 1958, Greenberg 1990, Anderson 2000, Grassberger and Reiter 2001) that has been studied at a molecular level in several instances (e.g. Mellenthin et al. 2006). Gene sequences have also been produced for a sister species, *L. cuprina*, due to the economic effect of myiasis on Australian sheep (East and Eisemann 1993). This means that gene sequence information could be obtained from the public domain, or easily sequenced in *L. sericata* by designing PCR primers from *L. cuprina* sequence.

The research detailed herein was designed to address the hypothesis that estimates of *L. sericata* age could be made more precise by including gene expression data in the prediction process. To determine the usefulness of gene expression, profiles of 12 genes were produced from a time-series collection of immature *L. sericata* that was comprised of 958 individual larvae and pupae. Since gene expression levels are continuously variable (e.g. Gibson and Weir 2005) they can be considered quantitative traits, which are subject to the influence of genetic and environmental factors (Mackay 2001, Conner and Hartl 2004). If gene expression is to be used to predict blow fly age, then influences on variation in expression must be accounted for. Accordingly, the potential for genetic effects was assessed by evaluating gene expression among three regional strains of *L. sericata* (originating from Davis, CA; East Lansing, MI; and Morgantown, WV). Likewise, the profiles produced at two rearing temperatures (20°C and 33.5°C) were assessed, to determine if temperature affects gene expression and the age estimates resulting from them. The data produced enabled the evaluation of gene expression as a means of predicting blow fly age, revealing valuable information regarding the use of transcript level abundances to estimate a PMI.

### *Materials and Methods*

#### **Species Identification, Fly Rearing, and Collections.**

*L. sericata* rearing methods and species confirmation are detailed in Chapter 3. Each collected individual was immediately frozen in RNAlater (Applied Biosystems), at -80°C after length, weight, and developmental stage were recorded.

**Sequencing of *L. sericata* Loci.** Several loci required sequencing before quantitative PCR primers could be designed. These included *β Tubulin 56 D*, *chitin synthase* (both detailed in Chapter 4), *acetylcholine esterase*, *ecdysone receptor*, *ultraspiracle*, *white*, *scalloped wings*, and *rhodopsin 3*. Sequences from the closest dipteran relative available at ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) were used to design PCR primers for the locus, targeting at least a 300 bp segment. Sequencing primers and the sequences used to design them are listed in Table 5.1. Sequencing methods were as in Chapter 4, with slight modifications of annealing temperature, extension times, and the number of cycles depending on the gene in question. Sequences were subsequently compared to known sequences, via BLAST comparison (NCBI), to confirm that the appropriate sequence had been obtained (Table 5.2).

#### **RNA Extraction.**

RNA was isolated from a subset of the 2559 individuals described in the Chapter 3. Five individuals per time point were examined from the Michigan and California replicates raised at 20°C or 33.5°C, and from the West Virginia replicates raised at 33.5°C. The RNA isolation method from Chapter 4, conducted in a 96-well format on an ABI PRISM 6100, was modified for use with larvae and pupae. Modifications stemmed from the fact that adding too much lysed tissue to the wells prevented solutions from being drawn through the filters. Flies were ground in 300 µL of lysis solution by hand with a sterile pestle. To prevent filter clogs, lysates from larvae greater than 10 mg in size and from pupae were diluted in RNA lysis solution (Applied Biosystems); 20 µL of larval lysates or 40 µL of pupal lysates, were placed into 300 µL of additional RNA lysis solution. The dilutions were determined in preliminary experiments to establish the

largest volume of preliminary lysate that did not clog the filters of the 96-well RNA isolation plates. The diluted 300  $\mu$ L lysates of individual larvae and pupae of any size were then drawn through a 96-well filter plate (Applied Biosystems), which removed large particles, larval cuticle, and pieces of puparium from the solution before it was added to the RNA isolation plate. All other steps were followed using the manufacturer suggested protocol, with a final RNA elution volume of 100  $\mu$ L. After RNA purification a DNase I reaction removed any remaining DNA contamination as described in Chapter 4.

### **Reverse Transcription and Quantitative PCR.**

cDNAs and controls were developed as in Chapter 4, except that a High Capacity cDNA Archive Kit (Applied Biosystems) was utilized, according to the manufacturer's instructions. Once the no reverse transcriptase controls were shown to be negative (DNA free) using the *rp49* primers, quantitative PCR was performed with primers for all genes (loci, qPCR primers, and primer concentrations are in Table 5.1). All qPCR products yielded one product of the appropriate size when checked by gel electrophoresis, and dissociation curves were consistent with a single product. Any reactions that had deviant dissociation curves were eliminated from the study.

Quantitative PCR was performed using the same cycling parameters as in Chapter 4, however quantitative PCR was set up and analyzed in a 384-well format using a Biomek 2000 Automated Workstation (Beckman Coulter), not by hand. qPCRs were performed with 10  $\mu$ L reactions, which consisted of 2  $\mu$ L of cDNA. Power SYBR Green PCR Master Mix was used in the reactions.

Samples of cDNA from an individual were divided into two wells and the average CT was used as the score for that gene and individual. Scores were standardized against an internal standard value derived from the average of the housekeeping gene CTs (*rp49* and  *$\beta$  tubulin 56D*) by subtracting the housekeeping score from the gene CT. In addition, reactions were standardized against each other by setting the CT for the positive control reactions to the same value on every 384-well plate in the experiment (positive control reactions were run with *rp49* primers and a standard cDNA sample).

### **Statistical Analyses.**

Statistics and graphs were produced using the R statistical package (R core development team). Several plots were used to understand the data. Gene expression was compared for each developmental stage using boxplots of standardized gene expression CT scores for each locus. Descriptions of boxplots can be found in Chapter 3. Standard CT values for each locus were plotted against minimum development percents (as detailed in Chapter 3), and locally weighted sum of squares curves were drawn through the non-linear data, allowing comparisons of average expression to temperature treatments and average expression among strains. For some graphs (see Results), a gene for which no transcript could be detected was assigned a CT of 50 (the maximum cycle tested) if the housekeeping genes gave a positive result. This kept the graphs from being skewed due to missing data. Values were then plotted against minimum development percents. Corrected expression plots were placed next to standard plots for a locus (no transcript results were removed). Left-right panel comparisons demonstrated an increase in the standardized CT value in the right panels at points where gene expression was absent.

GAMs (Hastie and Tibshirani 1990; Wood 2006) were developed to assess the effects of gene expression (quantitative and binary data), body size, developmental stage, temperature, and strain on predicting minimum development percent. Each variable was assessed on its own, or together with only the linear data (the stage, temperature, and strain variables were only included if the variable was not a significant predictor of age on its own), to determine its usefulness in predicting blow fly age. The variables were used in larger models to determine the best GAMs for predicting blow fly age (Table 5.3). The models construct smoothing curves for non-linear data, which are cumulatively used to predict the age of a fly. However, the distribution of data around that smoothed curve must also be identified. The distribution is applied to the smoothed curve via a link function, thus each model was assigned a distribution and link function based on the criteria discussed in Chapter 3. GAMs produce generalized cross validation (GCV) and percent deviance explained (PDE) statistics, which are also explained in Chapter 3.

The significances of specific terms in the models were determined. Genes were considered significant predictors of age if the estimated P-value associated with the individual gene by itself, or in combination with developmental stage and/or temperature, was below 0.025–0.006 (depending on the Bonferroni correction required). Once individual genes were established as significant predictors of development percent, they were used in combination with other terms in larger GAMs. Comparisons of the GCV and PDE scores from the models helped to determine if gene expression was useful in predicting blow fly age. Diagnostic plots of error in the models were also constructed and are explained in Chapter 3.



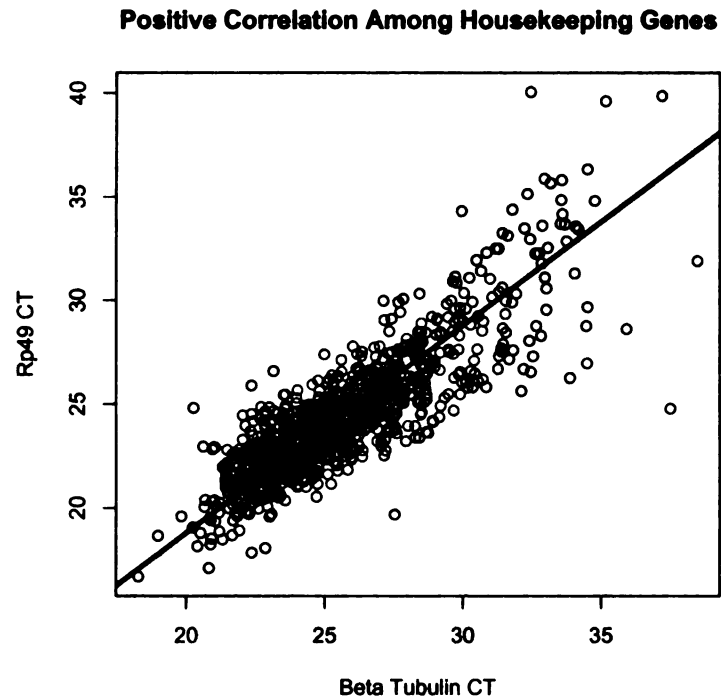
## *Results*

### **Gene Sequences.**

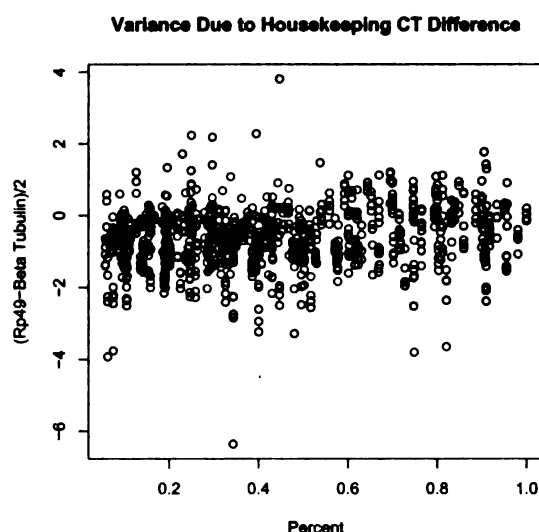
Fly strains were identified as *L. sericata* (Chapter 3). The sequenced genes underwent a BLAST search on the NCBI website ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) (Table 5.2). All exhibited a close match to the same gene in another fly species. Two genes also demonstrated 99–100% matches to *L. sericata* sequences that were published after this project commenced.

### **Quantitative PCR and Statistical Modeling.**

The housekeeping genes demonstrated a strong positive correlation with each other ( $r = 0.84$ ), indicating that they were good surrogates for estimating RNA concentrations (Figure 5.1). A line with a slope of 1 and a Y-intercept of  $-1.2$  (the average CT difference between the two genes) ran directly through the plot of *rp49* CT versus the *tubulin* CT (Figure 5.1). The difference derived from the use of the average of both housekeeping genes resulted in standardization scores that were well within the range of standardized CT values for the other genes analyzed; indicating that most of the variance in CT scores was not due to variation in housekeeping gene transcript levels (Figure 5.2 as compared to Figures 5.3–5.41.).



**Figure 5.1. The positive correlation between housekeeping genes.** The CT scores followed a line with a slope of 1 (after accounting for the average CT difference of 1.2) with a correlation of 0.84.



**Figure 5.2. Variance in gene expression due to the standardization against housekeeping genes.** Since the average of both housekeeping genes was used for standardization, one half of the difference between *rp49* and the *beta tubulin* gene was plotted throughout development. The variance in this plot was relatively constant and indicates that an approximate difference in expression of up to 4 fold (2 CT units) could be explained by variation in housekeeping gene expression.

Of the 2559 flies detailed in Chapter 3, 1025 were used to produce gene expression profiles. Once ~100 samples had been profiled, preliminary plots of the data indicated that *wg*, *scl*, and *rh3* were unlikely to provide useful information (Figures 5.3–5) and were dropped from further study. 958 of the samples yielded partial or full profiles of the remaining 9 genes. These were comprised of 48 first instar larvae, 79 second instar larvae, 135 third instar larvae, 334 postfeeding third instar larvae, and 362 pupae. There were 260 and 272 individuals profiled from the CA and MI strains (respectively) raised in the 20°C treatments. Likewise 149, 121, and 156 individuals from the CA, MI, and WV strains (respectively) were profiled from the 33.5°C treatments. Full gene expression profiles were obtained for 501 samples.

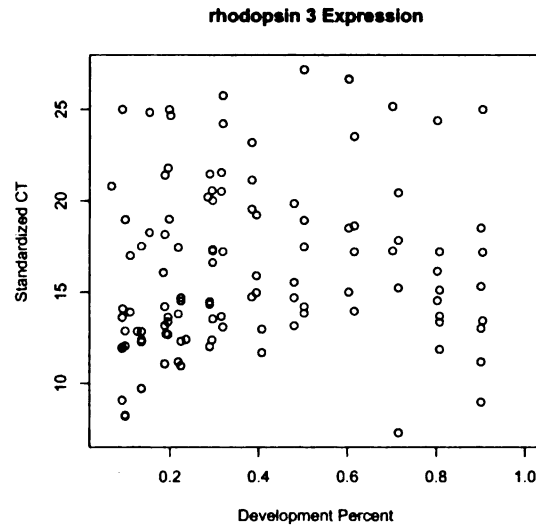
Plots were constructed to show the expression levels of each gene by developmental stage and by development percent. GAMs were used to assess the usefulness of individual gene expression profiles (quantitative and binary) for predicting blow fly age. Table 5.3 lists all GAMs investigated, with models 2–14 assessing the relative usefulness of individual genes. Table 4 lists the significance and degrees of freedom for terms in the models. The results of GAMs and plots for individual genes are listed below (Tables 5.3–4; Figures 5.3–41).



**Figure 5.3. Standardized gene expression of *scalloped wings* throughout the immature development of *L. sericata*.** No apparent pattern of gene expression was observed throughout development.



**Figure 5.4. Standardized gene expression of *wingless* throughout the immature development of *L. sericata*.** No informative pattern of gene expression was observed throughout development.



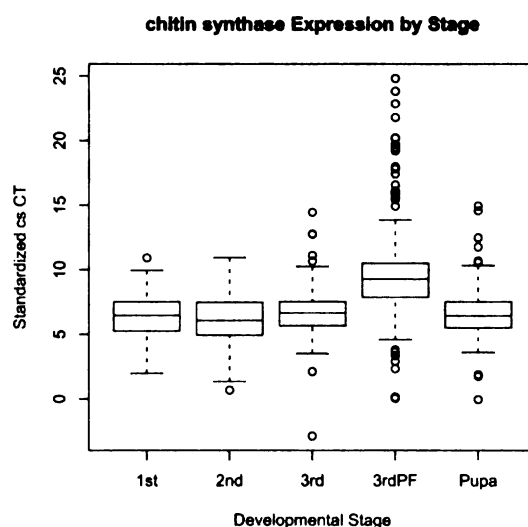
**Figure 5.5. Standardized gene expression of *rhodopsin 3* throughout the immature development of *L. sericata*.** No apparent pattern of gene expression was observed throughout development.

### **Expression of Informative Genes.**

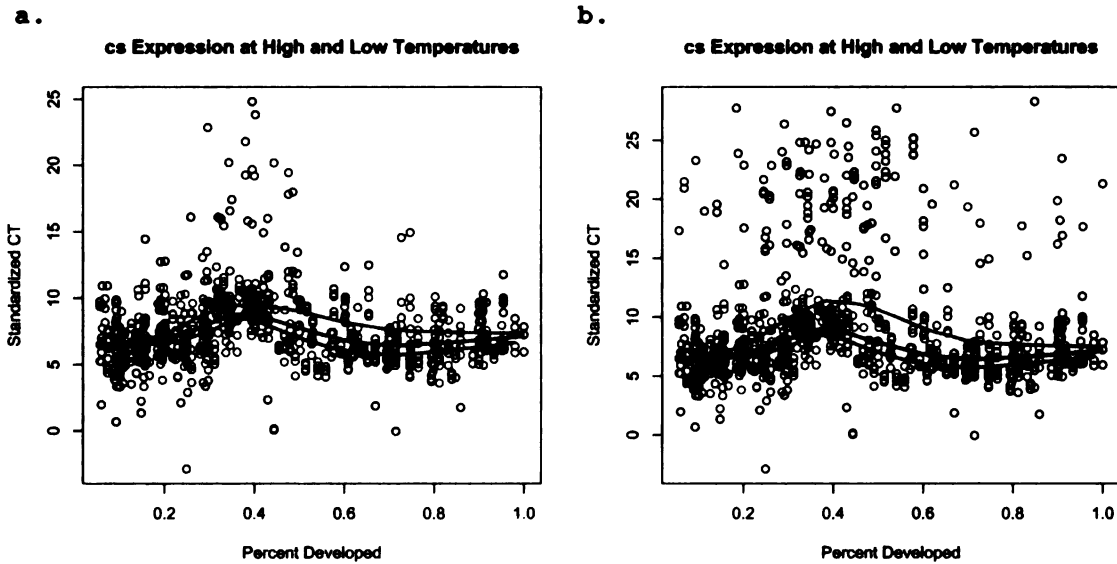
The most informative genes for forensic purposes are those that have stark changes in regulation during development, particularly the latter portion of development, where aging flies is most difficult. Within the third instar: *cs*, *ecr*, *hsp60*, *hsp90*, *rop-1*, *w*, and *usp* were up- or down-regulated between feeding and postfeeding third instars. The most informative overall were *cs* and *ecr*. *cs* was strongly down-regulated during the postfeeding third instar, and it was often not expressed at all (or expressed at an undetectable level). Conversely, *ecr* was highly up-regulated during the postfeeding stage. The reliable expression differences of *cs*, *ecr*, and, to a lesser extent, the other genes noted above are useful indicators of the postfeeding condition and can be used as molecular markers for that stage. The overall behavior of each gene was as follows.

The first gene evaluated was *cs* (Figures 5.6–9; note throughout that higher CT scores indicate lower gene expression levels). Of the 958 individuals assessed, *cs* was

not detected in 102. The gene was expressed at consistent levels, except during the postfeeding third instar, when it was expressed at lower levels (Figure 5.4) or not at all. *cs* tended to be expressed at higher levels at lower temperatures (Figure 5.5), with the MI strain expressing less *cs* at lower temperatures than the CA strain (Figure 5.6), while no expression level differences were apparent among the three strains at the high temperature treatment (Figure 5.7). The CA strain exhibited a stronger tendency to not express *cs* during the postfeeding third instar at 33.5°C (Figure 5.7). GAMs assessing age with CT scores for this gene were only significant predictors of development percent when included with developmental stage (Table 5.4). The GAM that included the binary *cs* term (Model 2) exhibited a 0.3% higher PDE than the GAM that assessed age in terms of stage alone (Model 1), while the quantitative data (Model 6) exhibited an increase of 1.6% in PDE (Table 5.3).

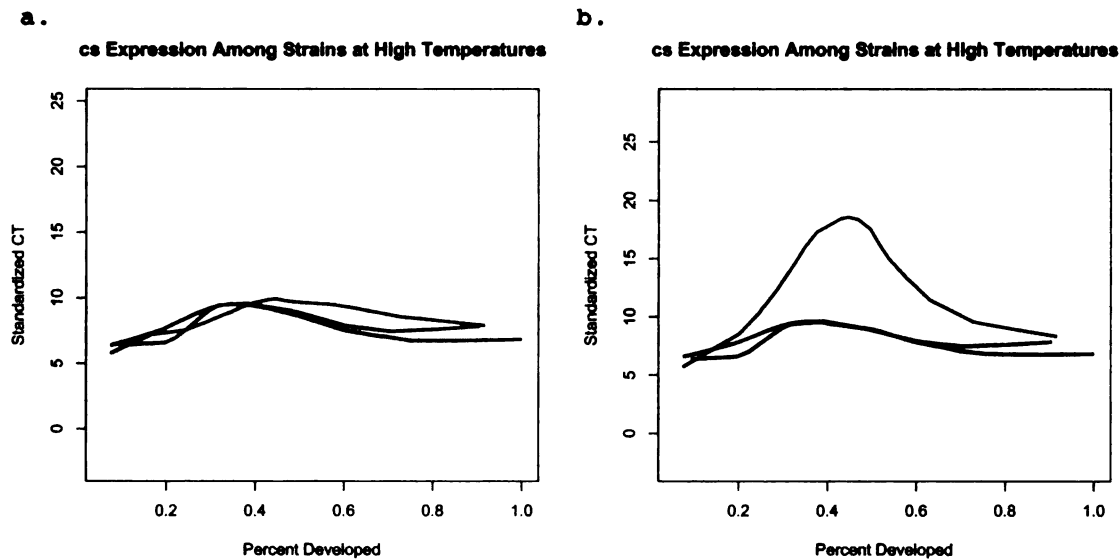


**Figure 5.6. Standardized *chitin synthase* CT scores plotted by stage.** The postfeeding third instar stands out as expressing significantly less *cs* than the other stages.

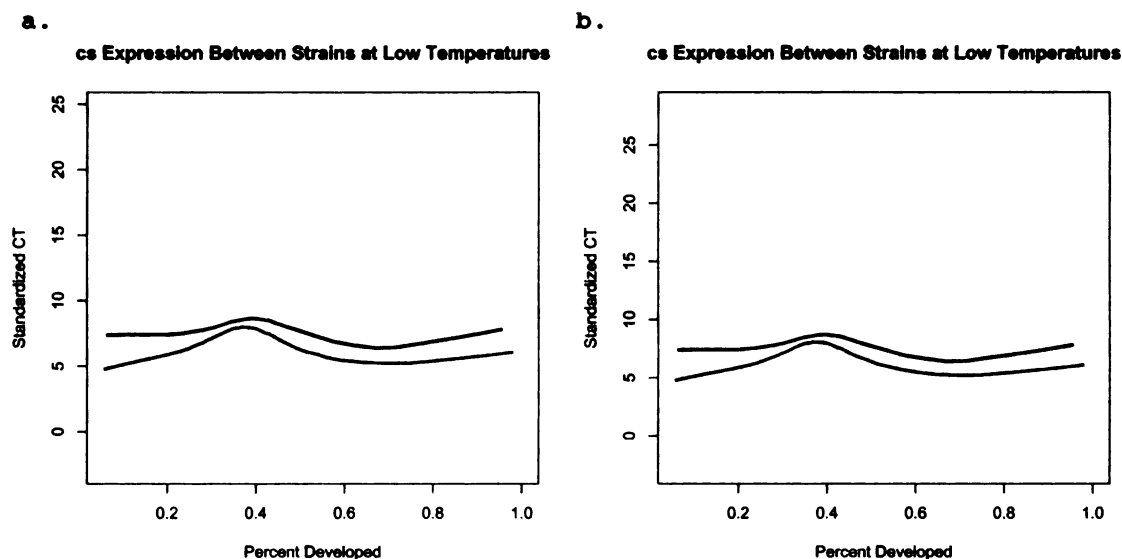


**Figure 5.7. Standardized *chitin synthase* CT scores plotted by development percent for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The shift in the red line between (a) and (b) indicates a cluster of individuals that did not express the gene at the time point where gene expression is at its lowest levels for this locus.**





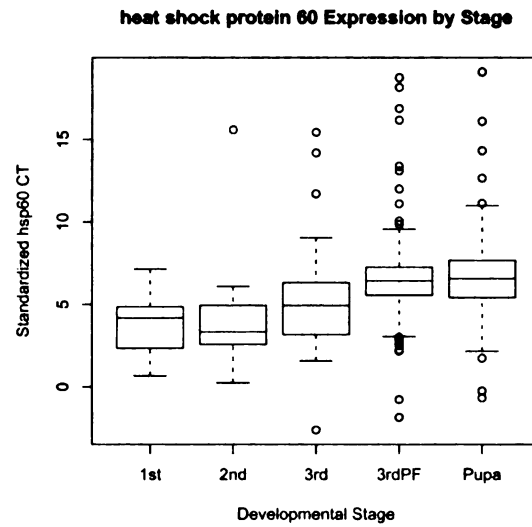
**Figure 5.8. Standardized *chitin synthase* CT scores plotted by development percent at 33.5°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The shift in the red line between (a) and (b) indicates a cluster of individuals that did not express the gene at that time point.**



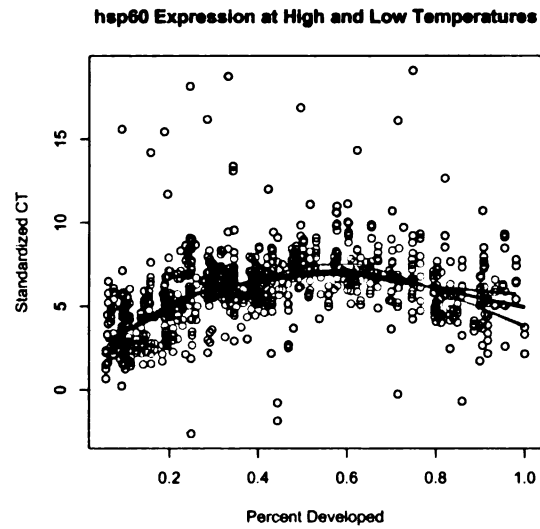
**Figure 5.9. Standardized *chitin synthase* CT scores plotted by development percent at 20°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. Few individuals did not express the gene at this temperature. The CA strain expressed more of *cs* than the MI strain, though in the same pattern.**

Plots of *hsp60* expression are in Figures 5.10–13. *hsp60* was not detected in 106 samples. When grouped by stage, *hsp60* was expressed at its highest levels in the first two instars, at intermediate levels during the feeding portion of the third instar, and at its lowest levels during the postfeeding third instar and pupal stages (Figure 5.10). When plotted against development percent, *hsp60* expression decreased from hatching until early pupation, and then increased in abundance until eclosion. Expression of *hsp60* was not affected by temperature (Figure 5.11). However, the CA strain expressed the gene at higher levels (though in the same pattern) than the MI strain at 20°C (Figure 5.13). While the sample size was small, at 33.5°C there was no obvious difference in expression among the strains, but the CA strain did express the gene at relatively high levels around 40 percent development (Figure 5.12). The gene was a significant predictor of

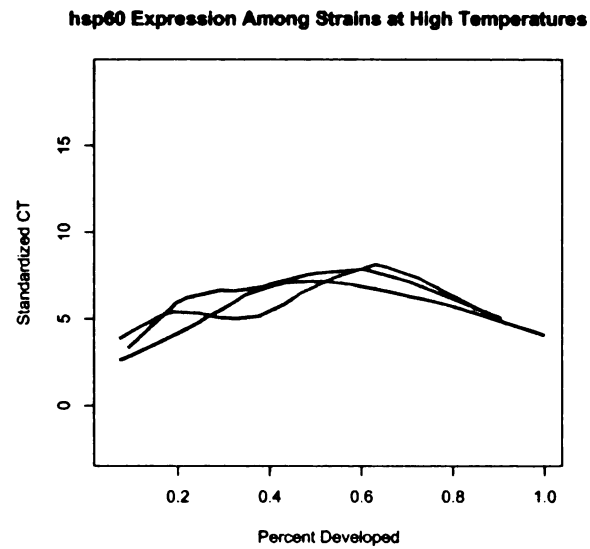
development percent in the GAM assessing its utility (Table 5.4). The GAM that included the *hsp60* term (Model 7) exhibited a PDE of 14.7% (Table 5.3).



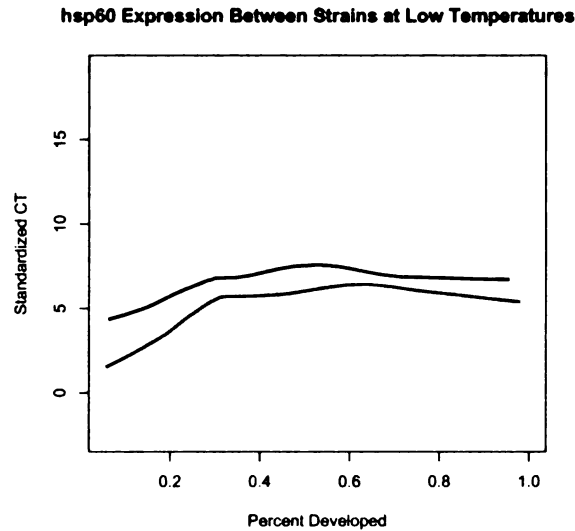
**Figure 5.10. Standardized *heat shock protein 60* CT scores plotted by stage.** Gene expression was highest during the first two stages and lowest during the last two stages.



**Figure 5.11. Standardized *heat shock protein 60* CT scores plotted by development percent for all individuals that expressed the gene.** The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The expression levels of this gene decrease from hatching, until 60 percent development (early pupation), then increase until eclosion.



**Figure 5.12. Standardized *heat shock protein 60* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The strains expressed the gene in the same general pattern, though the CA strain had higher expression around 40 percent development.

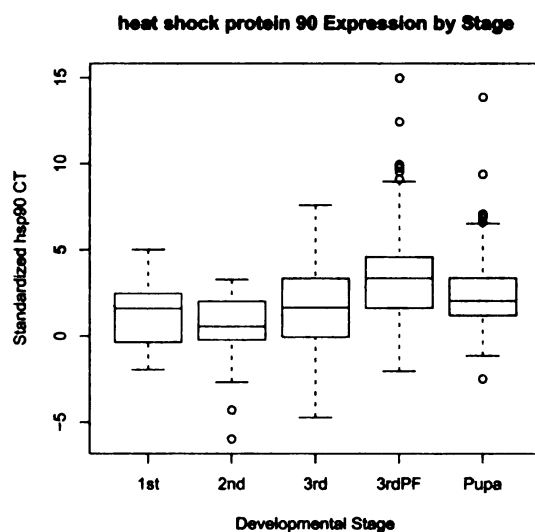


**Figure 5.13. Standardized *heat shock protein 60* CT scores plotted by development percent at 20°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more of this gene than the MI strain, at this temperature. Few individuals did not express the gene at this temperature.

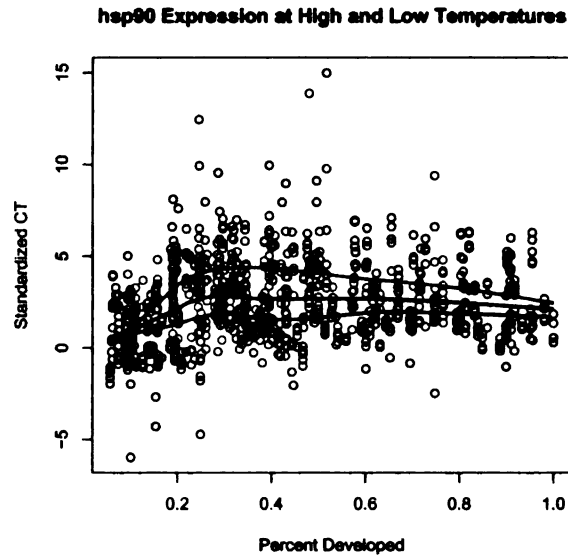
Plots of *hsp90* expression are in Figures 5.14–17. *hsp90* was not detected in 11 samples. When grouped by stage, *hsp90* was expressed at stable levels except during the feeding portion of the third instar, when it was expressed at lower levels than the other stages (Figure 5.14). When plotted against development percent, *hsp90* expression demonstrated different expression patterns depending on the temperature at which individuals were raised (Figure 5.15). During the first quarter of development (approximately) there was a decrease in transcript abundance, which was much more pronounced at 33.5°C than at 20°C. At the high temperature transcript abundance increases until eclosion, but at the lower temperature expression was maintained at a constant level. At 33.5°C, there was no obvious difference in expression among strains, though the point of minimal *hsp90* expression occurred at a later time in the CA strain than it did in the other two (Figure 5.16). At 20°C both strains demonstrated a similar

pattern of *hsp90* expression, with the CA strain expressing the gene at elevated levels.

The gene was a significant predictor of development percent when included in a GAM assessing development percent with CT scores for *hsp90* along with temperature and developmental stage, which were also significant terms in that model (Table 5.4). The GAM (Model 8) exhibited a PDE increase of 0.6% compared to Model 1 (Table 5.3).

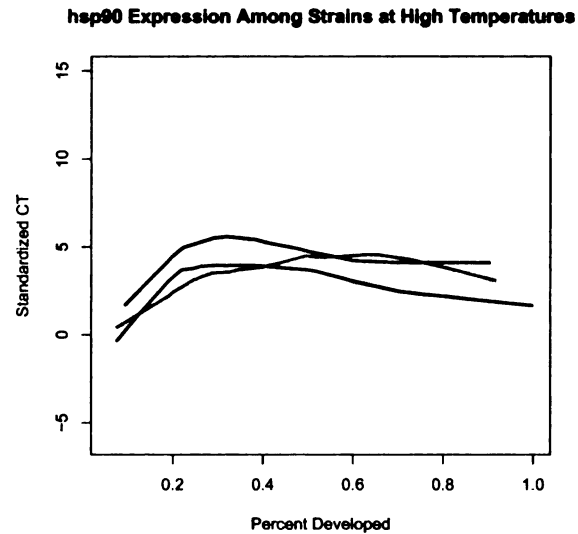


**Figure 5.14. Standardized *heat shock protein 90* CT scores plotted by stage. Gene expression was lowest during the postfeeding third instar.**

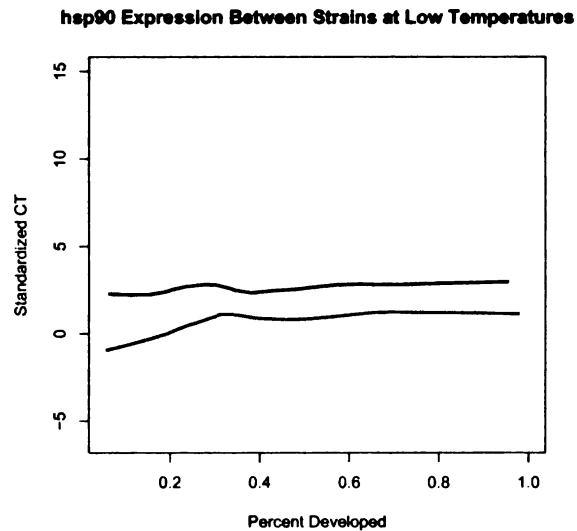


**Figure 5.15. Standardized *heat shock protein 90* CT scores plotted by development percent for all individuals that expressed the gene.** The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The expression levels of the gene decreased from hatching, until approximately 25 percent development (when postfeeding larval development is attained), then increased until eclosion when flies were raised at the high temperature. At the low temperature, expression was higher than at high temperatures, and expression was maintained at a steady level after 25 percent development.





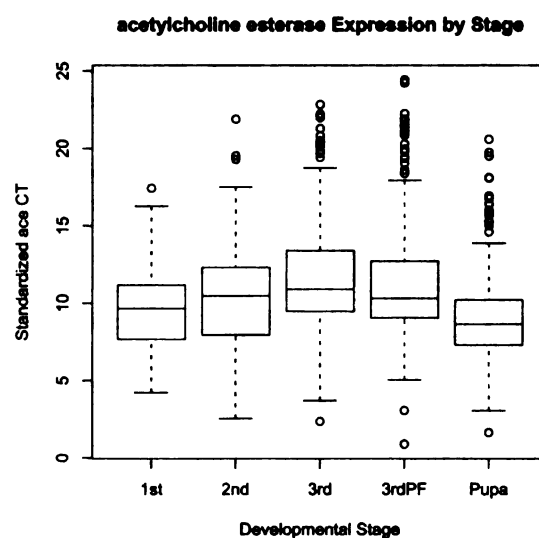
**Figure 5.16. Standardized *heat shock protein 90* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. Expression decreased until the postfeeding third instar (around 25 percent) then increased until eclosion. The strains expressed the gene in the same general pattern, though the time of minimum expression occurred later in the CA strain.



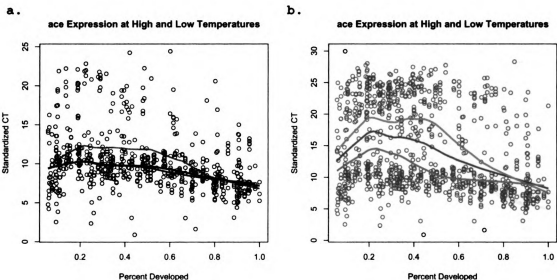
**Figure 5.17. Standardized *heat shock protein 90* CT scores plotted by development percent at 20°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more of this gene than the MI strain, at this temperature.

Plots of *ace* expression are in Figures 5.18–21. *ace* was not detected in 322 samples. When grouped by stage, *ace* was expressed in decreasing levels through the third instar, and then it increased in expression until eclosion (Figure 5.18). When plotted against development percent, *ace* expression demonstrated two different expression patterns depending on the temperature at which individuals were raised (Figure 5.19). During the first quarter of development (approximately) there was a decrease in transcript abundance, which was much more pronounced at 33.5°C than at 20°C. At 33.5°C transcript abundance increased until eclosion, but at the lower temperature expression was maintained at a constant level. At 33.5°C there was no obvious difference in expression among the strains. Though sample sizes per strain were small, the point of minimal *ace* expression occurred at a later time in the CA strain than it did in the other two strains (Figure 5.20). In the 20°C replicates, both strains

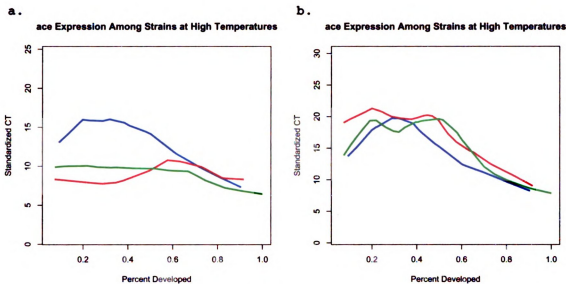
demonstrated a similar pattern of expression, with the CA strain expressing the gene at elevated levels (Figure 21). In a GAM (Model 9), *ace* was a significant predictor of development percent when temperature and developmental stage were included in the model. The binary term was significant by itself (Table 5.4). The non-parametric GAM (Model 9) exhibited a PDE increase of 1.7% compared to Model 1 (Table 5.3). The GAM that included binary *ace* expression (Model 3) demonstrated a PDE of 3.6% (Table 5.3).



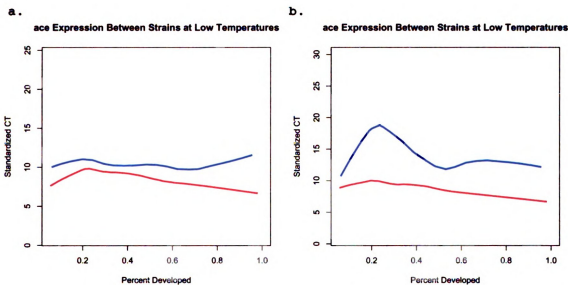
**Figure 5.18. Standardized *acetylcholine esterase* CT scores plotted by stage. Gene expression decreased through the third instar, then increased until eclosion.**



**Figure 5.19. Standardized *acetylcholine esterase* CT scores plotted by development percent for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The black line is the lowest curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The shift in the lines between (a) and (b) indicates a cluster of individuals that do not express the gene at the time point where gene expression is at its lowest levels for this locus. Many individuals did not express *ace* between approximately 20 and 50 percent development (late third instar through early pupation). This time period mirrors the period of least *ace* expression at the high temperature treatment. Expression increased through development in the low temperature treatments.**

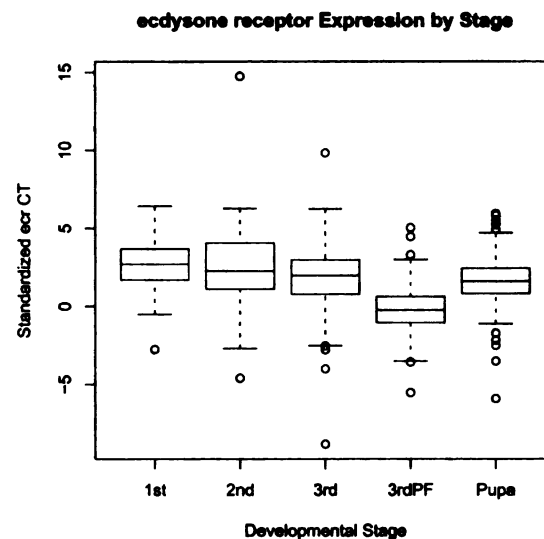


**Figure 5.20. Standardized *acetylcholine esterase* CT scores plotted by development percent at 33.5°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The shift in these lines between (a) and (b) indicates a cluster of individuals that do not express the gene at that time point.**

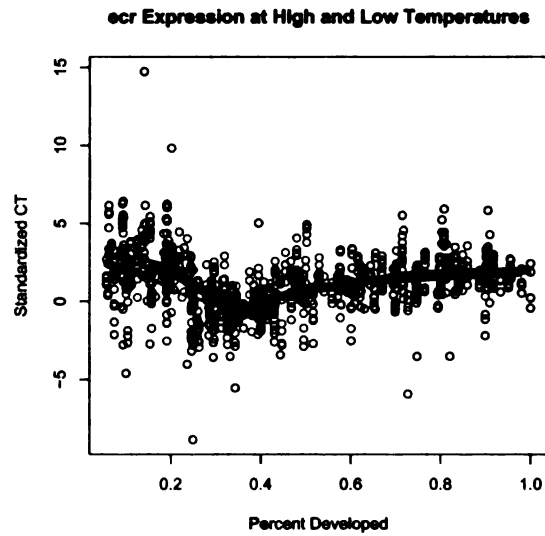


**Figure 5.21. Standardized *acetylcholine esterase* CT scores plotted by development percent at 20°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more *ace* than the MI strain (a). MI was also more likely to not express *ace*.**

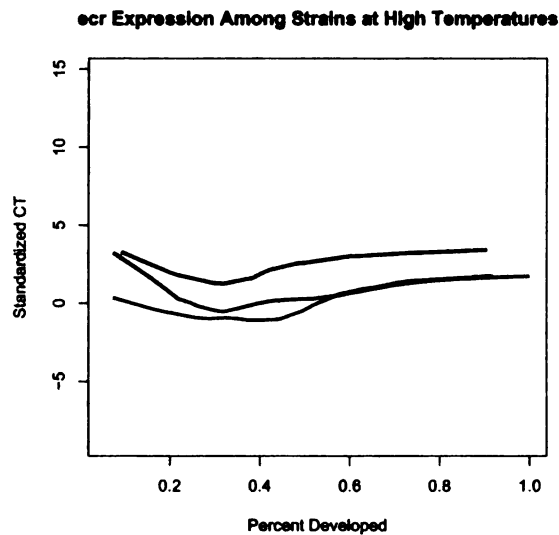
Plots of *ecr* expression are in Figures 5.22–25. *ecr* was not detected in 17 samples. When grouped by stage, *ecr* was expressed in slightly increasing levels until the postfeeding third instar, when expression was significantly greater than other stages, then expression decreased through pupation until eclosion (Figure 5.22). When plotted against development percent, *ecr* expression increased until achieving maximum expression at approximately 35 percent, then expression decreased. This pattern was the same for cohorts raised at both temperatures (Figure 5.23). However, at 33.5°C, the MI strain expressed less *ecr* than the other strains (Figure 5.24) and at 20°C the MI strain expressed less *ecr* than the CA strain, except at the point of maximum expression, when both strains converged to the same expression level for this gene (Figure 5.25). A GAM assessing age with *ecr* CT scores was a significant predictor of development percent (Table 5.4). This GAM (Model 10) exhibited a PDE of 6.6% (Table 5.3).



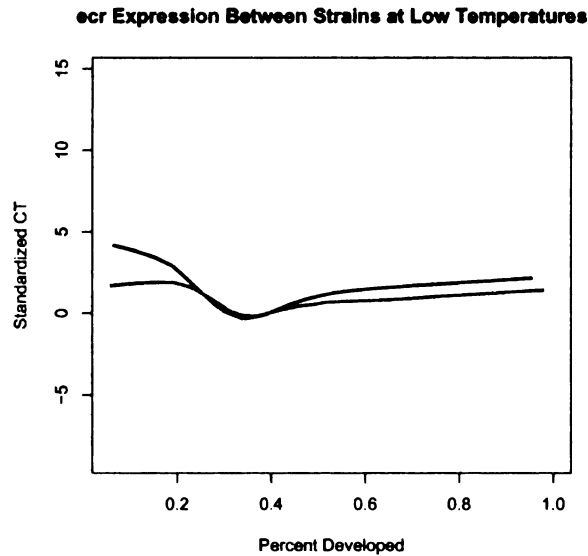
**Figure 5.22. Standardized *ecdysone receptor* CT scores plotted by stage.** Gene expression increased through the postfeeding third instar, then decreased until eclosion.



**Figure 5.23. Standardized *ecdysone receptor* CT scores plotted by development percent for all individuals that expressed the gene.** The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The expression levels of this gene increase from hatching, until approximately 35 percent development (during the postfeeding third instar), then decrease until eclosion.



**Figure 5.24. Standardized *ecdysone receptor* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The strains expressed the gene in the same general pattern, though the MI strain expressed less *ecr* than the other strains.

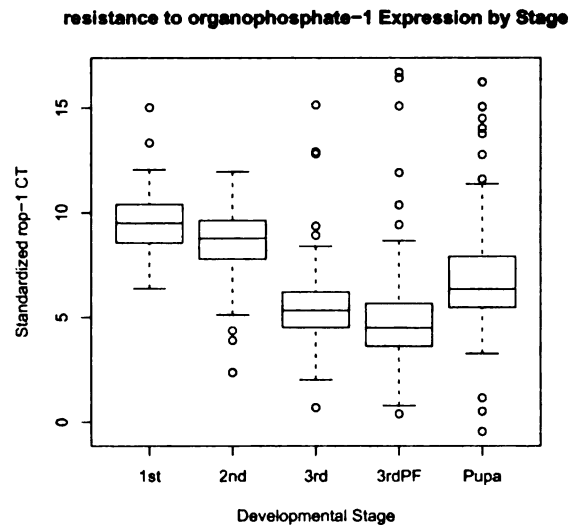


**Figure 5.25. Standardized *ecdysone receptor* CT scores plotted by development percent at 20°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more of this gene than the MI strain, at this temperature, though they converge to the same maximum expression level during the postfeeding third instar.

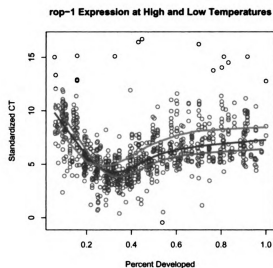
Plots of *rop-1* expression are in Figures 5.26–29. Of the 958 individuals assessed, *rop-1* was not detected in 99 samples. When grouped by stage, *rop-1* was expressed in increasing levels through the postfeeding third instar, then expression decreased through pupation until eclosion (Figure 5.26). When plotted by development percent, *rop-1* increased in abundance until approximately 25 percent development (when raised at 33.5°C), or until 35 percent development (when raised at 20°C) (Figure 5.27). The shift in these curves closely mirrored the shift in body size (seen in Figure 3.3), which did not occur with all genes. After achieving maximum expression, the abundance of message decreased until eclosion. At high temperatures, there was no obvious difference in expression of *rop-1* among the strains (Figure 5.28). However, at the low temperature treatment, expression was consistently higher in the CA strain compared to the MI strain (Figure 5.29). In the GAM assessing age with *rop-1* CT



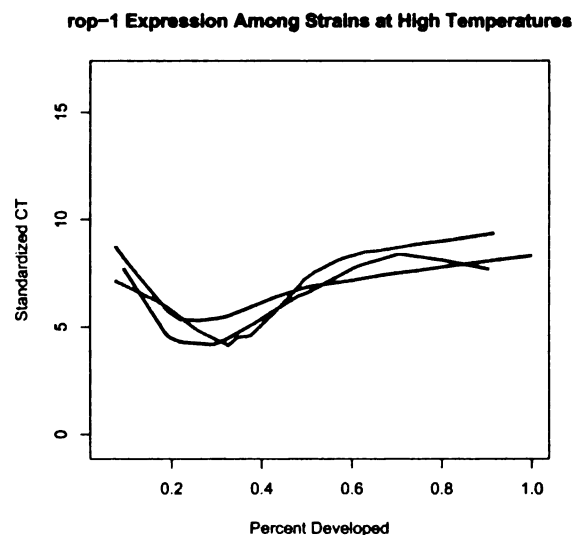
scores, the gene was a significant predictor of development percent when included with stage (Table 5.4). Though both terms were significant predictors of development percent, the GAM (Model 11) exhibited no PDE increase when compared to Model 1. However, the GCV score was lower for Model 11 when compared to Model 1 (Table 5.3), indicating that the inclusion of *rop-1* will help make better estimates of development percent.



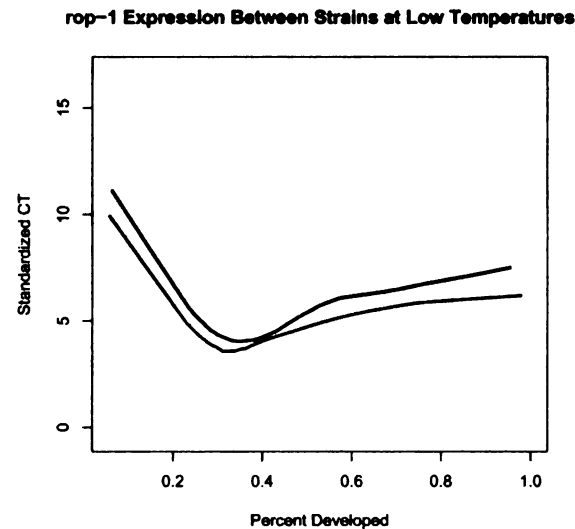
**Figure 5.26. Standardized *resistance to organophosphate 1* CT scores plotted by stage.** Gene expression increased through the postfeeding third instar, then decreased until eclosion.



**Figure 5.27. Standardized *resistance to organophosphate 1* CT scores plotted by development percent for all individuals that expressed the gene.** The black line is the lowess curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The expression levels of this gene increase from hatching until approximately 25 percent development at high temperatures and until approximately 35 percent development at low temperatures, then gene transcript abundance decrease until eclosion.



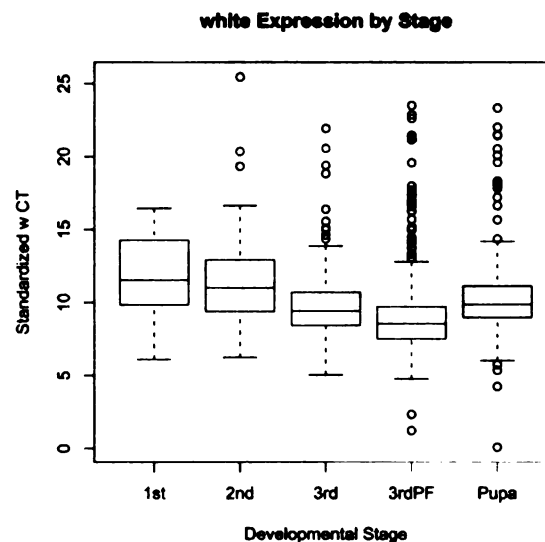
**Figure 5.28. Standardized *resistance to organophosphate 1* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The strains expressed the gene in the same general pattern.



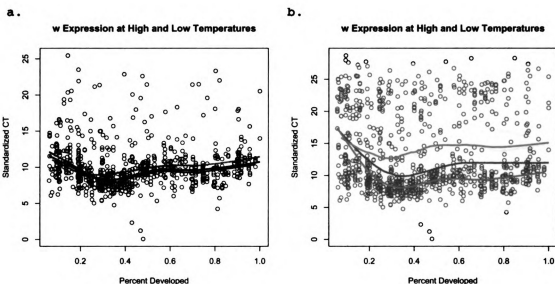
**Figure 5.29. Standardized *resistance to organophosphate 1* CT scores plotted by development percent at 20°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more of the gene than the MI strain, at this temperature, though both followed the same pattern.

Plots of *w* expression are in Figures 5.30–33. *w* was not detected in 233 samples. When grouped by stage, *w* was expressed in increasing amounts from the first instar through the postfeeding third instar, then expression decreased in pupae (Figure 5.30). When plotted by development percent, *w* increased in abundance until approximately 25 percent development (when raised at 33.5°C), or until 35 percent development (when raised at 20°C) (Figure 5.31), though this pattern was subtler than in the *rop-1* plots. Individuals from the low temperature treatments were also more likely to express *w* than individuals from high temperature treatments (Figure 5.31). After achieving maximum expression, the abundance of the transcript decreased until eclosion. At high

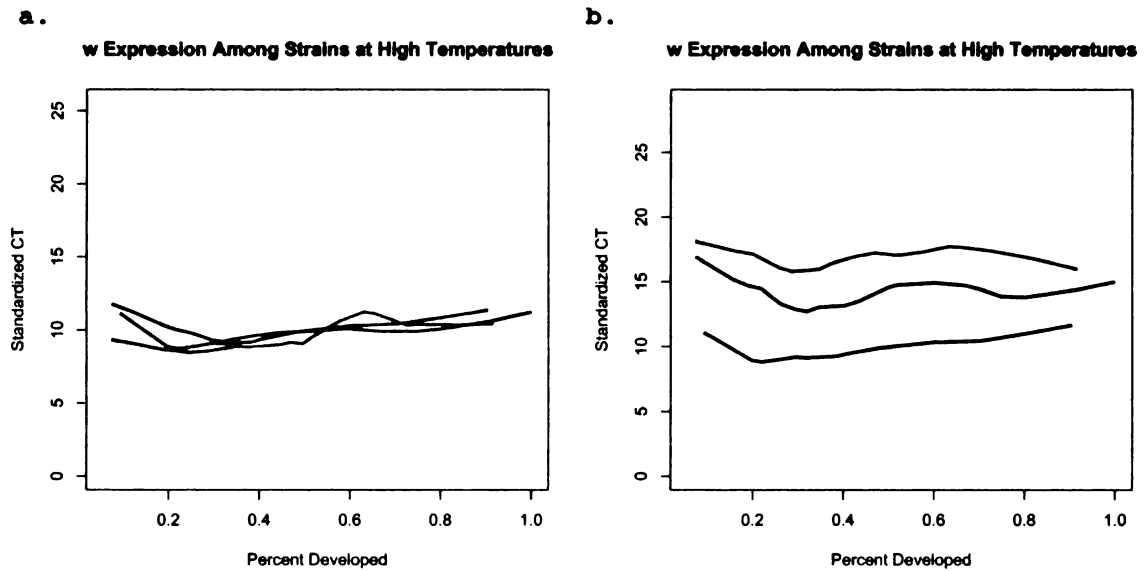
temperatures, there was no obvious difference in expression of *w* among the strains (Figure 5.32), but the CA strain was the least likely to express the gene at that temperature and the MI strain was the most likely to express it. However, at the low temperature treatment, expression was consistently higher in the CA strain compared to the MI strain, with the MI strain likely to not express the gene during pupation (Figure 5.33). In GAMs assessing both the binary and non-parametric expression of *w*, the gene was a significant predictor of development percent (Table 5.4). Binary expression (Model 4) explained 0.43% of the deviance in the data. When *w* CT scores were used to predict development percent (Model 12), a PDE of 3.55% was attained (Table 5.3).



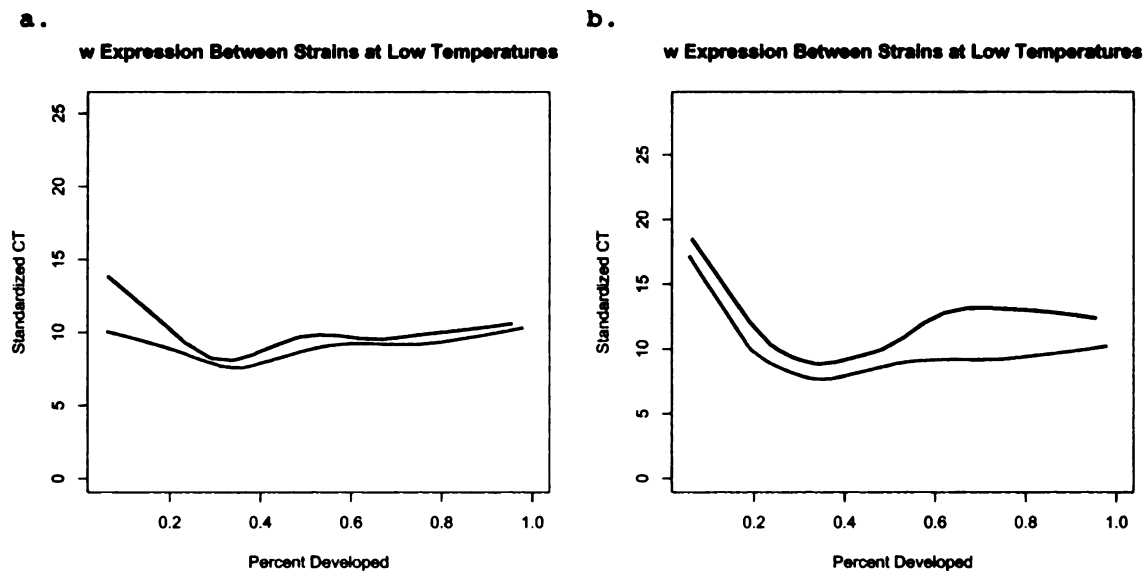
**Figure 5.30. Standardized *white* CT scores plotted by stage.** Gene expression increased through the postfeeding third instar, then decreased until eclosion.



**Figure 5.31. Standardized *white* CT scores plotted by development percent for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50.** The black line is the lowest curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The shift in the lines between (a) and (b) indicates a cluster of individuals that did not express the gene at the time point where gene expression is at its lowest levels for this locus. The expression levels of the gene increased from hatching until approximately 25 percent development at high temperatures and until approximately 35 percent development at low temperatures, then gene transcript abundance decrease until eclosion. High temperature treatments were less likely to express the gene.



**Figure 5.32. Standardized *white* CT scores plotted by development percent at 33.5°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The shift in the lines between (a) and (b) indicates that CA was most likely to not express *w* and MI was most likely to express the gene.**

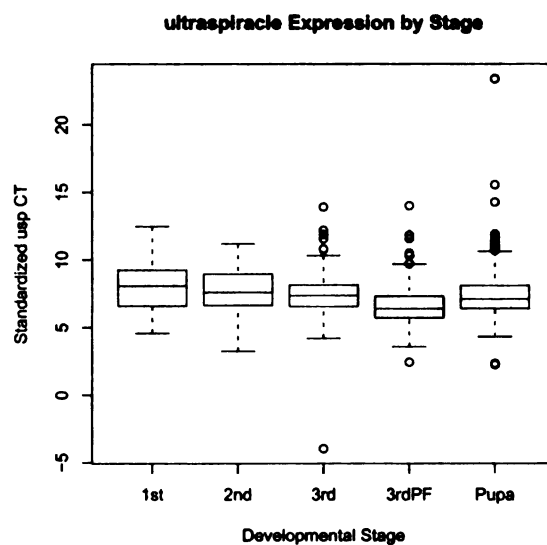


**Figure 5.33. Standardized *white* CT scores plotted by development percent at 20°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more *w* than the MI strain (a). MI was also more likely to not express *w* during pupation (the last half of development).**

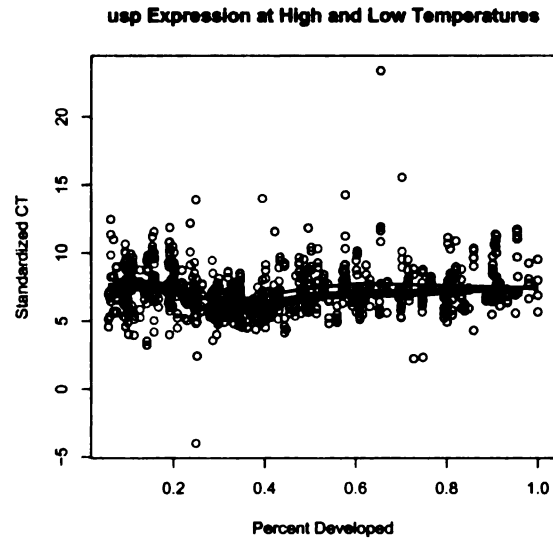
Plots of *usp* expression are in Figures 5.34–37. *usp* was not detected in 109 samples. When grouped by stage, *usp* was expressed in increasing amounts from the first instar through the postfeeding third instar, and then expression decreased in pupae (Figure 5.34). When plotted by development percent, *usp* increased in concentration from hatching until the postfeeding third instar, then expression decreased through pupation (Figure 5.35), with no obvious difference in expression among temperature treatments. The three strains exhibited little difference in expression at 33.5°C (Figure 5.36). However, at 20°C MI and CA expressed *usp* in a similar pattern, with CA consistently expressing more of the gene than the MI strain. A GAM assessing expression of *usp* was a significant predictor of development percent when stage was



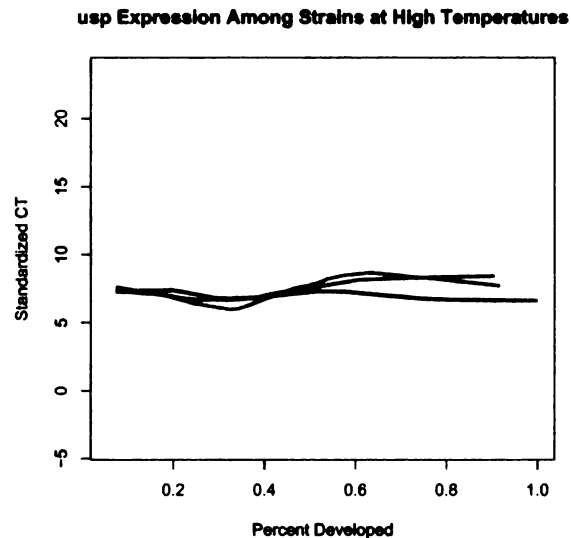
included in the model (Table 5.4). When *usp* CT scores were used to predict development percent (Model 13), a PDE increase of 0.6% was attained (Table 5.3).



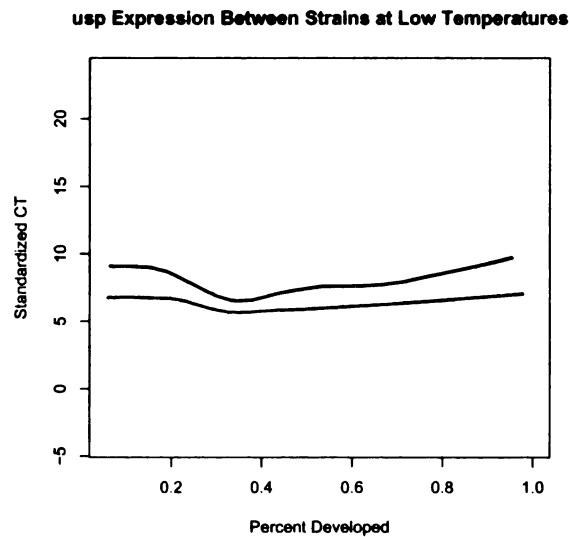
**Figure 5.34. Standardized *ultraspiracle* CT scores plotted by stage.** Gene expression increased through the postfeeding third instar, then decreased until eclosion.



**Figure 5.35. Standardized *ultraspiracle* CT scores plotted by development percent for all individuals that expressed the gene.** The black line is the lowest curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The expression levels of the gene increased from hatching until approximately 35 percent development, then gene transcript abundance decreased until eclosion.



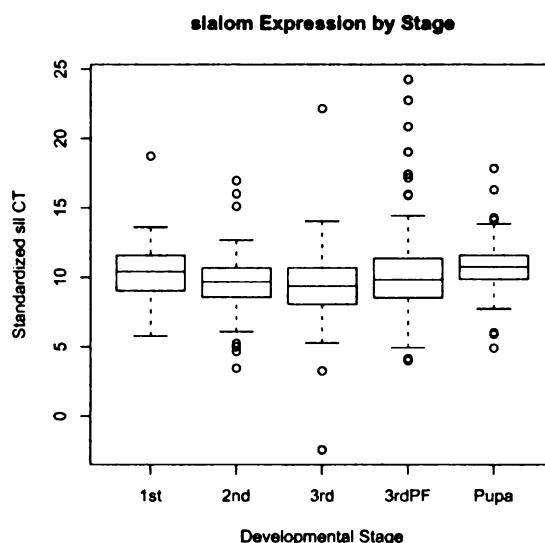
**Figure 5.36. Standardized *ultraspiracle* CT scores plotted by development percent at 33.5°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The strains expressed the gene in the same general pattern.



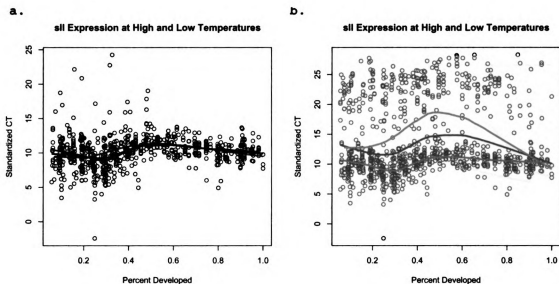
**Figure 5.37. Standardized *ultraspiracle* CT scores plotted by development percent at 20°C for all individuals that expressed the gene.** The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The CA strain expressed more of the gene than the MI strain, though both followed a similar pattern.

Plots of *sll* expression are in Figures 5.38–41. Of the 958 individuals sampled, 281 did not express the gene. When plotted by stage, expression increased through the feeding portion of development, then decreased until eclosion (Figure 5.38). When plotted by development percent, expression increased until approximately 25 percent development, and then decreased until approximately 50 percent development, when expression levels increased again until eclosion (Figure 5.39). There was little difference in expression patterns between temperatures, but individuals from high temperature treatments were much less likely to express the gene (Figure 5.39). When expression at the high temperature was assessed, it was clear that the lack of expression occurred in both the CA and WV strains (Figure 5.40). The expression of *sll* at low temperatures was similar among the CA and MI strains, but MI was much less likely to express the gene

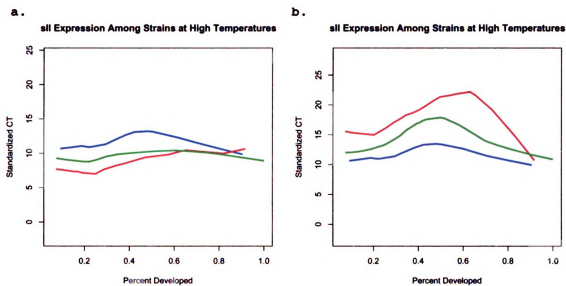
(Figure 5.41). When GAMs utilized *sll* expression to predict development percent, both the binary and quantitative expression levels of the gene were significant predictors of age (Table 5.4). The GAM that utilized standardized *sll* CT scores (Model 14) exhibited a PDE of 9.17%. The GAM using the binary expression data (Model 5) required the assessment of stage, strain, and temperature with *sll* to achieve an increase of 0.3% in PDE compared to Model 1 (Table 5.3).



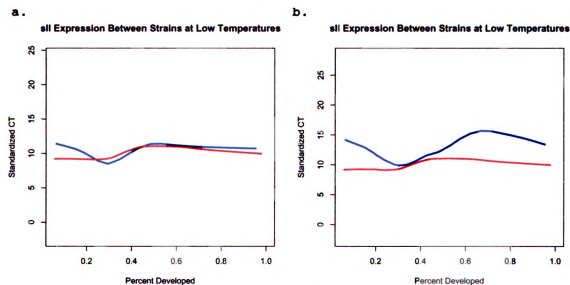
**Figure 5.38. Standardized *slalom* CT scores plotted by stage.** Gene expression increased through the third instar, then decreased until eclosion.



**Figure 5.39. Standardized *slalom* CT scores plotted by development percent for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50.** The black line is the lowest curve for all individuals, the blue line is the curve for expression at 20°C, and the red line is the curve for expression at 33.5°C. The shift in the lines between (a) and (b) indicates a cluster of individuals that did not express the gene at the time point where gene expression was at its lowest levels. The expression levels of the gene increased from hatching until approximately 25 percent development, then gene transcript abundance decreased until approximately 50 percent, and then increase again until eclosion. High temperature treatments were less likely to express the gene.



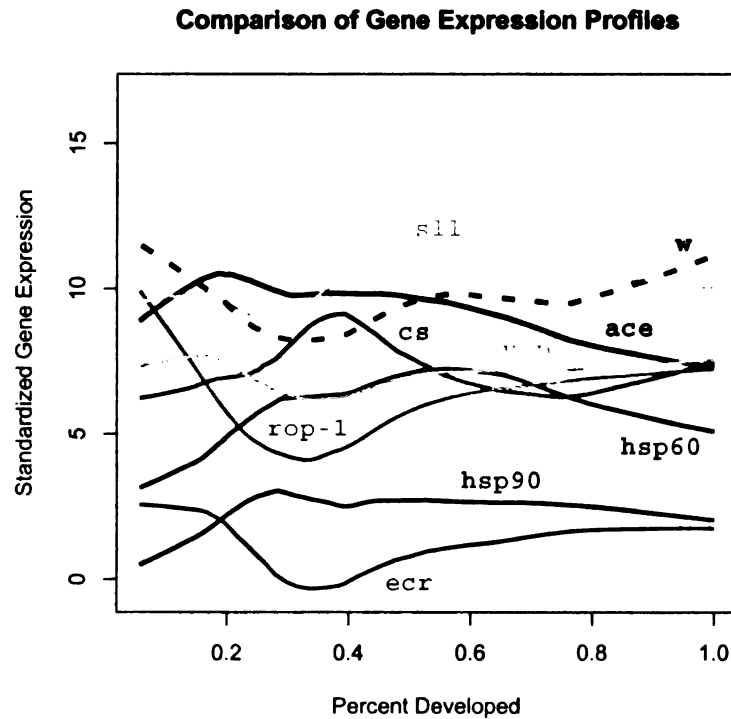
**Figure 5.40. Standardized *slalom* CT scores plotted by development percent at 33.5°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The green line indicates gene expression in the WV strain. The shift in the lines between (a) and (b) indicates that CA was most likely to not express *sl* and MI was most likely to express the gene.**



**Figure 5.41. Standardized *slalom* CT scores plotted by development percent at 20°C for (a) all individuals that expressed the gene and (b) in all 958 individuals, with NA expression values assigned a CT of 50. The red line indicates gene expression in the CA strain. Blue indicates gene expression in the MI strain. The strains expressed the gene in a similar pattern (a). MI was more likely to not express *sl*.**

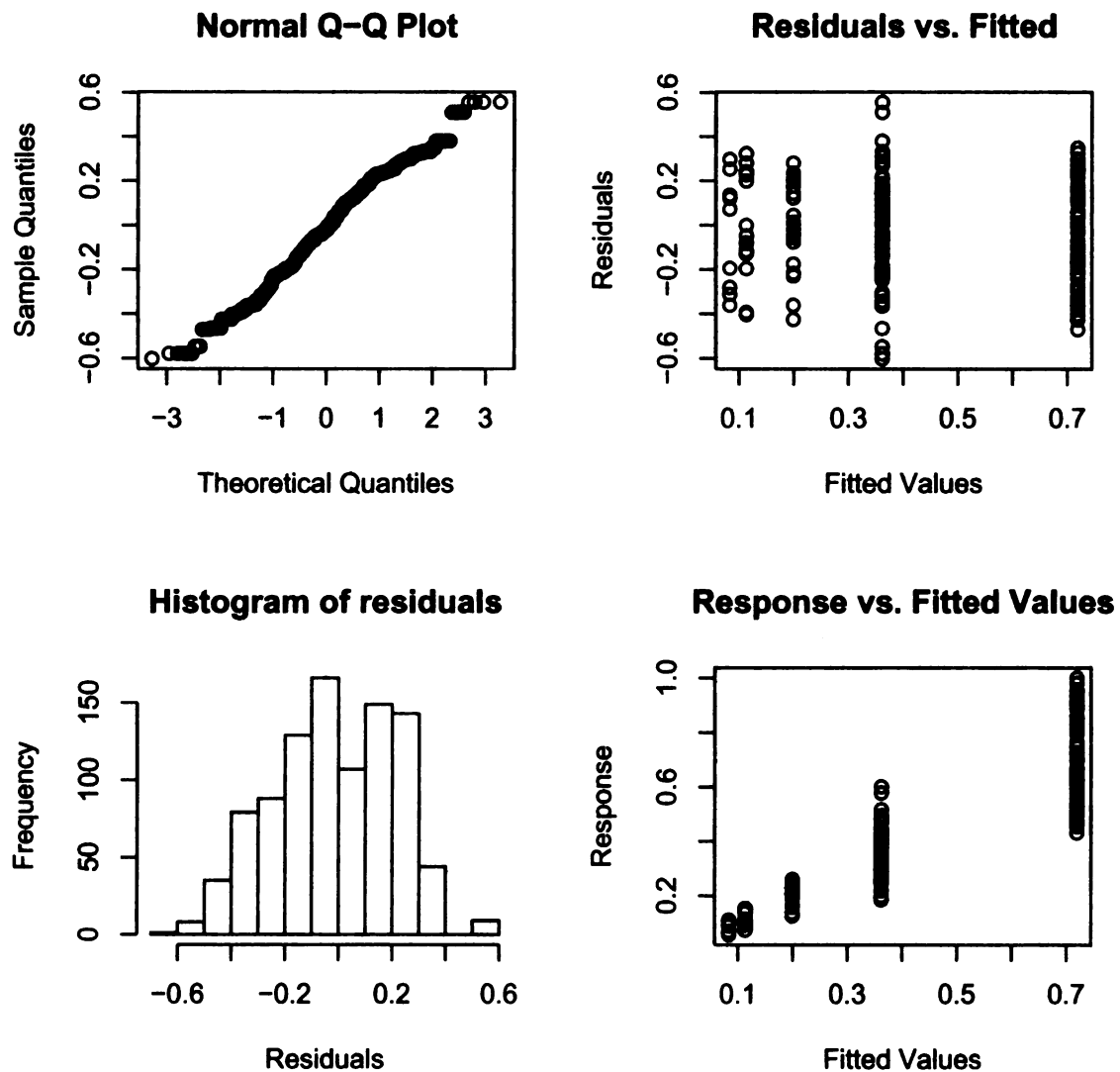
The expression profiles of all developmentally regulated genes are shown in Figure 5.42. All GAMs that included expression data from multiple genes (Models 16–23) were found to have lower GCV scores than the control models. Likewise, all but one model (Model 19 which used a gaussian distribution with gene expression data) exhibited an increase in PDE compared to the control models (92.1%– 95.7% PDE compared to 88.2%– 91.8%; Table 3). Model 19 had a PDE of 91.3%, which was 0.5% lower than Model 15 and 3.1% better than the PDE for Model 1. (See information on specific stages below.)

When diagnostic plots of the models were compared the improvement generated by incorporating gene expression was substantial. Figures 5.43–50 depict diagnostic plots, which assess error for specific models tested in this experiment. The control models (Figures 5.43 and 5.44) show non-gene expression data only, while the rest display control data and some form of gene expression data used to predict development percent. The lower right quadrant of the plots depicts the predicted (Fitted) versus true (Response) development percents of all individuals used to make the model. In model 15 (which used non-gene data, Figure 5.44), error increased throughout development and a gap existed between larval and pupal predictions. All gene expression models decreased error and closed the gap in predictions that were present in Figures 5.43 and 5.44.

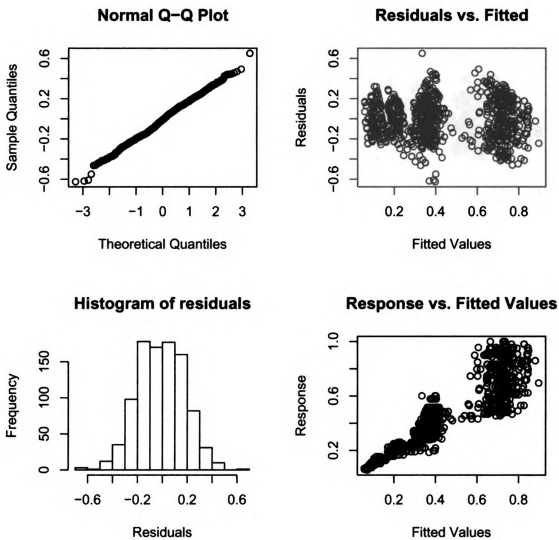


**Figure 5.42. The combined standardized expression (CT score) patterns for all nine genes throughout *L. sericata* immature development.** The lowest curves for each gene represent expression for all individuals that expressed the gene. The combined expression patterns of the genes were used to predict blow fly age with GAMs.

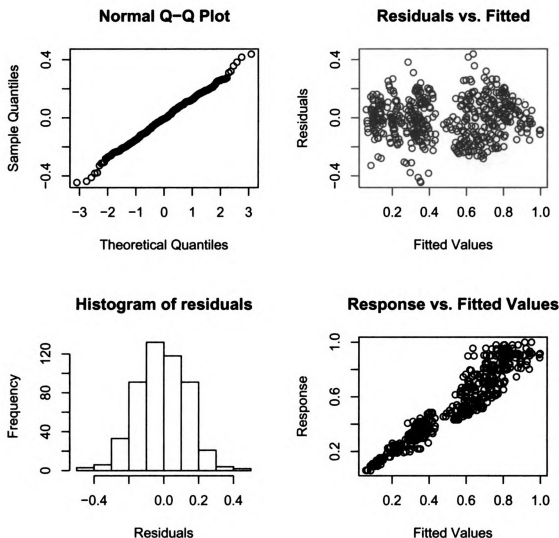




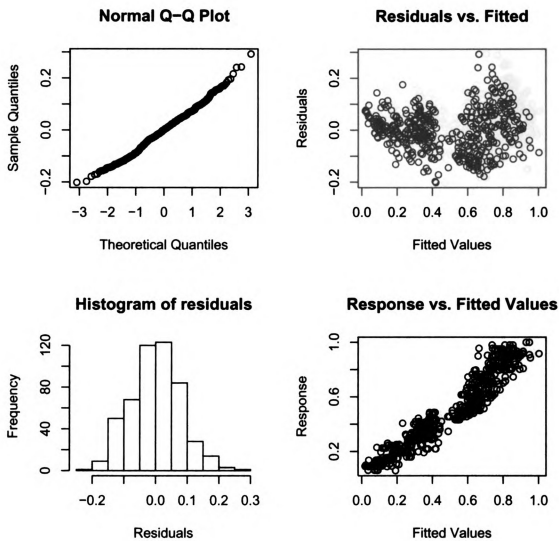
**Figure 5.43. Diagnostic plot for Model 1, which used developmental stage to predict *L. sericata* development percent (on a scale of 0–1). A gamma distribution was used with this model. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. Error increased with age, and gaps exist between predictions for each developmental stage.**



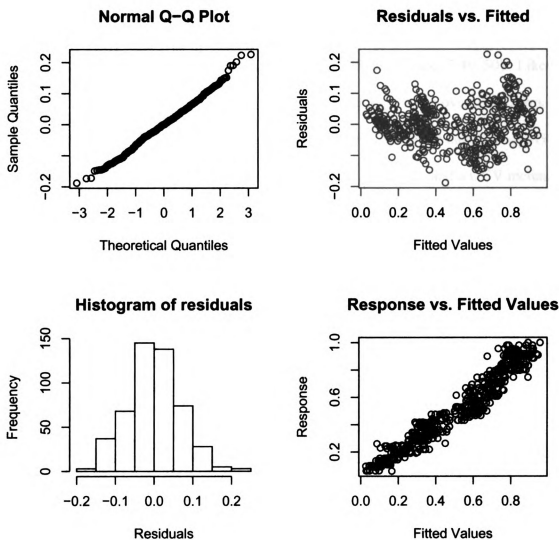
**Figure 5.44. Diagnostic plot for Model 15, which used developmental stage, strain, temperature, length, and weight to predict *L. sericata* development percent (on a scale of 0–1). A gamma distribution was used with this model. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. Error increased with age, and a gap exists between predictions for postfeeding third instar and pupal ages.**



**Figure 5.45. Diagnostic plot for Model 18, which used developmental stage, strain, temperature, length, weight, binary gene expression for four genes, and quantitative gene expression for nine genes to predict *L. sericata* development percent (on a scale of 0–1).** A gamma distribution was used with this model. Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to other models and the gap between predictions for postfeeding third instar and pupal ages has shrunk.

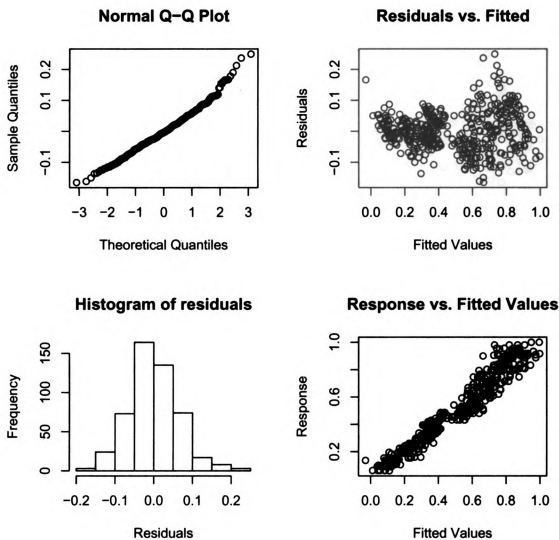


**Figure 5.46. Diagnostic plot for Model 19, which used the same parameters as Model 18 to predict *L. sericata* development percent (on a scale of 0–1), but a gaussian distribution was used.** Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to Model 15 and the gap between predictions for postfeeding third instar and pupal ages has shrunk. However, compared to Model 18, this model exhibits more error in predictions of younger individuals.

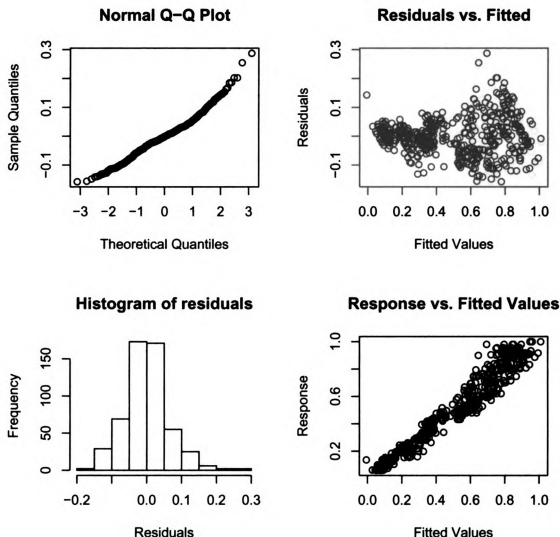


**Figure 5.47. Diagnostic plot for Model 20, which used the same parameters as Model 18 to predict *L. sericata* development percent (on a scale of 0–1), but a gaussian distribution was used with this model and all genes expression scores were anchored against *hsp60* expression.** Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to Model 15 and the gap between predictions for postfeeding third instar and pupal ages has been eliminated. However, compared to Model 18, this model exhibits more error in predictions for younger individuals, though it is an improvement over predictions made with Model 19.

The last two models were similar to Model 21, but strain, temperature, and non-significant terms were removed. When compared to the models that preceded them, there was little change in the diagnostic plots (Figure 5.48–49 and Figures 5.49–50). Likewise, the statistical evaluations of the models revealed little change in the overall performance of the models compared to the model that preceded them. Model 22 had PDE and GCV scores of 94.6% and 0.0059, which represented no change in PDE and a GCV increase of 0.0003 compared to Model 21. Likewise, the removal of temperature (Model 23 compared to Model 22) resulted in PDE and GCV scores of 94.7% and 0.0059 for Model 23, which represented a 0.1% increase in PDE and no change in GCV compared to Model 22.

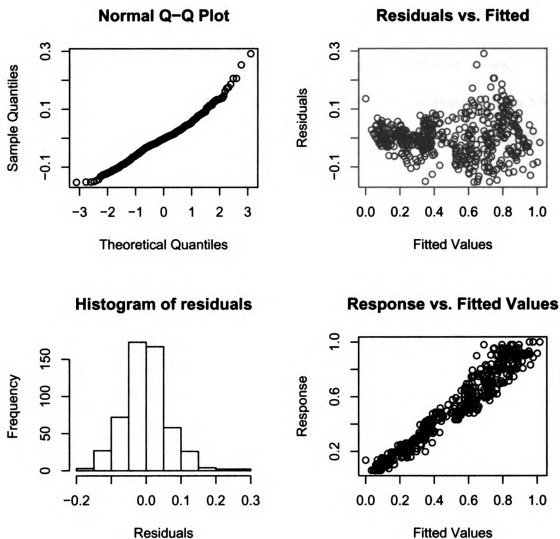


**Figure 5.48. Diagnostic plot for Model 21, which used the same parameters as Model 18 to predict *L. sericata* development percent (on a scale of 0–1), but a gaussian distribution was used with this model and all genes expression scores were anchored against length measurements.** Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to Model 15 and the gap between predictions for postfeeding third instar and pupal ages has been eliminated. However, compared to Model 18, this model exhibits more error in predictions for younger individuals, though it is an improvement over predictions made with Model 19.



**Figure 5.49. Diagnostic plot for Model 22, which used the same parameters as Model 21 to predict *L. sericata* development percent (on a scale of 0–1), but parameters that were non-significant predictors of age in Model 21 (length, weight, cs, w, strain) were removed.** Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. The increase in error with age has diminished compared to Model 15 and the gap between predictions for postfeeding third instar and pupal ages has been eliminated. However, compared to Model 18 this model exhibits more error in predictions for younger individuals, though it is an improvement over predictions made with Model 19. Removing strain, cs, and w, had little effect on the predictions made with the model compared to predictions made with Model 21.

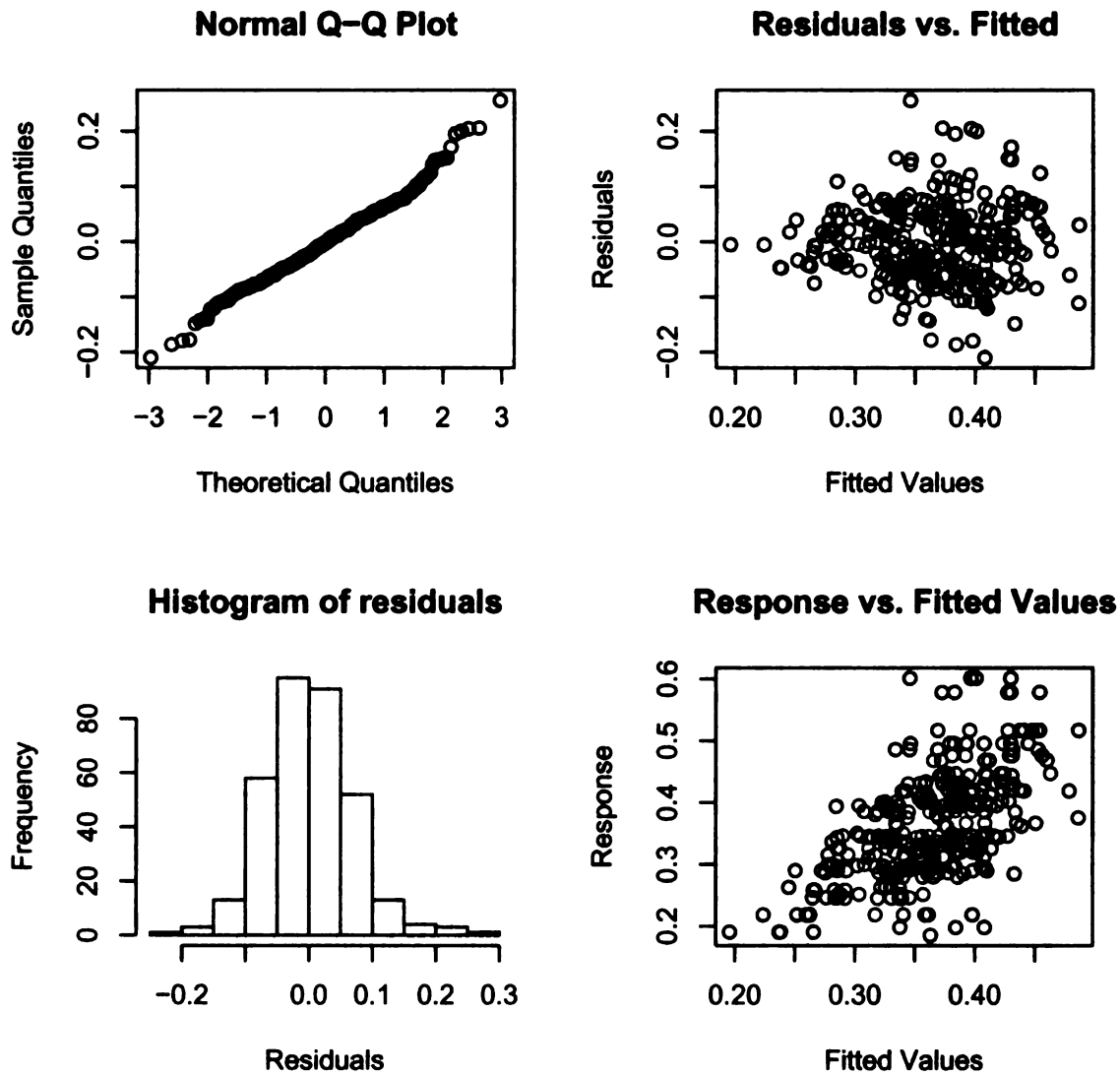




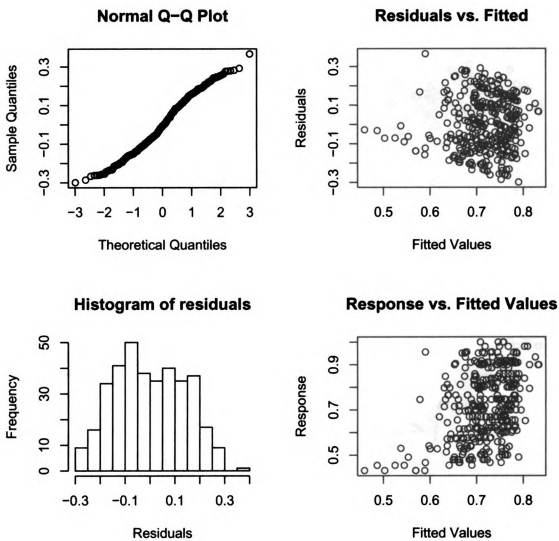
**Figure 5.50. Diagnostic plot for Model 23, which used the same parameters as Model 22 to predict *L. sericata* development percent (on a scale of 0–1), except that temperature was removed from this model.** Descriptions for each panel type can be found in Figure 3.4. Predicted (Fitted) values represent a range of true (Response) values. Removing temperature had little effect on the predictions made with this model compared to predictions made with Model 22.

Finally, the differences between GAMs with or without gene expression data were examined for the most difficult to age stages, the postfeeding third instar and pupation. Here the influence of expression data was most vivid. As mentioned, there is very little

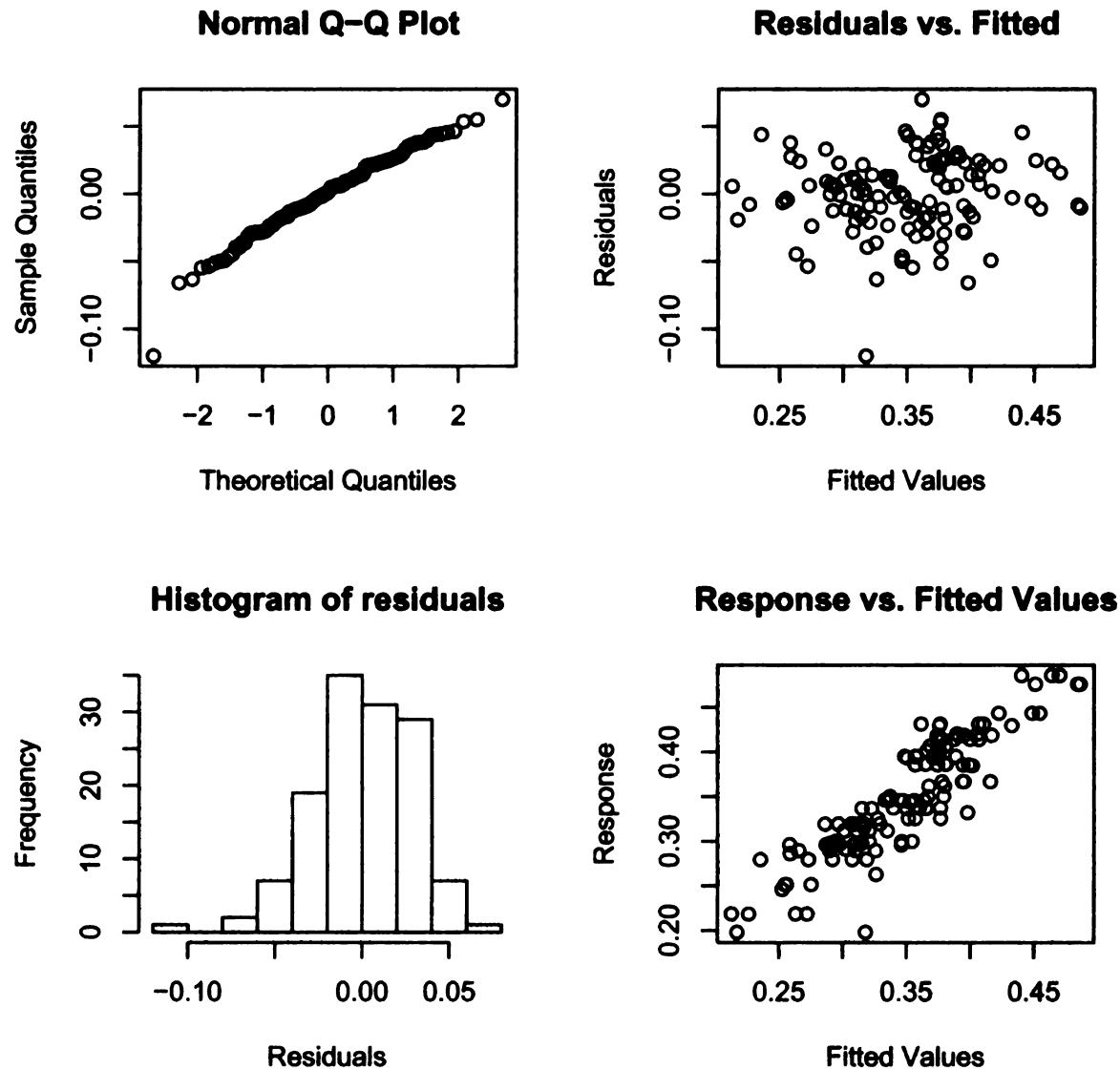
age information to be gained during the latter stages of development, as the animal has ceased changing in size. This leaves large errors in age estimates using conventional aging techniques (including stage/length/weight/temp/strain), with PDE values plummeting to 36.2% for third instars (Figure 5.51) and 15.8% for pupae (Figure 5.52). When gene expression data are included, these jump to 79.8% (Figure 5.53) and 78.2% (Figure 5.54) respectively.



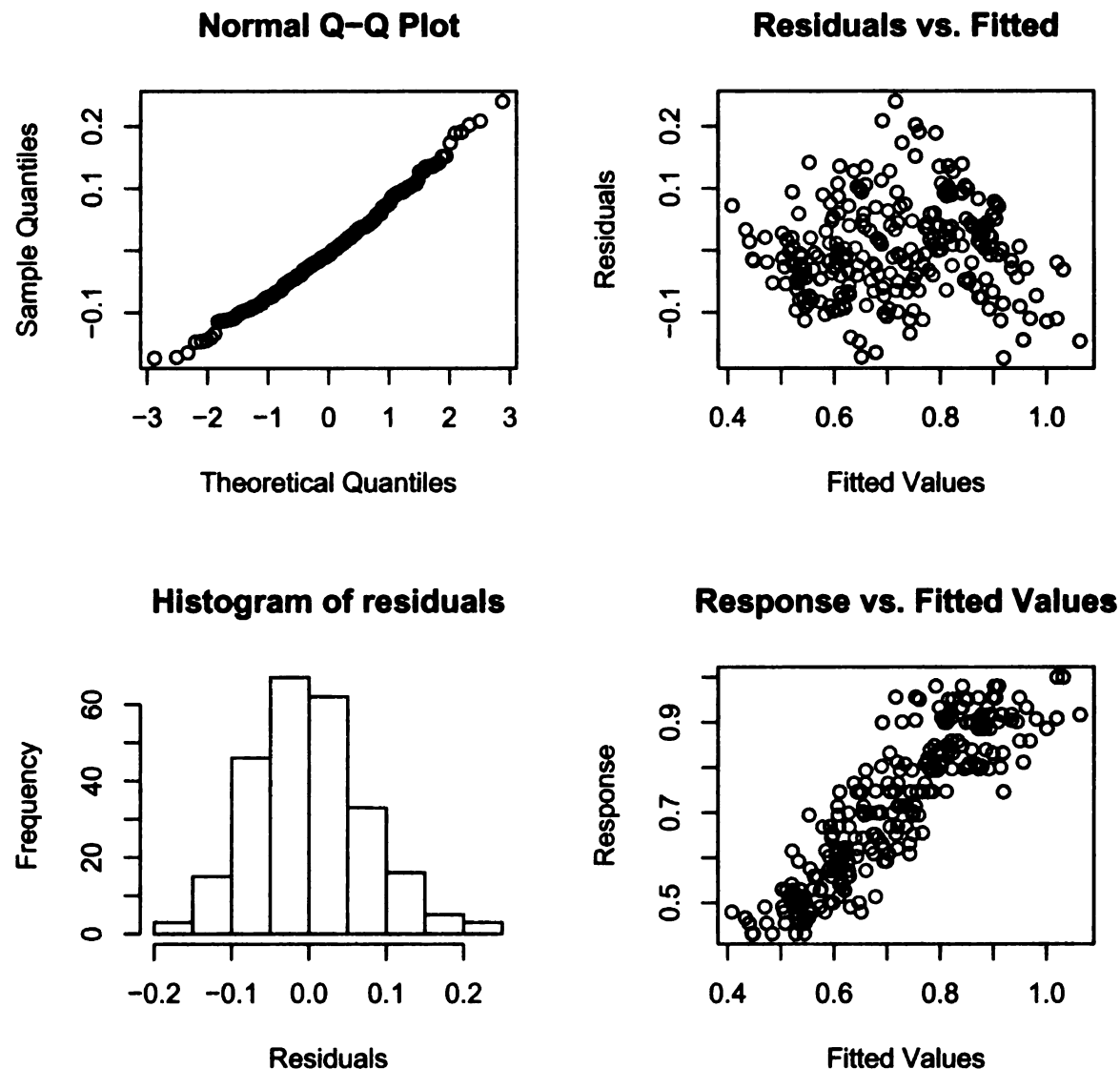
**Figure 5.51. Diagnostic plot for postfeeding third instar incorporating stage/length/weight/ temp/strain.** Descriptions for each panel type can be found in Figure 3.4. Note the extensive scatter in response vs. fitted values.



**Figure 5.52. Diagnostic plot for pupae incorporating stage/length/weight/ temp/strain.** Descriptions for each panel type can be found in Figure 3.4. Note the extensive scatter in response vs. fitted values.



**Figure 5.53. Diagnostic plot for postfeeding third instar incorporating stage/length/weight/ temp/strain and expression data.** Descriptions for each panel type can be found in Figure 3.4. Note the tightening in response vs. fitted values over Figure 5.51.



**Figure 5.54. Diagnostic plot for postfeeding third instar incorporating stage/length/weight/ temp/strain and expression data.** Descriptions for each panel type can be found in Figure 3.4. Note the tightening in response vs. fitted values over Figure 5.52.

### *Discussion*

Two important conclusions, regarding PMI predictions made with insect evidence, can be drawn from the research presented in this section. The first, and most important, is that the hypothesis that gene expression data can improve predictions of blow fly age was mathematically and graphically demonstrated (Tables 5.3, 5.4, and

Figures 5.43–54). The second is that the inclusion of strain and temperature into a GAM used to predict development percents of immature *L. sericata* had little effect on the outcome of predictions. Accordingly, it is reasonable to expect that the technique can be applied to different populations, raised at variable temperatures, with little overall effect on the results of a prediction.

The effective use of genetic data was most impressive during the third instar and pupation. This was anticipated, as it is these stages that are the longest, and are where only broad estimates of fly age can currently be made. The first two instars are typically about a day in length (temperature dependent), thus knowledge of developmental stage alone is an excellent predictor of age, and is where most of the precision in knowing stage originates. Size is helpful during the feeding portion of the third instar, however once the larvae cease growing, precision tumbles. This is not fully reflected in the general GAMs, as they look across development, not at specific parts of development. Once comparisons within these stages were done however, the utility of gene expression data became obvious. Relatively little was learned from the standard traits in postfeeding third instar larvae, with only about 36% of the deviance in the data being explained. This stems from the body size change (shrinkage) that occurs once feeding ceases and the increase in variation that occurs with the shrinkage. During pupation, where size changes little or not at all, the PDE dropped to 16%. Addition of gene expression data made both of these values jump (80% and 78% respectively). It is where standard aging techniques are at their worst that genetic data provided the greatest improvement in predicting blow fly age.

In addition to increasing the precision of blow fly age predictions, the molecular methodology is favorable to the standard approach for other reasons. First, the use of GAMs and their related statistics allows for a detailed understanding and description of error, an important part of the Daubert requirements for scientific evidence (see Chapter 3 for more on this point). Second, almost any forensic laboratory that is qualified to work with DNA can conduct gene expression analyses, assuming they are using quantitative PCR (many laboratories are). Finally, with robotics and microarrays (Arbeitman et al. 2002, Beckstead et al. 2005) the technique can be automated and miniaturized. All of these qualities make the analysis of gene expression data a powerful future tool for forensic entomological PMI predictions.

There are several non-forensic benefits that have resulted from this research as well. The information on expression levels of the insecticide-related genes *cs*, *ace*, *ecr*, *usp*, and *rop-1* may aid in the control of *L. sericata* and other fly species. The first four genes are several of the main targets of insecticides (Ware and Whitacre 2004), as they all involve integral aspects of fly biology, that cannot be lost without causing death: *cs* is critical to the formation of an insect cuticle (Tellam et al. 2000), *ecr* and *usp* form a heterodimer molecule that is necessary for the proper progression of development (Henrich and Brown 1995), *ace* is critical to the function of nerves that interact with the neurotransmitter acetylcholine (Chen et al. 2001), and mutants for *rop-1* (an alpha carboxylesterase) can enable resistance to organophosphates (Newcomb et al. 2005), which are insecticides that target acetylcholine-binding nerves. The high-resolution data of the expression levels of all of these genes, throughout the immature lifecycle, reveals an interesting picture that will be useful in designing insect control programs.



Each insecticide-interacting gene varies in concentration throughout development. Two genes (*cs* and *ace*) are even likely to not be expressed at all during the third instar. Periods of high or low expression of a gene may indicate points in development where different chemicals may be more or less effective in killing these insects. It will be possible, using this and similar research, to identify stages of development that are likely to be more or less resistant to a specific insecticide. In the case of *ace*, the third instar was a period of low expression for this gene, thus organophosphates may be differentially effective as a poison for that stage than for the other immature stages. Such information will enable a more educated decision as to the type of insecticide to use for pest control, based on the life stage that will be targeted. Additionally, 20°C increased *ace* expression, which, again, may alter the effectiveness of organophosphates. If it is found that certain temperatures are more conducive to insect control, operations that use insecticides will know the temperatures to maintain their facilities to ensure maximal effectiveness of insecticide treatments. Temperature effects on insecticide resistance may also be used to guide the seasonal use of these chemicals by applying them when they are most likely to kill pests.

The gene expression profiles also provide a high-resolution picture of transcript level abundance throughout development, which is information that is useful to evolutionary biologists. One current goal of evolutionary and medical biologists is to determine the nature of evolution in gene regulation and phenotypes, through comparative genomics (Collins 1998, Yeates and Wiegmann 2005). Evolutionary biologists aim to gain insight into speciation and adaptation (i.e. Skaer et al. 2002, McGregor 2005, Yeates and Wiegmann 2005), while the medical field will need to

understand how to transfer treatments for diseases in mammalian model organisms to humans or predict when treatments will not transfer across species (Varki and Altheide 2005, Khaitovich et al. 2006). Both fields will benefit from emerging comparisons of gene regulation among related species.

Dipteran biology is uniquely poised to be a leading contributor to the emerging field of comparative genomics. The order contains *D. melanogaster*, which is an important model organism with a sequenced genome. Currently, there are eleven other *Drosophila* genomes that are complete ([www.flybase.org](http://www.flybase.org)), as well as projects to sequence a number of other fly genomes and to sequence 50 genomes in *D. melanogaster*. In addition, the pest status of a number of other Dipterans, including mosquitoes, the Tse-tse fly (Diptera: Muscidae), and the Mediterranean fruit fly *Ceratitis capitata* (Diptera: Tephritidae) have resulted in two complete mosquito genomes (*Anopheles gambiae* and *Aedes aegypti*) and several ongoing genome sequencing projects. Given the propensity of the Diptera for developing into pests, along with the plummeting costs of sequencing, the list of published fly genomes is likely to grow. All of this genomic sequence allows for powerful comparisons of any gene of interest among species. As an example, spliceosomes were recently compared among 11 insect genomes, revealing important similarities and evolutionary differences among their snRNA genes (Mount et al. 2007). Of the genomes compared, 8 were Dipteran (6 were Drosophilids).

Currently, most research in comparative genomics involves sequence comparisons, though the tools exist now to investigate how gene expression affects phenotype evolution as well. Already, such a comparison within the family Calliphoridae

has demonstrated that acrostichal bristle types are determined by heterochrony in the expression of proneural genes during the growth of imaginal discs (Skaer et al. 2002). As technology and genomic sequences become available, more of these sorts of comparisons will occur.

The data reported here address several questions that will be critical for understanding gene expression evolution: How much variation in expression is there? What are the distributions of gene expression levels? How does genotype affect an expression profile? How does the environment affect an expression profile?

The first two questions are important for several reasons. The nature of gene expression variation will be important for the choice of statistical tools used to analyze and compare profiles. Additionally, selection can only occur on variable traits, so information on gene expression variation (if it affects a phenotype) is an indicator of variation that selection can act upon. Unfortunately, information on gene expression variation throughout development, which can be used to answer these questions, is rare. Typically, comparisons of gene expression are limited by cost so temporal profiles tend to have very few replicates at any one time-point. As an example, Arbeitman et al. (2002) performed two replicate analyses at any one time-point. Likewise, studies that include sufficient replication to characterize distributions of variation in gene expression tend to focus on one or a few time points (Powell 1997, Montooth et al. 2003, and Tarone et al. 2005).

This dissertation produced temporal profiles of gene expression, with sufficient replication throughout the immature lifecycle to develop a detailed picture of how transcript abundances of the 9 genes in the research vary throughout maturation. Based

on Figures 5.6–5.41, several trends were apparent. First, variation in the expression levels of all of these genes, at any point in development, was at least one order of magnitude (one CT unit is approximately a two fold change in expression). Second, variation tended to be uniform throughout development for any given locus (with several exceptions). Third, different loci exhibited different amounts of variation in expression levels. Fourth, there was typically a non-normal (gamma) distribution for expression levels, with most loci demonstrating a tail along the low concentration side of the average CT score.

The information on gene expression variation produced in this dissertation has several important repercussions. First, models of gene expression will likely need to account for ample variation in gene expression (ten fold or more) and should assume a gamma distribution before a normal distribution. Second, for any locus the distribution and variation in expression of that gene will be specific to that locus. Also, variation at one time point can reasonably be assumed to be uniform throughout development, unless there is specific data to suggest otherwise. Last, there is sufficient variation in gene expression for selection to act upon. Such knowledge will be important as researchers develop models to explain the effects of gene expression variation within a pathway or network.

Understanding the nature of genotype and environment effects on gene expression will also be important in future comparative genomic endeavors. In this research, there were consistent differences in transcript abundances for several loci that varied by genotype and temperature, resulting in the subtle but significant effects of strain and temperature on GAM predictions of age (Table 5.3). Variation ranged from very little in

the case of *hsp60* (which only varied by strain at low temperatures) to very large in the case of *ace* (which was affected by temperature and strain). One key difference between temperature and strain effects was that changes in expression among strains tended to result in a shift of the same pattern along the standardized CT axis (i.e. *rop-1* in Figure 5.29), while temperature differences could result in vertical shifts in expression (i.e. *cs* in Figure 5.7), completely different patterns in expression (i.e. *hsp90* in Figure 5.15), or a shift in expression along the time axis (i.e. *rop-1* in Figure 5.27).

Knowledge of effects on expression variation will aid future research in several ways. First, the importance of environmental effects was highlighted, as they seemed to exert more influence on some loci than genetic differences among strains. Moreover, the types of changes that the two temperatures produced were varied (shifts along both axes and new expression patterns), while the effects of genetic differences were to maintain the overall pattern of expression but shift the average concentration of transcripts produced by a locus, by less than one order of magnitude. Such knowledge will be important to future comparative genomic research as it will enable investigators to design, analyze, and interpret experiments with these effects in mind.

The results from this chapter have contributed to the bodies of knowledge in several different fields of biology. Basic information on the variation in gene expression, throughout development, (for the 9 loci investigated) will be useful in designing, analyzing, and interpreting comparative genomic data. For the five genes that interact with insecticides, this information will be valuable in designing insect control programs for *L. sericata* and similar blow flies that will maximize the effectiveness of the insecticides used to control the insect. Most importantly, the main hypothesis of this

research, that reliable gene expression throughout development can help produce more precise predictions of blow fly age, was demonstrated.

Quantitative PCR and sequencing primers used in these experiments. Tub= $\beta$  *Tubulin 56 D*. Rp49=*rp49*. ChS=*chitin synthase*. Hsp60=*heat shock protein 60*. Hsp90=*heat shock protein 90*. AcE=*acetylcholine esterase*. EcR=*ecdysone receptor*. Usp=*ultraspiracle*. Rop=*resistance to organophosphate 1*. W=*white*. Sll=*slalom*. Wg=*wingless*. Scl=*scalloped wings*. Rh3=*rhodopsin 3*. Template refers to accession numbers of the sequences used to construct the primers. NA in the template column indicates sequences that were obtained by this research, that are not yet submitted to NCBI. nM refers to the end concentration, in nM, of primers in a PCR reaction.

**Table 5.1. Primers used for sequencing and qPCR for all loci in this experiment.**

<b>Primer</b>	<b>Function</b>	<b>Sequence</b>	<b>Template</b>	<b>nM</b>
Tub R1	Sequencing	CACCAGATCGTTCATGTTGC	NM_166357	1000
Tub F1	Sequencing	CGAGACCTACTGCATCGACA	NM_166357	1000
qTub R	Quantitative PCR	ACCAGGCATGAAAAAGTGAAGAC	EF056211	400
qTub F	Quantitative PCR	TCCGTAAATTGGCCGTCAAC	EF056211	400
qRp49 F	Quantitative PCR	ACAATGTTAAGGAAGTCAAGTTTGG	AB118976	400
qRp49 R	Quantitative PCR	GGAGACACCGTGAGCGATTT	AB118976	400
ChS F2	Sequencing	GAAGTGCCTATACCCGTGGA	AF221067	1000
ChS R2	Sequencing	GGATGTAAACACGCCGCTAT	AF221067	1000
qChS F	Quantitative PCR	GCCGACGGAGAACCTATACCA	EF056212	66.67
qChS R	Quantitative PCR	GATGGCTGTCATTGTGGGTACA	EF056212	400
qHsp60 F	Quantitative PCR	CATCATTCCCGCCCTTGA	AB118971	66.67
qHsp60 R	Quantitative PCR	ATCTTCGGCAATAATGACCAAAG	AB118971	400
qHsp90 F	Quantitative PCR	AAGATCATTTGGCTGTCAAGCA	AB118970	400
qHsp90 R	Quantitative PCR	AGAAGGGCACGGAATTCAAGT	AB118970	400
AcE F4	Sequencing	TATATGGGCTCCAGCAAAGG	U88631	1000
AcE R4	Sequencing	ATGGTACCCGATTGCATCAT	U88631	1000
qAcE F	Quantitative PCR	CACCGGTTATGCCAGGTTTT	NA	66.67
qAcE R	Quantitative PCR	TGATCCCAAAGGCCAACATT	NA	400
EcR F4	Sequencing	TTTCACCCTCGAGCAGTCTT	U75355	1000
EcR R3	Sequencing	CTTCTTTTCGCGTCGTTTC	U75355	1000
qEcR F	Quantitative PCR	GCATGCGGCCGGAAT	NA	400
qEcR R	Quantitative PCR	GCGTCGTTTCATTGCACACT	NA	400
Usp F1	Sequencing	CGCAGGAGATAAAGCCAGAC	AY007213	1000
Usp R1	Sequencing	TGGTGTGCGACGTGCATATT	AY007213	1000
qUsp F	Quantitative PCR	CGAGCAAAAAGCCGAATCAC	NA	400
qUsp R	Quantitative PCR	TGCCTACGCGCAAAAAGG	NA	400
qRop F	Quantitative PCR	GCCCCACTGTTGAGCCATA	AY691501	400
qRop R	Quantitative PCR	CCCGAGGATGTTTGGGTAAAG	AY691501	400
W F1	Sequencing	ACCGATCCTCCGCTCTTAAT	U38899	1000
W R1	Sequencing	TGATATCCAAGAACGCCACA	U38899	1000
qW F	Quantitative PCR	ACAACAGCCAAGACTTGGACATAG	NA	400
qW R	Quantitative PCR	GCGCCCAGTGTCCTACCA	NA	400
qSII F	Quantitative PCR	TCCAACGGCCACAATCTTAAGTA	AY926574	66.67
qSII R	Quantitative PCR	CGTTTAGGTGTTGCCGCAAT	AY926574	400
qWg F	Quantitative PCR	TGTCTGTTTCTGTACGGTGAA	AY926575	400
qWg R	Quantitative PCR	TTATCGCCAATAACACGGAAATT	AY926575	66.67
ScI F4	Sequencing	CGCCATTGTGAACGTGATAC	U58977	1000
ScI R3	Sequencing	GCGAAAGCCAAAACCTACGAG	U58977	1000
qScI F	Quantitative PCR	CGGAAGCGGCAGATTTTT	NA	400
qScI R	Quantitative PCR	TTCTCCGGGATTGGTGACA	NA	400
Rh3 F2	Sequencing	CGGCAAATCCTTATCGAAAT	AJ878411	1000
Rh3 R3	Sequencing	ACAAAACGTCCCCAACTTTC	AJ878411	1000
qRh3 F	Quantitative PCR	ACTACGAATGCTTTTATTGCCTTATG	NA	66.67
qRh3 R	Quantitative PCR	GCTTTGCCAGTGAGTCATTTTACC	NA	400



**Table 5.2. Gene sequencing results.**

<b><u>Gene</u></b>	<b><u>Species</u></b>	<b><u>Size</u></b>	<b><u>Percent</u></b>
<i><math>\beta</math> tubulin 56 D</i>	<i>Glossina morsitans morsitans</i>	635	86
<i>chitin synthase</i>	<i>Lucilia cuprina</i>	713	98
<i>acetylcholine esterase</i>	<i>Lucilia cuprina</i>	369	99
<i>ecdysone receptor</i>	<i>Lucilia cuprina</i>	350	98
	<i>Lucilia sericata</i>	102	100
<i>ultraspiracle</i>	<i>Lucilia cuprina</i>	683	95
	<i>Lucilia sericata</i>	508	99
<i>white</i>	<i>Lucilia cuprina</i>	861	95
<i>scalloped wings</i>	<i>Lucilia cuprina</i>	884	96
<i>rhodopsin 3</i>	<i>Calliphora vicina</i>	311	85

Species indicates the dipteran species that best matched a sequence. Size indicates the length in base pairs of match results for the sequence BLAST on the NCBI website. Percent indicates the percent similarity of that BLAST comparison.

**Table 5.3. Generalized additive models assessed in this experiment.**

<u>Model</u>	<u>Model Parameters</u> [Distribution (link)]	<u>GCV</u>	<u>PDE</u>	<u>N</u>
1	Percent = Stage [gamma(log)]	0.049	88.2	958
2	Percent = bins+Stage [gamma(log)]	0.049	88.5	958
3	Percent = binace [gamma(log)]	0.33	3.6	958
4	Percent = binw [gamma(log)]	0.35	0.43	958
5	Percent = bins//+Stage+Strain+Temp [gamma(log)]	0.048	88.5	958
6	Percent = s(cs)+Stage [gamma(log)]	0.046	89.5	856
7	Percent = s(hsp60) [gamma(log)]	0.41	14.7	852
8	Percent = s(hsp90)+Temp+Stage [gamma(log)]	0.048	88.8	947
9	Percent = s(ace)+Temp+Stage [gamma(log)]	0.046	89.9	636
10	Percent = s(ecr) [gamma(log)]	0.33	6.65	941
11	Percent = s(rop-1)+Stage [gamma(log)]	0.046	88.2	859
12	Percent = s(w) [gamma(log)]	0.31	3.55	725
13	Percent = s(usp)+Stage [gamma(log)]	0.047	88.8	849
14	Percent = s(sif) [gamma(log)]	0.34	9.17	677
15	Percent = Stage +Strain + Temp+ s(Length)+s(Weight)+s(Length,Weight) [gamma(log)]	0.038	91.8	958
16	Percent= Stage+Strain+Temp+s(L)+s(W)+s(L,W)+bin[caws] [gamma(log)]	0.037	92.1	958
17	Percent= Stage+Strain+Temp+s(L)+s(W)+s(L,W)+s(genes) [gamma(log)]	0.024	95.1	501
18	Percent= Stage+Strain+Temp+s(L)+s(W)+s(L,W)+s(genes) + bin[caws][gamma(log)]	0.024	95.1	501
19	Percent= Stage+Strain+Temp+s(L)+s(W)+s(L,W)+s(genes) + bin[caws] [gaussian(identity)]	0.0079	91.3	501
20	Percent= Stage+Strain+Temp+s(L)+s(W)+s(L,W)+s(hsp60,genes) + bin[caws] [gaussian(identity)]	0.007	93.7	501
21	Percent= Stage+Strain+Temp+s(L)+s(W)+s(L,W)+s(Length,genes) + bin[caws] [gaussian(identity)]	0.0056	94.6	501
22	Percent= Stage+Temp+s(L,W)+s(Length,genes) + bin[caws] [gaussian(identity)] no cs or w	0.0059	94.6	534
23	Percent= Stage+s(L,W)+s(Length,genes) + bin[caws] [gaussian(identity)] no cs or w	0.0059	94.7	534

The various models used to predict minimum development percent (Percent), the distributions applied to each model (gamma or gaussian), and the link function that applied the distribution to the model (identity or log). bin*locus* indicates the binary data for that gene. bin[caws] means all four significant binary variables (cs, ace, w, and sif) were included. s(*locus*) indicates a smoothing curve for that gene was analyzed. s(L) or s(W) indicates smoothing curves for length or weight were analyzed. s(variable,variable) indicates that a smoothed contour surface composed of the two variables was used to predict development percent. GCV is the generalized cross validation score for that model. PDE is the percent deviance explained by a model. N is the number of individuals included in that model.

**Table 5.4. Estimated significance and degrees of freedom for variables in all GAMs.**

<b>Model</b>	<b>Variable</b>	<b>Linear</b>	<b>chi-sq</b>	<b>df/edf</b>	<b>P-value</b>
1	Stage	Y	8816.2	4	<0.0001
2	Stage	Y	9031.3	4	<0.0001
	binary <i>chitin synthase</i>	Y	21.087	1	<0.0001
3	binary <i>acetylcholine esterase</i>	Y	46.911	1	<0.0001
4	binary <i>white</i>	Y	5.209	1	0.023
5	Stage	Y	9025.6	4	<0.0001
	Strain	Y	4.2436	2	0.12
	Temp	Y	22.828	1	<0.0001
	binary <i>slalom</i>	Y	8.0417	1	0.0047
6	Stage	Y	8797	4	<0.0001
	<i>s(chitin synthase)</i>	N	23.093	3.99	0.0001
7	<i>s(heat shock protein 60)</i>	N	156.36	6.96	<0.0001
8	Stage	Y	7926.3	4	<0.0001
	Temp	Y	28.806	1	<0.0001
	<i>s(heat shock protein 90)</i>	N	25.187	6.52	0.0005
9	Stage	Y	6502.8	4	<0.0001
	Temp	Y	12.568	1	0.0004
	<i>s(acetylcholine esterase)</i>	N	30.128	7.17	0.0001
10	<i>s(ecdysone receptor)</i>	N	82.179	5.46	<0.0001
11	Stage	Y	7309.2	4	<0.0001
	<i>s(resistance to organophosphate 1)</i>	N	62.44	5.47	<0.0001
12	<i>s(white)</i>	N	28.493	6.15	0.0001
13	Stage	Y	8167.2	4	<0.0001
	<i>s(ultraspiracle)</i>	N	27.006	3.58	<0.0001
14	<i>s(slalom)</i>	N	83.34	5.85	<0.0001
15	Stage	Y	483.29	4	<0.0001
	Strain	Y	23.608	2	<0.0001
	Temp	Y	94.234	1	<0.0001
	<i>s(Length)</i>	N	39.327	7.85	0.11
	<i>s(Weight)</i>	N	13.253	8.02	<0.0001
	<i>s(Length,Weight)</i>	N	68.628	15.2	<0.0001
16	Stage	Y	460.93	4	<0.0001
	Strain	Y	16.914	2	<0.0001
	Temp	Y	115.13	1	<0.0001
	<i>s(Length)</i>	N	21.342	7.13	0.0039
	<i>s(Weight)</i>	N	15.804	4.91	0.0073
	<i>s(Length,Weight)</i>	N	80.472	21.9	<0.0001
	binary <i>chitin synthase</i>	Y	32.382	1	<0.0001
	binary <i>acetylcholine esterase</i>	Y	0.1385	1	0.71
	binary <i>white</i>	Y	0.7359	1	0.39
	binary <i>slalom</i>	Y	0.0736	1	0.79

**Table 5.4. Continued.**

<b>Model</b>	<b>Variable</b>	<b>Linear</b>	<b>chi-sq</b>	<b>df/edf</b>	<b>P-value</b>
<b>17</b>	Stage	Y	232.56	4	<0.0001
	Strain	Y	3.9962	2	0.14
	Temp	Y	62.513	1	<0.0001
	s(Length)	N	30.504	5.61	<0.0001
	s(Weight)	N	8.6437	1.27	0.0054
	s(Length,Weight)	N	37.219	11.8	0.0003
	s( <i>chitin synthase</i> )	N	8.7356	2.98	0.034
	s( <i>heat shock protein 60</i> )	N	52.67	5.75	<0.0001
	s( <i>heat shock protein 90</i> )	N	16.536	1.92	0.0003
	s( <i>acetylcholine esterase</i> )	N	80.381	7	<0.0001
	s( <i>ecdysone receptor</i> )	N	21.546	1	<0.0001
	s( <i>resistance to organophosphate 1</i> )	N	44.075	5.07	<0.0001
	s( <i>white</i> )	N	2.9499	1.63	0.17
	s( <i>ultraspiracle</i> )	N	23.967	5.25	<0.0001
	s( <i>slalom</i> )	N	1.5654	1	0.21
<b>18</b>	Stage	Y	232.56	4	<0.0001
	Strain	Y	3.9962	2	0.17
	Temp	Y	62.513	1	<0.0001
	s(Length)	N	30.504	5.61	<0.0001
	s(Weight)	N	8.6437	1.27	0.0054
	s(Length,Weight)	N	37.219	11.8	0.0003
	s( <i>chitin synthase</i> )	N	8.7356	2.98	0.034
	s( <i>heat shock protein 60</i> )	N	52.67	5.75	<0.0001
	s( <i>heat shock protein 90</i> )	N	16.536	1.92	0.0003
	s( <i>acetylcholine esterase</i> )	N	80.381	7	<0.0001
	s( <i>ecdysone receptor</i> )	N	21.546	1	<0.0001
	s( <i>resistance to organophosphate 1</i> )	N	44.075	5.07	<0.0001
	s( <i>white</i> )	N	2.9499	1.63	0.17
	s( <i>ultraspiracle</i> )	N	23.967	5.25	0.0004
	s( <i>slalom</i> )	N	1.5654	1	0.21
	binary <i>chitin synthase</i>	Y	101.58	1	<0.0001
	binary <i>acetylcholine esterase</i>	Y	101.58	1	<0.0001
	binary <i>white</i>	Y	101.58	1	<0.0001
	binary <i>slalom</i>	Y	101.58	1	<0.0001

**Table 5.4. Continued.**

<b>Model</b>	<b>Variable</b>	<b>Linear</b>	<b>chi-sq</b>	<b>df/edf</b>	<b>P-value</b>
<b>19</b>	Stage	Y	107.61	4	<0.0001
	Strain	Y	5.7489	2	0.057
	Temp	Y	22.098	1	<0.0001
	s(Length)	N	0.3813	1	0.54
	s(Weight)	N	4.9028	1	0.027
	s(Length,Weight)	N	51.222	14.5	<0.0001
	s( <i>chitin synthase</i> )	N	4.8273	1	0.029
	s( <i>heat shock protein 60</i> )	N	73.642	4.34	<0.0001
	s( <i>heat shock protein 90</i> )	N	24.305	3.64	<0.0001
	s( <i>acetylcholine esterase</i> )	N	89.688	6.41	<0.0001
	s( <i>ecdysone receptor</i> )	N	18.27	1	<0.0001
	s( <i>resistance to organophosphate 1</i> )	N	51.348	6.19	<0.0001
	s( <i>white</i> )	N	0.6523	1	0.42
	s( <i>ultraspiracle</i> )	N	35.758	3.86	<0.0001
	s( <i>slalom</i> )	N	2.1667	2.15	0.37
	binary <i>chitin synthase</i>	Y	23.006	1	<0.0001
	binary <i>acetylcholine esterase</i>	Y	23.006	1	<0.0001
	binary <i>white</i>	Y	23.006	1	<0.0001
	binary <i>slalom</i>	Y	23.006	1	<0.0001
<b>20</b>	Stage	Y	120.09	4	<0.0001
	Strain	Y	11.099	2	0.0042
	Temp	Y	10.942	1	0.001
	s(Length)	N	6.7474	3.38	0.11
	s(Weight)	N	4.8594	1.2	0.038
	s(Length,Weight)	N	24.463	9.38	0.0054
	s( <i>heat shock protein 60, chitin synthase</i> )	N	41.961	17.8	0.0015
	s( <i>heat shock protein 60, heat shock protein 90</i> )	N	39.926	7.75	<0.0001
	s( <i>heat shock protein 60, acetylcholine esterase</i> )	N	63.749	7.54	<0.0001
	s( <i>heat shock protein 60, ecdysone receptor</i> )	N	56.883	15	<0.0001
	s( <i>heat shock protein 60, resistance to organophosphate 1</i> )	N	58.023	15.4	<0.0001
	s( <i>heat shock protein 60, white</i> )	N	0.5825	1	0.46
	s( <i>heat shock protein 60, ultraspiracle</i> )	N	50.482	9.58	<0.0001
	s( <i>heat shock protein 60, slalom</i> )	N	0.2966	1.09	0.63
	binary <i>chitin synthase</i>	Y	18.171	1	<0.0001
	binary <i>acetylcholine esterase</i>	Y	18.171	1	<0.0001
	binary <i>white</i>	Y	18.171	1	<0.0001
	binary <i>slalom</i>	Y	18.171	1	<0.0001

**Table 5.4. Continued.**

<b>Model</b>	<b>Variable</b>	<b>Linear</b>	<b>chi-sq</b>	<b>df/edf</b>	<b>P-value</b>
<b>21</b>	Stage	Y	167.71	4	<0.0001
	Strain	Y	20.991	2	<0.0001
	Temp	Y	16.04	1	<0.0001
	s(Length)	N	0.7976	1	0.37
	s(Weight)	N	11.685	2.99	0.0091
	s(Length,Weight)	N	0.5645	0.37	0.21
	s(Length, <i>chitin synthase</i> )	N	2.7732	1	0.097
	s(Length, <i>heat shock protein 60</i> )	N	130.35	9.69	<0.0001
	s(Length, <i>heat shock protein 90</i> )	N	64.9	10.1	<0.0001
	s(Length, <i>acetylcholine esterase</i> )	N	103.54	14.2	<0.0001
	s(Length, <i>ecdysone receptor</i> )	N	38.517	1	<0.0001
	s(Length, <i>resistance to organophosphate 1</i> )	N	56.472	12	<0.0001
	s(Length, <i>white</i> )	N	1.1918	1	0.28
	s(Length, <i>ultraspiracle</i> )	N	107.65	14.9	<0.0001
	s(Length, <i>slalom</i> )	N	30.064	7.77	0.0002
	binary <i>chitin synthase</i>	Y	17.721	1	<0.0001
	binary <i>acetylcholine esterase</i>	Y	17.721	1	<0.0001
	binary <i>white</i>	Y	17.721	1	<0.0001
	binary <i>slalom</i>	Y	17.721	1	<0.0001
<b>22</b>	Stage	Y	207.64	4	<0.0001
	Temp	Y	12.111	1	0.0006
	s(Length,Weight)	N	15.495	5.48	0.013
	s(Length, <i>heat shock protein 60</i> )	N	143.36	11	<0.0001
	s(Length, <i>heat shock protein 90</i> )	N	45.018	8.54	<0.0001
	s(Length, <i>acetylcholine esterase</i> )	N	96.334	13.1	<0.0001
	s(Length, <i>ecdysone receptor</i> )	N	67.053	13.6	<0.0001
	s(Length, <i>resistance to organophosphate 1</i> )	N	81.328	10.5	<0.0001
	s(Length, <i>ultraspiracle</i> )	N	111.35	15.1	<0.0001
	s(Length, <i>slalom</i> )	N	20.687	6.9	0.0046
	binary <i>chitin synthase</i>	Y	5.1045	1	0.024
	binary <i>acetylcholine esterase</i>	Y	17	1	<0.0001
	binary <i>white</i>	Y	0.0717	1	0.79
	binary <i>slalom</i>	Y	17	1	<0.0001

**Table 5.4. Continued.**

<b>Model</b>	<b>Variable</b>	<b>Linear</b>	<b>chi-sq</b>	<b>df/edf</b>	<b>P-value</b>
<b>23</b>	Stage	Y	193.9	4	<0.0001
	s(Length,Weight)	N	18.197	5.39	0.0042
	s(Length, <i>heat shock protein 60</i> )	N	140.01	13.3	<0.0001
	s(Length, <i>heat shock protein 90</i> )	N	38.366	7.27	<0.0001
	s(Length, <i>acetylcholine esterase</i> )	N	104.41	15.4	<0.0001
	s(Length, <i>ecdysone receptor</i> )	N	74.27	14.8	<0.0001
	s(Length, <i>resistance to organophosphate 1</i> )	N	70.912	10.3	<0.0001
	s(Length, <i>ultraspiracle</i> )	N	142.34	16.4	<0.0001
	s(Length, <i>slalom</i> )	N	18.154	7.02	<0.0001
	binary <i>chitin synthase</i>	Y	4.0422	1	0.045
	binary <i>acetylcholine esterase</i>	Y	10.675	1	0.0012
	binary <i>white</i>	Y	0.0231	1	0.88
	binary <i>slalom</i>	Y	10.675	1	0.0012

The significance and degrees of freedom for all variables in all GAMs assessed in this project. For genetic data, the simplest model is shown in which a gene was statistically significant. Model 21 contains all available data; after that model, non-significant variables were removed. All terms were significant predictors of age in at least one model. Linear indicates whether (Y) or not (N) a term is linear. Df=degrees of freedom. edf=estimated degrees of freedom. P-value=estimated P-value. See the Generalized Additive Models section for greater details on models.

## **CHAPTER 6: VALIDATION OF GENE EXPRESSION BASED PREDICTIONS OF BLOW FLY AGE WITH BLIND PREDICTIONS**

### *Introduction*

The results presented in Chapter 5 indicated that predictions of blow fly development percent made with gene expression data are more precise than current forensic entomology approaches. However, the predicted performance of a model does not necessarily ensure that predictions will follow the expected pattern. To prove that a statistical model is a valuable predictor of development percent it is necessary to validate it (Scheiner and Gurevitch 2001), in this case by estimating the ages of unknown immature blow flies in a blind study. If the predicted ages of individuals plot along a True = Predicted line, then a model can be considered validated for use. In this chapter, the predicted ages of *L. sericata* collected on rats were used to validate the use of gene expression for PMI estimates.

### *Methods*

#### **Strain Collection and Identification.**

To avoid the effects of inbreeding resulting from over-winter rearing, a new collection of *L. sericata* was caught out of doors on the Michigan State University campus in May of 2006. Individuals were identified visually and by *COI* sequence, as described in Chapter 3.

#### **Collection of Unknowns.**

Egg masses were placed in the mouths of three CO<sub>2</sub> asphyxiated rat carcasses as described in Developmental Plasticity. Two cohorts were raised in incubators, one at



20°C and one at 33.5°C. Rearing conditions were the same as in Chapter 2. The third cohort was raised in a sealed terrarium under outdoor ambient conditions. The terrarium was covered with a tightly fitting foam lid, which had a hole cut in the middle to allow air in. The hole was sealed by a screen, which kept other flies out of the terrarium.

Several individual larvae or pupae were collected daily from each of the three cohorts. An independent researcher recorded the ages of the individuals, which were stored in RNAlater (Applied Biosystems) at –80°C. Pupae were collected as described in Chapter 3 for the stable temperature treatments; in the terrarium no attempt was made to separate pupae by age. Cohorts were checked daily until adults were observed, which was recorded as the minimum development time for the cohort; all development percents were calculated based on minimum development time.

For the ambient temperature experiment, temperature variance was noted using a HOBO data logger (Onset Computer, Bourne, MA). Temperature data were used to calculate accumulated degree hours (ADH) by subtracting a base temperature of 10°C from the recorded temperature at any hour and adding the hourly values together. This resulted in an accumulated degree hour (ADH) calculation. The total ADH until eclosion was used for the minimum development time and the ADH at the time of collection was divided by that number to calculate minimum development percent for the ambient temperature cohort.

### **Converting RNAlater Size Measurements to Live Size Measurements.**

Since the body size measurements in the previous section were based on live *L. sericata*, and the unknown individuals were presented for analysis stored in RNAlater, it was necessary to develop a protocol for converting size measurements of individuals

stored in solution to an estimate of live size. Unanalyzed individuals (N=400) from the MI collections in Chapter 3 were remeasured and reweighed after approximately one year of storage in RNAlater at  $-80^{\circ}\text{C}$ . The new measurements and weights were plotted against the recorded live data. Regression equations for the comparisons were used to convert the sizes of the unknown individuals to usable data for GAMs.

### **Molecular Biology and Statistics.**

Gene expression CT scores were obtained as described in the previous section, which were used to predict the ages of blind study individuals. Several predictions were made for each individual, beginning with a GAM that included length, weight, and developmental stage to predict development percent. Likewise, GAMs were produced utilizing body size measurements, developmental stage, and all available gene expression information. Either the gamma or gaussian distributions were applied to the models, depending on which better satisfied the assumptions of error (discussed in Chapter 3). Models were also developed with gene CT scores, in which each gene expression term was anchored [i.e. the term  $s(\text{variable}, \text{gene})$  was used] against length measurements [ $s(\text{length}, \text{gene})$ ], as well as *hsp60* CT scores [ $s(\text{hsp60}, \text{gene})$ ] when *hsp60* was detected (see Table 5.3).

Two to four predictions were generated for each individual. Development percents were predicted for all individuals based on body size and developmental stage data. An additive model was also developed for each individual, using all available data. If an individual was raised at a known temperature, then temperature was included in the model. Strain was not included, as few investigators will have that information. If a diagnostic plot indicated that anchoring a prediction with length was not likely to

improve upon other predictions made with body size and stage (based on the description of model types below), then a length anchored prediction was not used.

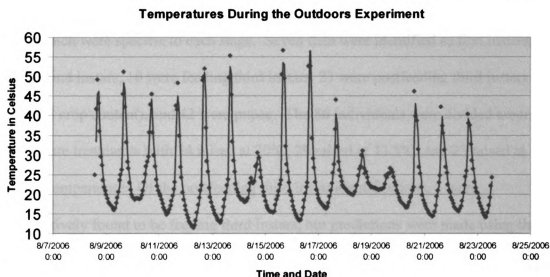
Predicted values of development percent were plotted against true development percent, and were visualized with the  $Y = X$  (True = Predicted) line  $\pm 10\%$ . A GAM was classified as a Type 1 model if its diagnostic plot was similar to Model 15 from the previous section, or as a Type 2 model if the diagnostic plot indicated no gap in predictions and a decrease in the error throughout development. Predictions from Type 1 and Type 2 models were plotted and compared to predictions made with body size and stage information only.

## *Results*

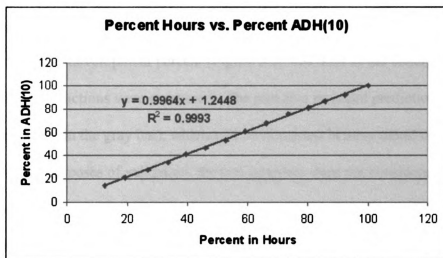
The new strain of Michigan *L. sericata* was identified by visual identification and through *COI* sequence. A BLAST search on the NCBI website revealed that a 753 base pair fragment of the *COI* gene was most similar to another *L. sericata* sequence, with 100% similarity. The closest match to non-*L. sericata* sequence was a 99% match to that gene in *L. cuprina*, with a 7 base pair mismatch.

Ambient temperatures spiked each day, then decreased to an evening low (Figure 6.1). Despite the fluctuations, the ADH based development percents were very similar to the purely temporal development percents (Figure 6.2). Because ADH is generally used in forensic entomology they were used for the analyses detailed below.





**Figure 6.1. Temperature plots for the ambient temperature experiment.** Temperatures spiked in the afternoon as sunlight was directly on the terrarium.



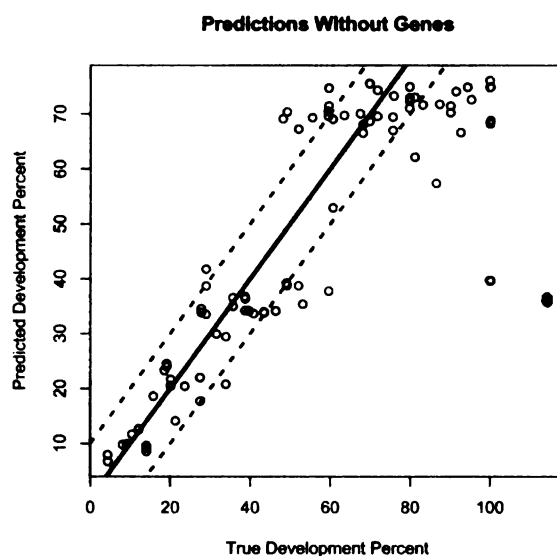
**Figure 6.2. The regression of ADH (base 10) percents versus percent development in hours.** The values were strongly correlated, with an R-squared and slope almost 1. However, the 1.2448 shift up the Y axis indicates that ADH is a better measure of developmental progress.

Ninety individuals were used to obtain cDNA samples for the blind study. Four of these were non-pupators (see the next section) and 86 were normally developing larvae

and pupae. Body size measurements were calculated based on the equations in Tables 6.1–2, which were specific to each stage. Seven flies were identified as first instars, 4 were second instars, 10 were feeding third instars, 23 were postfeeding third instars (based on crop content), and 42 were pupae. The 86 individuals were divided among temperature treatments with 34 raised at 20°C, 29 raised at 33.5°C, and 23 raised at ambient temperature. Of the postfeeding third instar identifications, 5 were retrospectively found to be feeding third instars, but predictions were made using the identified stage, as an investigator would not have been able to recognize a misidentified stage. Forty-nine individuals generated data that could be used in Type 2 unanchored models. Fifty-six individuals received *hsp60* anchored predictions. The ages of 71 flies were predicted based on models that used gene expression data anchored by length.

The plot for predicted versus true age for a GAM using developmental stage and body size to predict development percent (Figure 6.3) was used as the benchmark against which all other predictions were compared. The plot also included predictions for the four non-pupators (in the gray bar), which will be discussed in more detail in the next section. For the purposes of validation, the non-pupators were not considered further. If the hypothesized increase in precision occurred as predicted, then the results of plots from the models that used gene expression to predict blow fly age should demonstrate an improvement over the results in Figure 6.3. Additionally, the error associated with these models was considered. Based on the diagnostic plots of gene expression GAMs (from the previous section), approximately 90% of the data should be within 10% of the true age. With realistic collection conditions, individuals that are younger than the minimum age are expected (only a subset of flies on a body can be the oldest), thus

underpredictions of age are anticipated. However, the minimum development time of a blow fly is the important information used by forensic entomologists, thus very little overprediction of blow fly age should be tolerated in a model that will be used to predict this datum. When body size and stage data were used to predict blow fly age (Figure 6.3), the predictions of larval age was within 10% of true development except for 6 individuals. However, many pupal ages were predicted as greater than 10% older or 10% younger than their true ages. Likewise, few individuals were predicted to be between 40 and 60 percent developed, and no individuals were predicted to be greater than 75% developed.

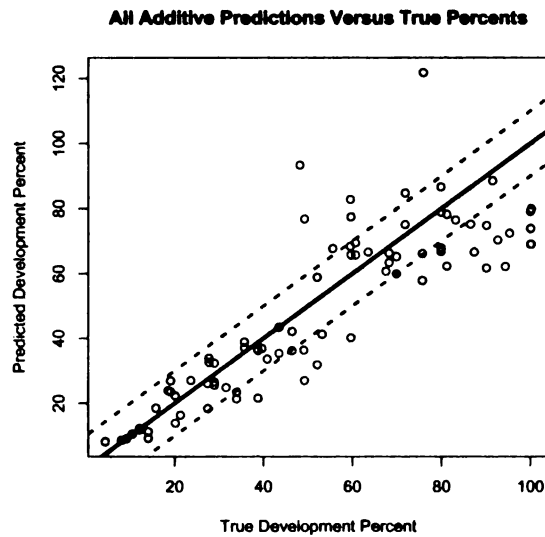


**Figure 6.3. Predicted versus true development percents for all 90 individuals in the blind study.** Predictions were made with GAMs that did not use gene expression information. The four points in the gray bar represent the four larvae that failed to pupate, and were eliminated from analyses in this section.

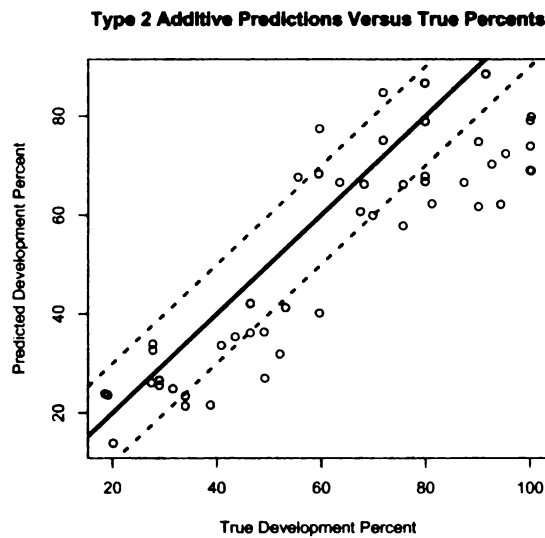
All gene expression models (Figures 6.4–10) improved upon the GAM predictions made with body size/stage data (Figure 6.3). When unanchored additive gene

expression predictions were compared to their true ages (Figure 6.4), 100% of the larvae were predicted as within 10% of their true ages or younger than their true age. Similarly, 85.7% of pupae were predicted as less than 10% older than their true development percents. Most importantly, pupal age predictions were more precise than the size/stage predictions of age, as seen by the reduced size of the gap between larval and pupal age predictions. Additionally, development percents were accurately predicted when true percents were greater than 80%. When only the Type 2 additive model predictions (which included models predicting development percent with complete, or almost complete, gene expression profiles) were plotted against their true development percents (Figure 6.5) 93.9% of individuals were predicted as within or less than 10% of their true ages. In the three cases that larvae were predicted as older than 10% of true age, the predictions were only marginally greater than 10% deviant from the true development percent, which was not the case for predictions of individuals from the preceding models (Figures 6.3–4). Length anchored predictions followed a similar pattern, overall and for Type 2 models (Figure 6.6–7). Over-predicted development percents (>10%) with Type 2 models of this type only occurred twice (again, with less deviation from the true age than the size/stage based model). However, models that used *hsp60*-anchored predictions produced too many imprecise predictions to be considered useful and were not considered further.

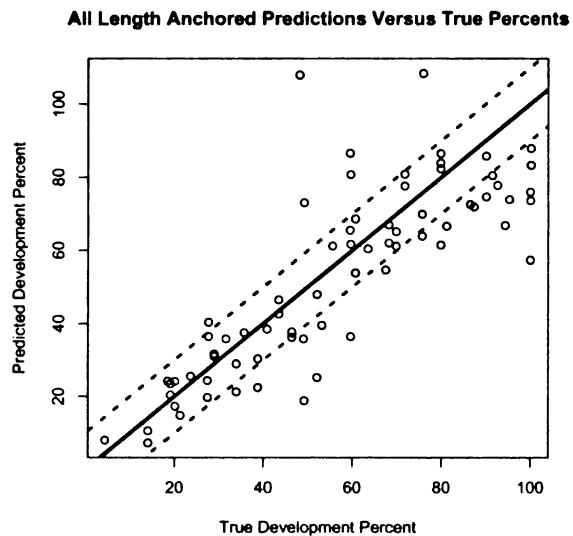




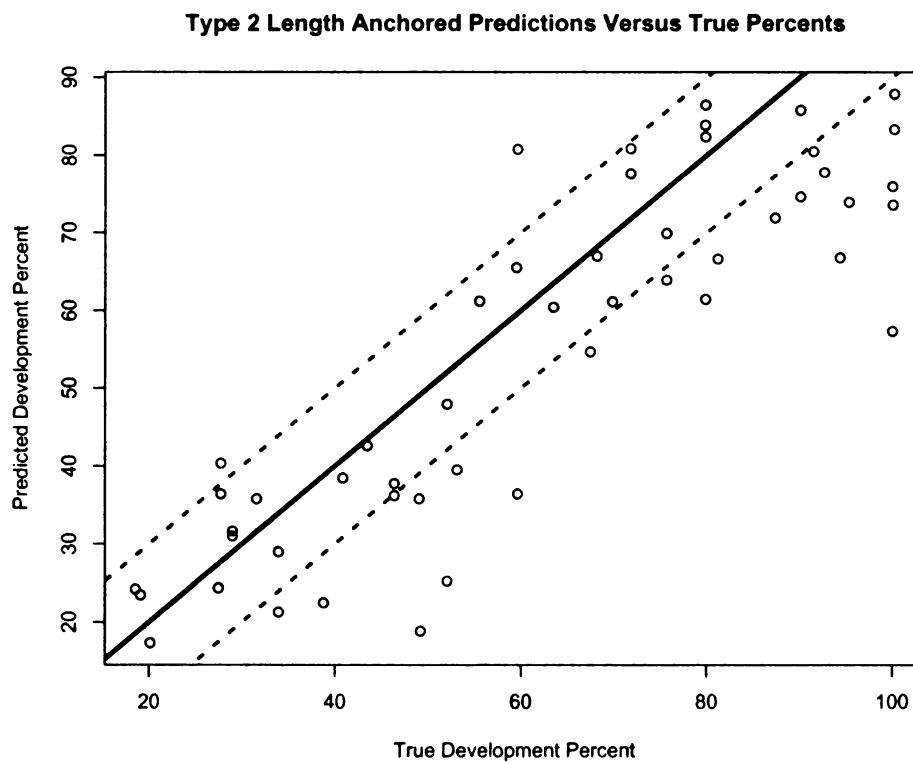
**Figure 6.4. Predicted versus true development percents for the 86 normally developing individuals in the blind study.** Predictions were made with GAMs that used unanchored gene expression terms.



**Figure 6.5. Predicted versus true development percents for 49 normally developing individuals in the blind study.** Predictions were made with GAMs that used unanchored gene expression terms and were Type 2 models.



**Figure 6.6. Predicted versus true development percents for 71 normally developing individuals in the blind study.** Predictions were made with GAMs that used length anchored gene expression terms.



**Figure 6.7. Predicted versus true development percents.** Predictions were made with GAMs that used length anchored gene expression terms and were Type 2 models.

The performances of Type 2 additive models were also evaluated by temperature, but it was not possible to determine if certain temperatures were associated with more accurate predictions than others; since 54.4% of profiles were predicted with unanchored additive Type 2 models and each temperature was only a third of the remaining 49 individuals there was little data to compare. The 20°C predictions seemed the most accurate, but more research (with larger sample sizes) will be required to establish whether or not temperature (as well as other environmental factors) affect predictions made with such data.

All models continued to demonstrate a bias toward predicting individuals to be younger than their true ages. Adding gene expression information did not eliminate underestimates of age; however, the frequency of underestimates was less than those made with the models in Figure 6.1.

### *Discussion*

The results of the blind study demonstrated the utility of using gene expression for predicting blow fly age, and the use of GAMs to predict blow fly age by incorporating gene expression data was validated. When complete or nearly complete gene expression profiles (Type 2 models) were used to predict blow fly age, the predictions were much more precise. Both gene based GAMs and GAMs that used length anchored gene expression terms provided more precise predictions of age, generating far more blow fly age estimates within 10% of actual age. In doing so, a number of problems with the current PMI prediction process have been addressed. Currently, the range of

development percents for pupal samples is ranges from ~40–100% (see Chapter 3) if standard entomological measurements are employed. The use of gene expression clearly decreases this range to within ~10% of the true development percent. Likewise, the predictions produced with gene expression GAMs yielded a more seamless transition of age estimates from postfeeding third instar to pupa. As well, estimates of ages greater than 10% above true age, which are the least desirable, remarkably decreased with these models as compared to the pupal predictions without gene expression data.

Error in predicting individuals to be younger than they were was more common. As this phenomenon occurred independent of the use of gene expression data it is not a signature shortcoming of using RNA profiles to predict age. Indeed, the aging bias occurred without the use of gene expression data. As previously mentioned in Chapter 2, *L. sericata* development time varies, with only some individuals developing at the maximum rate. Thus some individuals could truly be less mature than the most developed of their cohort. In natural conditions, underestimates of age will be likely; as females do not stop laying after the first female has oviposited on a corpse. The most important data, however, from a blow fly age estimate is the predicted minimum development time/PMI. Given that there is a maximum development rate, as long as a model is not likely to predict individual ages that are older than their true age, then the minimum development time will be precisely estimated, as was the case in the blind study. This problem can be largely overcome by sampling multiple individuals. The most developed can be used as an indicator of the minimum amount of time that flies could have colonized remains, thus avoiding or ameliorating a problem with underpredictions of age. Given the ability of GAMs to accurately predict blow fly

development percents using gene expression/body size/stage data, and with more precision than GAMs that employed traditionally used forensic entomology measures (size/stage), the application of these models in conjunction with transcript level information is an attractive option for predicting development percents of immature *L. sericata* for PMI estimates.

**Table 6.1. Regressions of live length against preserved length for each stage.**

Stage	Regression Equation	P-value	R <sup>2</sup>
1 <sup>st</sup> Instar	Length=1.062*RNA Later Length+0.3	<0.0001	0.76
2 <sup>nd</sup> Instar	Length=1.548*RNA Later Length-0.48	<0.0001	0.77
3 <sup>rd</sup> Instar (feeding)	Length=1.095*RNA Later Length+2.2	<0.0001	0.89
3 <sup>rd</sup> Instar (postfeeding)	Length=0.7371*RNA Later Length+5.2	<0.0001	0.59
Pupa	Length=0.9059*RNA Later Length+0.65	<0.0001	0.88

**Table 6.2. Regressions of live weight against preserved weight for each stage.**

Stage	Regression Equation	P-value	R <sup>2</sup>
1 <sup>st</sup> Instar	Weight=1.078*RNA Later Weight+0.07	<0.0001	0.87
2 <sup>nd</sup> Instar	Weight=1.077*RNA Later Weight-0.04	<0.0001	0.98
3 <sup>rd</sup> Instar (feeding)	Weight=1.009*RNA Later Weight+0.12	<0.0001	0.99
3 <sup>rd</sup> Instar (postfeeding)	Weight=1.047*RNA Later Weight-0.94	<0.0001	0.97
Pupa	Weight=0.9541*RNA Later Weight+0.23	<0.0001	0.98

## CHAPTER 7: GENE EXPRESSION ANALYSES OF NON-PUPATING LARVAE

### *Introduction*

Forensic entomology uses the predictable development times of blow flies (Diptera: Calliphoridae) to estimate a postmortem interval (PMI) by backtracking from the developmental stage of the most mature insects collected on a body to the time that the insects colonized the remains (Catts and Haskell 1990, Greenberg and Kunich 2002), a time that can approximate the PMI. Blow fly development is known to be plastic as different food sources and rearing treatments can delay development (Kaneshraja and Turner 2004, Clark et al. 2006, Tarone and Foran 2006) and any factors that contribute to plasticity can result in inaccurate PMI estimates.

In previous research on *L. sericata* (Diptera: Calliphoridae) (Meigen) development (Chapters 3 and 6), a curious form of developmental plasticity was observed in which postfeeding third instar larvae, as determined by behavior, the visibility of tissue in the crop, and gene expression levels (both *cs* and *ecr*), were present at a time when other members of their cohort had reached adulthood. The pupal stage between the postfeeding third instar and the imago is the longest in *L. sericata* immature development and can last for more than a week (Grassberger and Reiter 2001). Clearly, the collection and analysis of these abnormally old larvae, dubbed herein as “Peter Pan” larvae, could have a drastic effect on the accuracy of a PMI estimate if pupae were not collected at the death scene. A forensic entomologist is unlikely to miss pupal evidence, however; one may not be present at the time of collection, which may result in an untrained individual overlooking pupal samples. If pupae are overlooked, “Peter Pan” larvae may be included

in a larval sample from a death scene. Any means of identifying “Peter Pan” larvae will be useful in avoiding their analysis for PMI predictions and may also be an indication that pupae were missed in sampling. During the previous observation of the “Peter Pan” phenomenon (Chapter 3), 10 (or as many as were available) such individuals were collected, and placed in an RNA preserving fixative. Fifty five were profiled for the mRNA expression levels of 9 genes (Chapter 5), revealing useful information for identifying these larvae. The profiles of informative genes were then compared to the profiles of “Peter Pan” larvae collected in a blind study, to confirm that they can be used to successfully predict the condition.

### *Methods*

Ten “Peter Pan” larvae were collected from each replicate (Chapter 3), except in the case of samples from the CA strain, which is discussed below. The mortality of larvae in each cohort and the production of “Peter Pan” individuals were generally observed for each strain, though no formal measures were made. Samples were fixed in 250  $\mu$ L RNAlater (Applied Biosystems) and stored at  $-80^{\circ}\text{C}$ . RNA was isolated and converted to cDNA using previously described methods (Chapter 5). Quantitative PCR analyses of these larvae were then conducted, using primers specific for 9 genes as previously reported (Chapter 5). Quantitative PCR generates threshold cycle (CT) values, which can be used to determine the relative concentrations of different transcripts in an individual, that were standardized against housekeeping genes and a common cDNA sample, allowing for a comparison of CT scores of “Peter Pan” larvae to normally developing postfeeding third instars (Chapter 5). Quantitative PCR is based on the



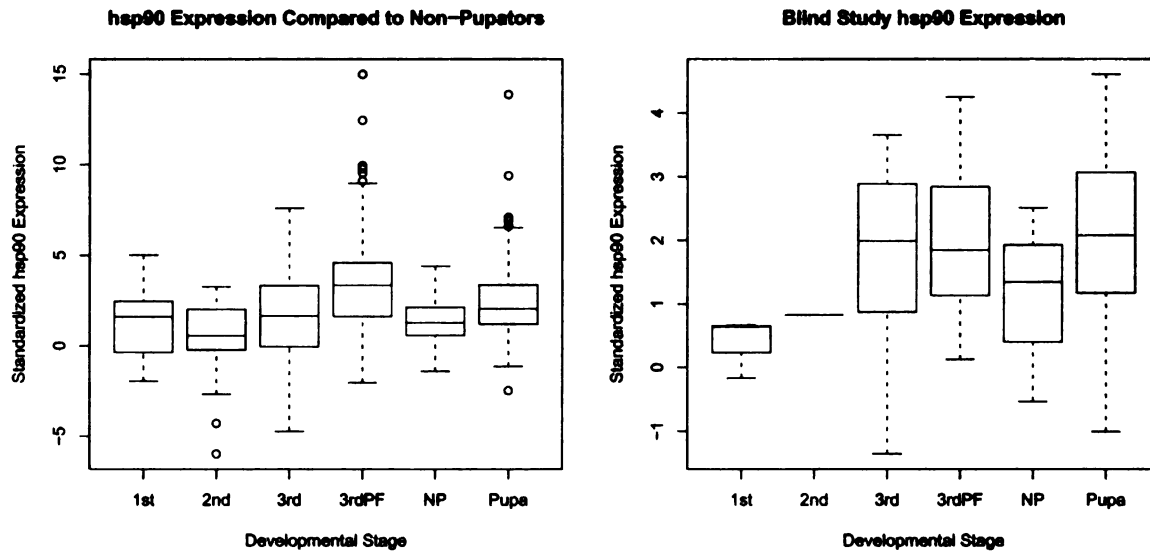
exponential increase of target DNA in a reaction, resulting in low CT scores for genes that are expressed at high levels. Analyzed individuals represented the three strains (CA, MI, WV) and two rearing temperatures (20°C and 33.5°C) studied in the previous experiment (Chapter 3). CT values of non-pupating larvae from 55 individuals were plotted by stage, to determine if any genes exhibited notable differences between normal postfeeding third instars and “Peter Pans”. For genes that exhibited differential expression between normal and non-pupating postfeeding third instars, expression was also compared to the profiles of “Peter Pan” larvae in a blind study (Chapter 6) to determine if it was possible to predict the state in unknown flies.

All boxplots of expression profiles were made in R (R core development team) with the blind study profiles shown in the right panel of the plots. Median values are represented by a horizontal line, the middle fiftieth percentile of CT scores is represented by the box, and the ninety-ninth percentile is contained within the whiskers of a plot. Expression profiles for a gene were considered different if the fiftieth percentile box for “Peter Pan” flies was beyond the median for that gene in postfeeding third instars.

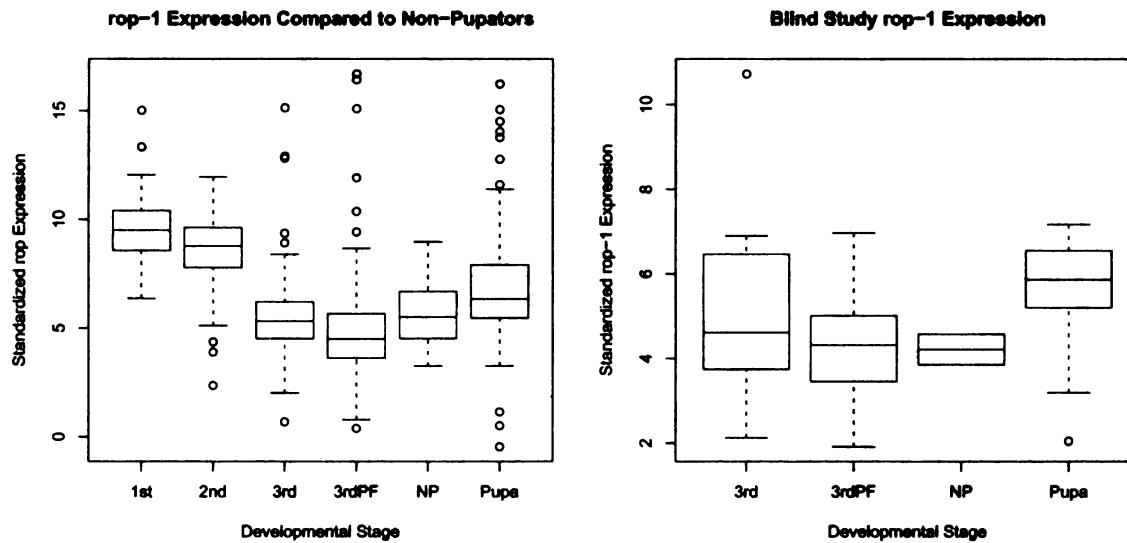
## *Results*

There were qualitative differences among the three strains studied (CA, MI, WV) for both the percent of 250 postfeeding third instars that were “Peter Pan” individuals, as well as mortality at 33.5°C. The CA strain was not likely to produce larvae that did not pupate and, as a result, 10 “Peter Pan” larvae could not be sampled from all of the CA replicates. Also, the WV strain was observed to experience much more mortality at the high temperature treatment than the other two strains.

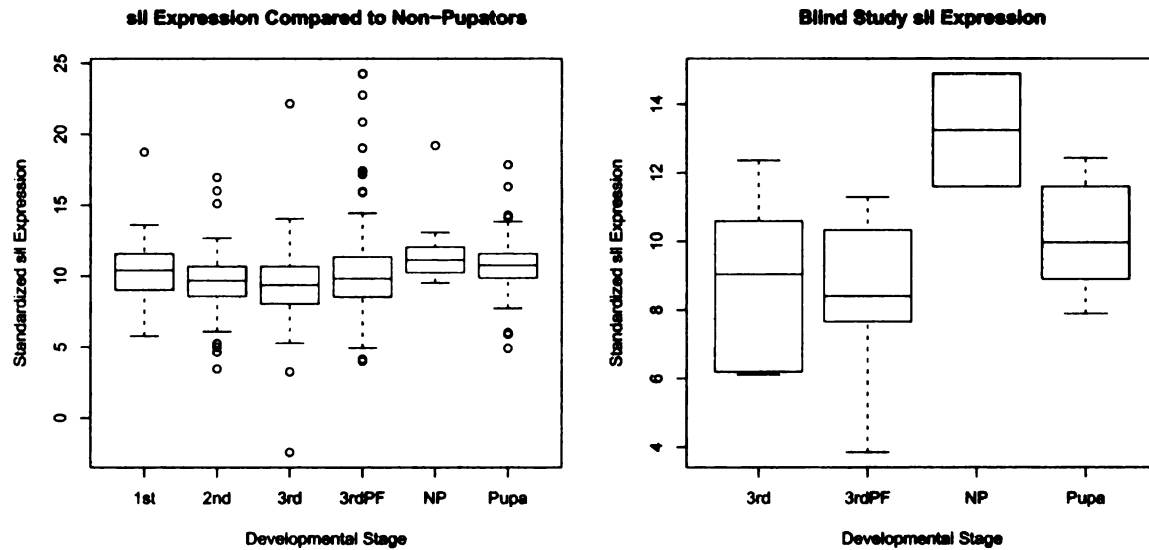
Nine gene expression profiles were produced for the “Peter Pan” larvae, allowing comparisons to the transcript level profiles of normally developing *L. sericata* in previous research (Chapter 5). Five of these genes were expressed within the range observed in postfeeding third instars, whereas four genes (*hsp90*, *rop-1*, *sll*, *usp*) were identified as having expression patterns that might be used to differentiate “Peter Pan” larvae from normal postfeeding third instars. The *hsp90* gene was up-regulated in larvae that had arrested development (Figure 7.1), while the other three loci were down-regulated as compared to postfeeding third instars (Figures 7.2–7.4). Expression plots (by stage) for these genes were shown with their counterpart plots in the blind study (Figures 7.1–4). Of the four genes that exhibited differential expression patterns in the 55 known non-pupators, three of them (*hsp90*, *sll*, *usp*) expressed similar, abnormal patterns in blind study “Peter Pans”, and could have been used to identify them as individuals that should not be incorporated into a PMI estimate.



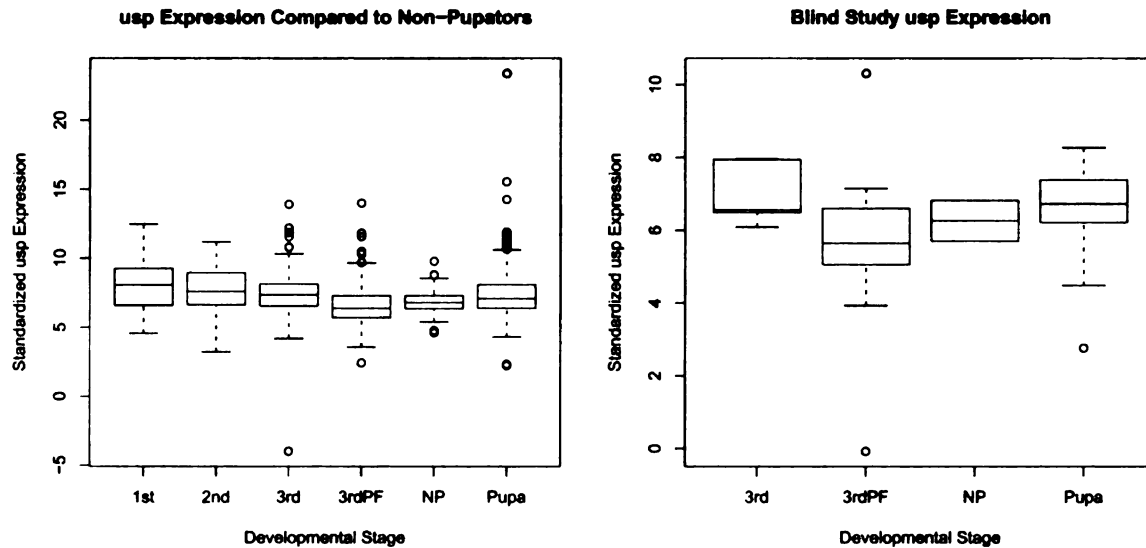
**Figure 7.1. Standardized gene CT scores for *hsp90*, plotted by stage, for all individuals from the previous section, for the 55 non-pupator individuals from this study, and the individuals in the blind study.** The left panel represents expression from a previous section (Chapter 5), with results from the 55 non-pupators (NP) placed next to postfeeding third instars (3rdPF). The right panel represents gene expression for the individuals in the blind study (Chapter 6), including the 4 non-pupators. 1<sup>st</sup> = first instar. 2<sup>nd</sup> = second instar. 3<sup>rd</sup> = feeding third instar. 3rdPF = postfeeding third instar. NP = non-pupator or “Peter Pan” larvae. On average, *hsp90* was expressed at higher levels in non-pupators than in postfeeding third instars.



**Figure 7.2. Standardized gene CT scores for *rop-1*, plotted by stage for all individuals from the previous section, for the 55 non-pupator individuals from this study, and the individuals in the blind study.** The left panel represents expression from a previous section (Chapter 5), with results from the 55 non-pupators (NP) placed next to postfeeding third instars (3rdPF). The right panel represents gene expression for the individuals in the blind study (Chapter 6), including the 4 non-pupators. 1<sup>st</sup> = first instar. 2<sup>nd</sup> = second instar. 3<sup>rd</sup> = feeding third instar. 3rdPF = postfeeding third instar. NP = non-pupator. On average, *rop-1* was expressed at lower levels in non-pupators than in postfeeding third instars, though this pattern was not apparent in the blind study.



**Figure 7.3. Standardized gene CT scores for *sll*, plotted by stage, for all individuals from the previous section, for the 55 non-pupator individuals from this study, and the individuals in the blind study.** The left panel represents expression from a previous section (Chapter 5), with results from the 55 non-pupators (NP) placed next to postfeeding third instars (3rdPF). The right panel represents gene expression for the individuals in the blind study (Chapter 6), including the 4 non-pupators. 1<sup>st</sup> = first instar. 2<sup>nd</sup> = second instar. 3<sup>rd</sup> = feeding third instar. 3rdPF = postfeeding third instar. NP = non-pupator. On average, *sll* was expressed at lower levels in non-pupators than in postfeeding third instars.



**Figure 7.4. Standardized gene CT scores for *usp*, plotted by stage, for all individuals from the previous section, for the 55 non-pupator individuals from this study, and the individuals in the blind study.** The left panel represents expression from a previous section, with results from the 55 non-pupators (NP) placed next to postfeeding third instars (3rdPF). The right panel represents gene expression for the individuals in the blind study, including the 4 non-pupators. 1<sup>st</sup> = first instar. 2<sup>nd</sup> = second instar. 3<sup>rd</sup> = feeding third instar. 3rdPF = postfeeding third instar. NP = non-pupator. On average, *usp* was expressed at lower levels in non-pupators than in postfeeding third instars.

## Discussion

### Potential Causes of the “Peter Pan” State.

The “Peter Pan” condition could result from a number of factors. Winter-induced diapause, a cessation of development brought on by cool temperatures and short day lengths, is known to occur in *L. sericata* and can vary widely among populations of this species (Greenberg and Kunich 2002). The light cycle used (12:12) in this study (see Chapter 3) and the low temperature treatment (20°C) are both capable of inducing this state (Tachibana and Goto 2004), indicating that “Peter Pan” larvae may be diapausing. Additionally, the CA strain produced less non-pupating larvae than the other strains, revealing differences in the genetic predisposition for producing larvae with arrested

development. Another difference among the strains also points to diapause, as the MI and WV strains were collected from regions that experience snow and prolonged periods of cold (thus necessitating an arrest of development to overwinter) and were more likely to produce “Peter Pan” larvae, suggesting that diapause is the cause of the “Peter Pan” state.

Alternatively, a summer diapause condition (known as estivation) might also be activated in these flies (e.g. Liu et al. 2006). The authors observed a delay in the development of the cotton bowl worm *Helicoverpa armigera* (Lepidoptera: Noctuidae) that was caused by high temperatures. Estivation seems unlikely in this instance, as 20°C is not warm enough to induce heat stress. Likewise, *L. sericata* are capable of developing at higher temperatures than 33.5°C (Grassberger and Reiter 2001). It should be noted though, that the WV strain experienced more mortality at the higher temperature treatment than the other two strains, indicating that heat stress may be possible for some strains at 33.5°C.

Last, the “Peter Pan” condition could be the result of mutations in genes that regulate development. As an example, *D. melanogaster insulin receptor (Inr)* mutations can delay larval development (Shingleton et al. 2005). It is possible that different alleles of genes that regulate development exist in natural populations in high enough frequencies to result in the observation of severely delayed development of larvae in some strains. There are myriad examples of stage-specific lethal mutants in *D. melanogaster* ([www.flybase.org](http://www.flybase.org)), including mutations of *dre4* (Sliter and Gilbert 1992), which is involved in the response to ecdysone, and E74A (Bialecki et al. 2002), which regulates the proper progression of immature developmental stages. Clearly, many genes

are required for the proper development of flies as they undergo molts and metamorphosis and genes that have mutant alleles with a lethal phenotype resulting in an inability of third instars to pupate could potentially explain the phenomenon of non-pupating larvae.

### **Interpreting Gene Expression Changes in “Peter Pan” Larvae.**

Some information on the “Peter Pan” condition can be gleaned from the gene expression changes that occur in these individuals, shedding light on the mechanism of the developmental delay. A potential molecular marker of diapause is *hsp90*, which was also profiled in this study, as it is up-regulated in larvae that are breaking diapause (Tachibana et al. 2005). *hsp90* expression in non-pupating individuals is commensurate with the breaking of diapause, since it was up-regulated in “Peter Pan” larvae. The “broken diapause” hypothesis is supported by the possibility that the initiation and breaking of diapause prolongs *C. vicina* development at low temperatures, as observed by Ames and Turner (2003). They noted an extremely long delay in development at very low temperatures, which was disproportionately large compared to the delay that is expected by decreasing temperature. The authors hypothesized that *C. vicina* begins and then breaks diapause when raised in environments that are close to the point at which the condition would be initiated, thereby increasing the delay in development. Such a delay could also be possible in *L. sericata* that were raised in environments that activate diapause, though the effects would be less pronounced in this experiment as the flies were not raised under “borderline” temperature conditions.

Alternatively, high *hsp90* levels could be symptomatic of a mutant condition that results in *hsp90* up-regulation. The simultaneous down-regulation of *usp* supports this



possibility, as *usp* and *ecr* interact to regulate other genes throughout fly development (Henrich and Brown 1995). These loci covaried with *hsp90* in 33.5°C temperature treatments (Chapter 5), raising the possibility that the gene is regulated by the ecdysone response, which is affected by juvenile hormone (JH) concentrations. High JH levels are known to prevent pupation (Kalthoff 2001) and *usp* is also a putative target of JH binding (Flatt et al. 2005). An interaction between *usp* and JH could explain the down-regulation of *usp* in flies that do not pupate if “Peter Pan” larvae were mutants for the JH response, possibly through faulty *usp* regulation.

The last gene that was down-regulated both in known and unknown samples was *sll*. It is expressed in the salivary glands of *L. sericata* larvae (Ali et al. 2005) that degrade during the postfeeding third instar. The “Peter Pan” state is an extremely prolonged postfeeding third instar, thus the down-regulation of *sll* in these flies is likely to be a result of extreme degradation of the salivary glands that express the gene, meaning that it is a good marker for the condition, but unlikely to reveal useful information as to why “Peter Pan” flies fail to pupate.

Given the reliable genetic regulatory changes associated with *L. sericata* larvae in arrested development, the cause of the “Peter Pan” state is probably due to one of two reasons: diapause or mutation. It seems most likely that the conditions of the experiment contributed the temporary initiation of diapause in *L. sericata* larvae. The light cycle (12:12) and one of the temperature treatments (20°C) are known to induce the state in on population of *L. sericata* (Tachibana and Goto 2004) and *hsp90* was up-regulated in larvae that experienced arrested development, in a manner that was consistent with the breaking of diapause (Tachibana et al. 2005). Alternatively, the inability to pupate may

be due to mutations in genes that regulate development, as supported by the down-regulation of *usp* in larvae that do not develop properly. If this is the case, there must be a selective reason for the maintenance of lethal alleles at high frequency in all three populations that were studied (CA, MI, WV). A selective advantage can occur if heterozygotes provide a beneficial phenotype as compared to homozygotes. Also, it is possible that there are lethal mutations that are linked to loci under strong positive selection, allowing them to “hitchhike” along with the gene under selection. Both scenarios are documented in population/quantitative genetic literature (reviewed in Conner and Hartl 2004) and could explain how lethal mutations may persist in three populations of blow flies, producing larvae that refuse to grow up.

#### **Forensic Analyses of Non-pupating Larvae.**

The analysis of “Peter Pan” larvae indicates that three of the nine loci examined, *hsp90*, *sll*, and *usp*, could be used to eliminate individuals in a state of arrested development from their use in age estimates, thus avoiding the increase in error associated with their inclusion. Based on the results, low *sll* and *usp* abundance, along with high *hsp90* concentrations, helped identify a “Peter Pan” larva. Recognition of expression patterns that are indicative of “Peter Pan” larvae will also enable investigators to identify collections that missed pupae, as “Peter Pans” would be collected at times when pupae and adults from their cohort should also be present. The analysis of expression levels for *sll*, *usp*, and *hsp90* may be instrumental in determining the developmental status of a postfeeding third instar larva and whether or not it should be used to predict a PMI.

## **CHAPTER 8: POPULATION AND TEMPERATURE EFFECTS ON BODY SIZE AND DEVELOPMENT TIME**

### *Introduction*

Understanding how ecological conditions drive the evolution of body size and development time is a fundamental topic of research in evolutionary biology (Oudman et al. 1991, Morey and Reznick 2000, Kingsolver 2001, Day and Rowe 2002, Hallas et al. 2002, Feder et al. 2003, Conner and Hartl 2004, Plaistow 2004, Gomez-Mestre and Buchholz 2006). One area of such research focuses on the fact that, when faced with limited resources, many organisms that go through a metamorphosis experience a tradeoff between developing quickly (ensuring survival to adulthood) and producing larger, and more successful adults by delaying maturation. Spadefoot toads, which are desert-dwelling and develop in vernal ponds, have been extensively studied in this capacity, with different species delaying development in low resource conditions, while others sacrifice body size in low nutrient conditions (Morey and Reznick 2000). Spadefoot toads have much faster development times than ancestral toad species and evolutionary relationships among them explain some of the differences in size and larval growth periods (Gomez-Mestre and Buchholz 2006). Though the tradeoff between body size and development time is documented, attempts to model it have been only partially successful, as some species will produce the largest individuals when growth is at its fastest, which led to hypotheses that developmental thresholds (minimum requirements for life history events to occur) must be involved in the physiological processes that determine body size and development time, with selection for certain thresholds resulting in different relationships between size and age at maturity (Day and Rowe 2001). The

effects of thresholds in development have been partially upheld by research in the mite species *Scancassania berlesei* (Michael), which noted the predicted shift in reaction norms for development that regulated by thresholds, but also observed effects of maternal nutrition and sex on the traits (Plaistow et al. 2004). Additionally, work on *Pieris rapae* has demonstrated that thermal performance curves are a result of selection pressures on development that can also affect the evolution of body size (Kingsolver 2000, Kingsolver 2001).

The ephemeral nature of blow fly resources (carrion and biological waste), which occur sporadically in the environment and disappear quickly, and the global distributions of some blow fly species could lead to selection for different body sizes and development times in different populations. One key feature of organisms that experience body size/development time tradeoffs is the evolution of plasticity (environmentally induced changes in phenotypes; Day and Rowe 2002), which has been demonstrated in the blow fly *Lucilia sericata* (Diptera: Calliphoridae) (Meigen) (Clark et al. 2006, Tarone and Foran 2006). Size and age at maturity in *L. sericata* are responsive to changes in laboratory rearing conditions such as moisture and type of food respectively. The evolution of plasticity and the potential for a tradeoff between body size and development time in this species mean that *L. sericata* is another possible model for studying the evolution of age and size at maturity.

Studying blow fly development also has applied uses. Blow flies cause myiasis (parasitization) in humans (Faulde 2001), sheep (East and Eisemann 1993), and cattle (Ramirez and Crystal 1975). Also, the predictable development of blow flies in optimal growth conditions can be used as a biological clock to help determine a postmortem

interval (PMI) in death investigations (Catts and Haskell 1990, Greenberg and Kunich 2002). Both myiasis and PMI predictions are intimately associated with the immature development of blow flies, thus any contribution to the understanding of development also contributes to the ability of researchers to improve PMI predictions and limit the damage caused by myiasis.

The research presented here stemmed from previous observations of developmental variation among populations of the forensically informative blow fly *L. sericata*. Different authors have reported varying minimum development times for this species, with some indicating faster minimum development times for this species (Kamal 1958, Greenberg 1991, Anderson 2000, Grassberger and Reiter 2001), as well as differences in the development times of Russian and US populations (but, with little detail explaining the support for the observation; Greenberg 1991). Some of the variation among studies may be due to differences in laboratory rearing conditions known to affect development time of *L. sericata* (Clark et al. 2006, Tarone and Foran 2006), though other factors may also be involved. In previous research, the effects of temperature and biological strain were considered throughout development (Chapter 3, Chapter 5), and consistent differences were observed between cohorts of *L. sericata* raised at different temperatures and among three biological strains originating from California, Michigan, and West Virginia. Low temperatures prompted larger third instar and pupal sizes (length and weight), and the California strain was consistently the largest, while the West Virginia strain was consistently smallest. Differences among the strains appeared to be due to variation in the timing of the cessation of feeding during the third instar (Figure 3.3). The observed effects of temperature and strain on the age and size at *L. sericata*

maturity suggest that both the environment and genotype are important considerations when predicting minimum development times of this species. As with body size, gene expression was also susceptible to temperature and strain effects (Chapter 5), further supporting the idea that different populations of *L. sericata* exhibit variable developmental patterns. Indeed, variation in body size and development time (the core phenotypes of interest to forensic entomologists) has been observed in multiple Cyclorrhaphan species (reviewed in Tarone and Foran 2006).

The explanation for temperature and strain effects on body size, development time, and gene expression levels, can be found in quantitative genetic theory (reviewed in Conner and Hartl 2004). Any continuously variable phenotype, such as length or minimum development time, is subject to the influence of genetics, the environment, and the interaction between the two, which can be detected through reaction norm plots and ANOVA (Scheiner and Gurevitch 2001, Conner and Hartl 2004). Because forensic entomologists use quantitative traits (development time, length, weight) to determine blow fly age predictions (Kamal 1958, Greenberg 1991, Anderson 2000, Grassberger and Reiter 2001), the phenotypes can be studied with a quantitative genetic approach to determine if and how the environment and genotype affect them. Knowledge of such factors can increase the accuracy of forensic entomological predictions of blow fly age by identifying genetic and environmental conditions that affect the maturation of the Calliphoridae.

Genetic variation for body size and development time must also be considered in terms of new approaches designed to more precisely predict development percents of blow flies based on gene expression data (Chapter 5). If minimum development times

are the same across populations, development percent predictions do not need to be calibrated to biological strain. However, if minimum development times are quite different across populations, predictions will be indicative of different ages for different strains, requiring a calibration appropriate to the strain of interest. A calibration would pose a problem to the forensic science community, as it would require a database of development times for any region where forensic entomology is employed or post priori studies to determine the specific minimum development time of a population of interest. Though the differences in body sizes and development times were consistent after hundreds of measurements for all strains/temperatures (within replicates), the number of replicates per temperature per strain (2) was small, and variability could theoretically be due to chance. To definitively demonstrate that the growth of individuals from each of the strains is affected by genotype and temperature, it was necessary to analyze a larger sample size.

The research presented here was designed to more rigorously test the hypothesis that genotype and the environment (temperature) affect the development time and body size of *L. sericata*. New flies were collected from Davis, CA; East Lansing, MI; and Morgantown, WV. Measurements of body size in 720 individuals and minimum development times from 60 replicates were analyzed to determine that biological strain and temperature affect these traits. Knowledge of such effects will be instrumental in decreasing error in PMI estimates made with blow fly data and may contribute to a better understanding of the evolutionary of tradeoff between body size and development time by providing data on this phenomenon in a new species.

## *Materials and Methods*

New collections of *L. sericata* were made in the summer of 2006 from the same locations as the flies studied in Chapter 3 (which occurred in 2005), and they were identified as *L. sericata*, through visual identification and *cytochrome oxidase 1* sequence (Tarone and Foran 2006). Ten cohorts of each strain were induced to lay eggs as previously described (Chapter 3). A cohort was assigned to a temporal block, which was composed of all cohorts laid on that day. Cohorts were split between 20°C and 33.5°C temperature treatments, and were raised as described in Chapter 3, except that pupae were left in the pupation substrate, not destructively sampled. For a replicate, twelve individual pupal lengths (to the nearest 0.5 mm) and weights (to the nearest 0.01 mg) were measured, as previously described (Chapter 3), on the 3<sup>rd</sup> day of pupation. Likewise, the minimum development time (to the nearest 0.5 hours) for a replicate was recorded.

This scheme resulted in 720 measurements of individual size: 240 for a strain, with 120 measurements in a temperature treatment. Similarly, 20 minimum development times per strain were recorded (10 per temperature treatment). Since a block was split among 20°C and 33.5°C treatments, which were composed of multiple replicates of a strain, the data were analyzed in R (R core development team) using a split-plot ANOVA (Scheiner and Gurevitch 2001), with strain nested within temperature, which was nested within temporal block. Mean, median, minimum, and maximum values for traits were reported for the 2006 study. Pupal data from Chapter 3 were also analyzed to determine if results were similar between experiments, but these data were not considered as thoroughly, as they were based on two replicates per strain.



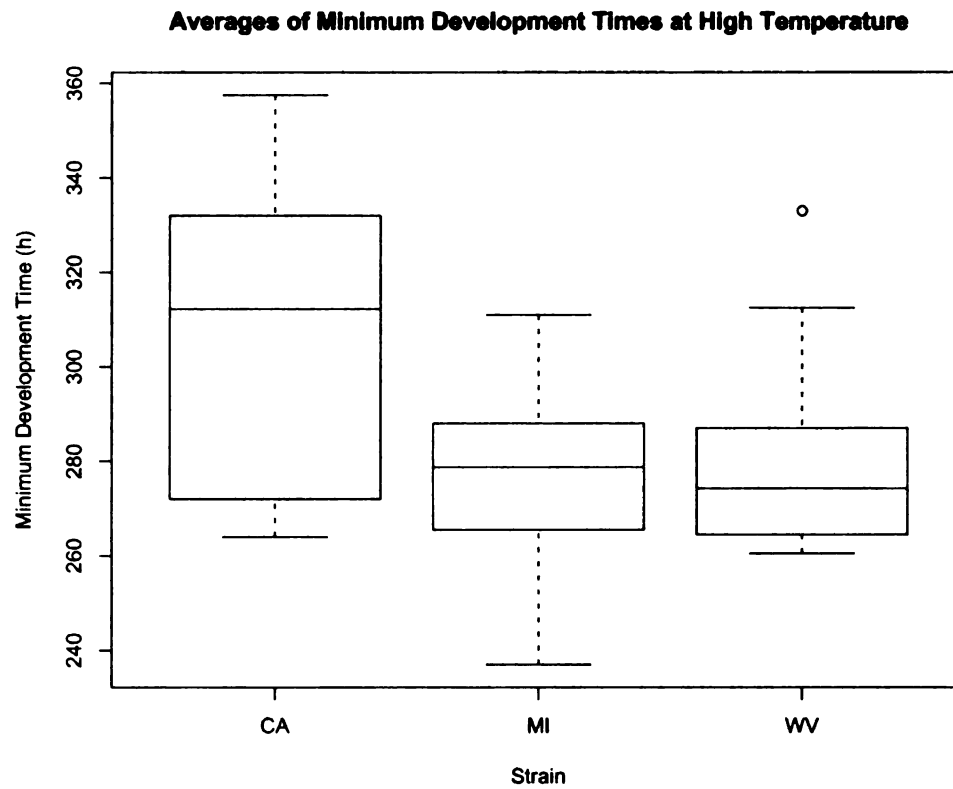
Boxplots (described in Chapters 2 and 7) and reaction norm plots were constructed in R to compare weight (a more precise measure of size than length, which was estimated to the nearest 0.5 mm) and development time variation among strains and between temperatures. For all reaction norm plots, the Davis, California strain (CA) is represented by red, the East Lansing, Michigan strain (MI) is represented by blue, and the Morgantown, West Virginia strain (WV) is represented by green.

### *Results*

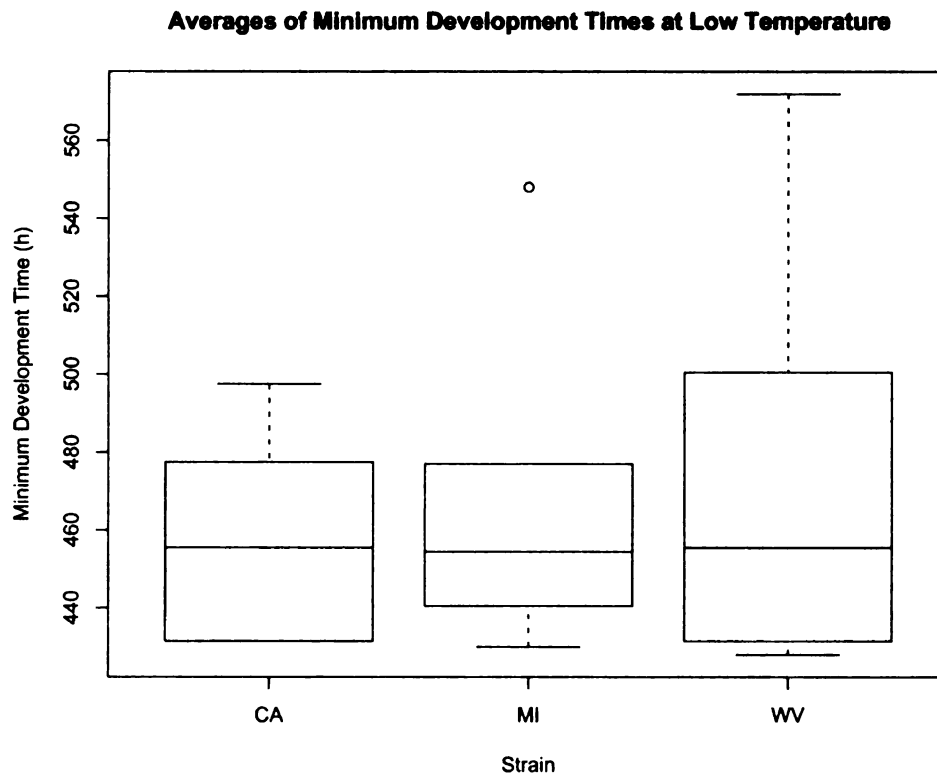
All three strains were identified as *L. sericata*. Sequence from all strains resulted in BLAST hits to other published *L. sericata cytochrome oxidase 1* sequences. A 361 bp sequence from the CA strain was 100% similar to known *L. sericata* sequence (accession number DQ868523). A 762 bp sequence from the MI strain was 100% similar to known *L. sericata* sequence (accession number L14947). A 494 bp sequence from the WV strain was 100% similar to known *L. sericata* sequence (accession number DQ868523). In all cases, the most similar sequences in the BLAST searches were *L. sericata* sequences, which were distinguishable from *L. cuprina* sequences by at least 1% sequence identity.

Minimum development times for the 2005 results were within the ranges of development times for the 2006 experiment. At 33.5°C, the mean minimum development time was 257.9 hours (minimum: 238 hours, maximum: 283.5 hours) for the 2005 experiment and 288.7 hours (median: 283.4 hours, minimum: 237 hours, maximum: 357 hours) for the 2006 experiment. When strains in the 2006 experiment were compared (Figure 8.1), CA developed slower (mean: 308 hours, median: 312.3 hours, minimum:

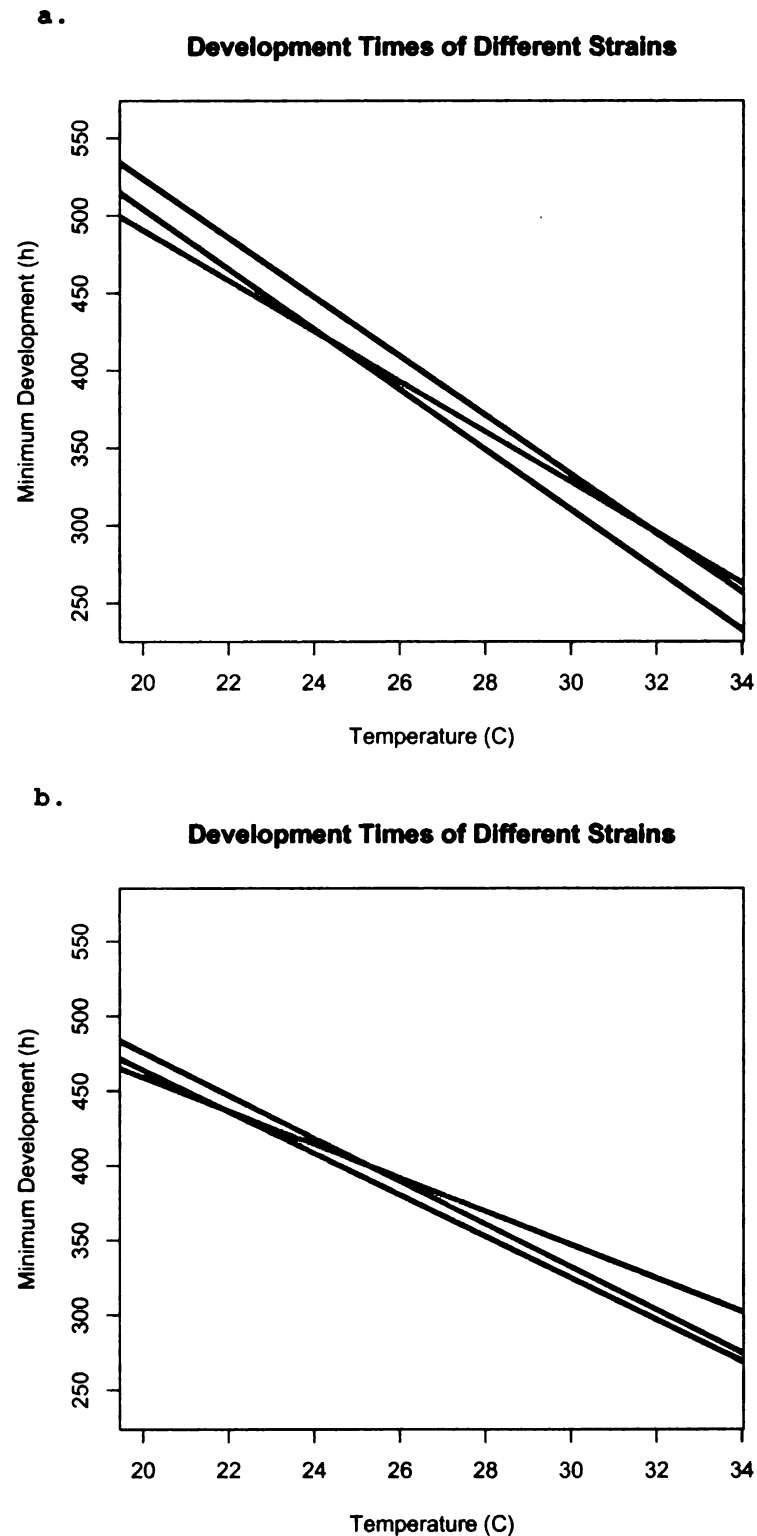
264 hours, maximum: 357.5 hours) than MI (mean: 276.1 hours, median: 278.8 hours, minimum: 237 hours, maximum: 311 hours) or WV (mean: 282.1 hours, median: 274.3 hours, minimum: 260.5 hours, maximum: 333 hours) in the high temperature treatment. At 20°C, the mean minimum development time for the 2005 experiment was 503.7 hours (minimum: 468.5 hours, maximum: 561 hours) and 466.1 hours (median: 455.5 hours, minimum: 428 hours, maximum: 572 hours, Figure 8.2) for the 2006 experiment. In the low temperature treatments, the CA strain (mean: 458.9 hours, median: 455.5 hours, minimum: 431.5 hours, maximum: 497.5 hours) developed, on average, faster than MI (mean: 463.9 hours, median: 454.5 hours, minimum: 430 hours, maximum: 548 hours) and WV (mean: 475.5 hours, median: 455.5 hours, minimum: 428 hours, maximum: 572 hours), as the others had some recorded minimum development times above 500 hours (Figure 8.2–3). Similarly, all rank order comparisons of mean development times among the strains were maintained between 2005 and 2006 with CA developing fastest at 20°C and slowest at 33.5°C, while MI grew fastest at 33.5°C and at an intermediate pace when raised at 20°C (Figures 8.1–3).



**Figure 8.1. Boxplots of minimum development time comparisons among strains at 33.5°C in 2006. The CA strain developed more slowly compared to the other strains.**



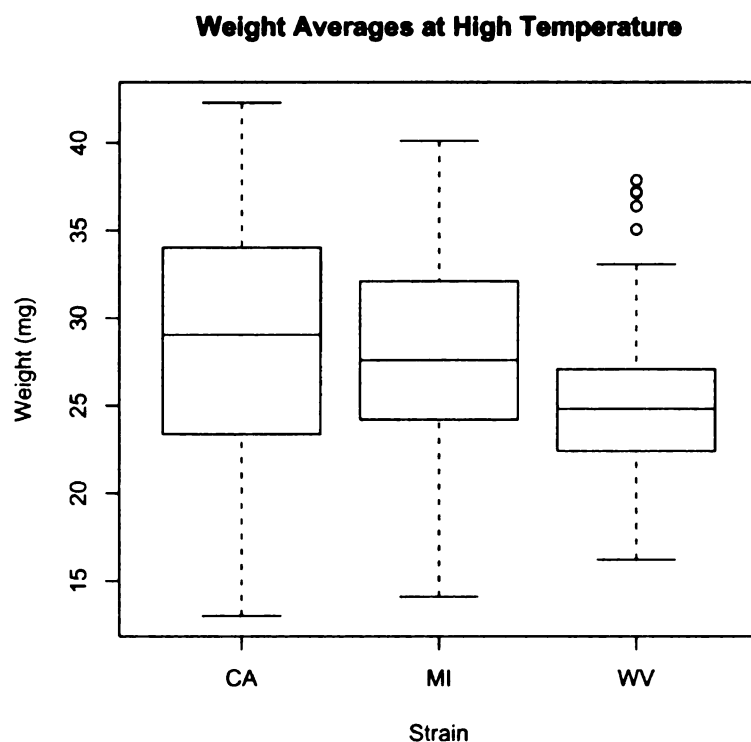
**Figure 8.2. Boxplots of minimum development times comparisons among strains at 20°C in 2006.** The strains exhibited similar median and fastest development times, but CA had a faster mean development time as compared to the other strains, which had some minimum development times that were longer than 540 h.



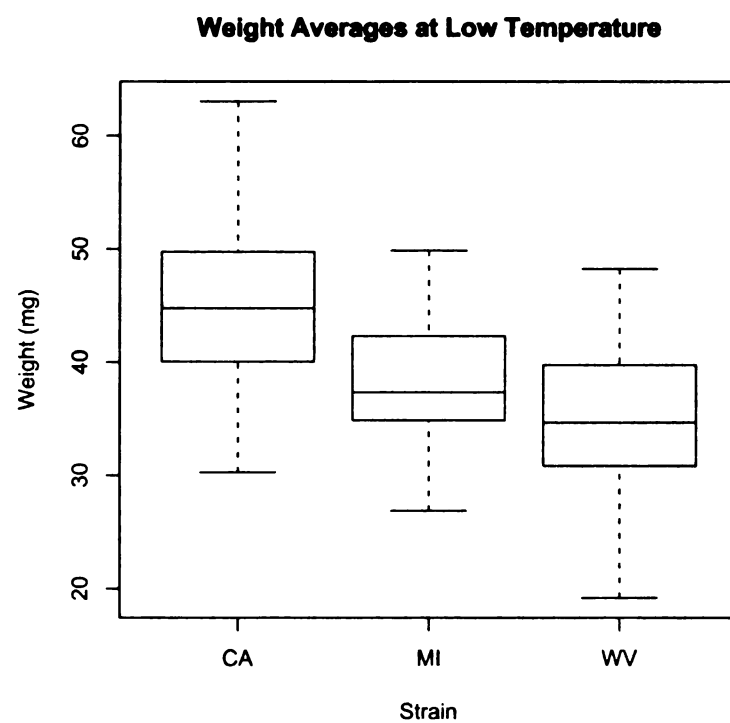
**Figure 8.3. A comparison of average minimum development times for each strain in 2005 (a) and in 2006 (b).** CA is red, MI is blue, WV is green. The MI mean minimum development time was consistently faster than WV, though their development times were comparable. The CA mean was fastest at 20°C and slowest at 33.5°C in both years. Development was faster at 33.5°C.

Body size comparisons between 2005 and 2006 were largely similar, though the 2006 experiment yielded larger animals than the 2005 experiment. When raised at 33.5°C, the mean length was 6.9 mm (minimum: 5.5 mm, maximum: 8 mm) in 2005 and 7.2 mm (median: 7 mm, minimum: 6 mm, maximum: 9 mm) in 2006, while mean weights were 23.12 mg (minimum: 10.66 mg, maximum: 49.14 mg) for 2005 and 27.29 mg (median: 38.3 mg, minimum: 19.18 mg, maximum: 63.03 mg) for 2006. Average weights at the high temperature were largest for the CA strain (mean: 28.63 mg, median: 29.04 mg, minimum: 13.01 mg, maximum: 42.28 mg), intermediate for MI (mean: 28.16 mg, median: 27.6 mg, minimum: 14.1 mg, maximum: 40.11 mg), and smallest for the WV strain (mean: 25.08 mg, median: 24.81 mg, minimum: 16.22 mg, maximum: 37.86 mg) in the 2006 data (Figure 8.4). When *L. sericata* was raised at 20°C mean length was 7.4 mm (minimum: 5.5 mm, maximum: 10 mm) for 2005 and 7.9 mm (median: 8 mm, minimum: 6.5 mm, maximum: 9.5 mm) for 2006 and average weights were 29.05 mg in 2005 (minimum: 10.66, maximum: 49.14) and 39.36 mg (median: 38.3 mg, minimum: 19.18 mg, maximum: 63.03 mg) in 2006. The mean weights of low temperature treatments in 2006 revealed CA to be the largest (mean: 44.8 mg, median: 44.77 mg, minimum: 30.27 mg, maximum: 63.03 mg), MI to be intermediate in size (mean: 38.33 mg, median: 37.33 mg, minimum: 26.86 mg, maximum: 49.86 mg), and WV to be the smallest strain (mean: 34.96 mg, median: 34.66 mg, minimum: 19.18 mg, maximum: 48.25 mg; Figures 8.5). Strain effects on body size followed this trend as well; when both measures for body size were compared among strains, the CA strain was

consistently the largest and the WV strain was consistently the smallest in 2006 (though the WV average length was longer than MI in 2005) (Figures 8.6–7).

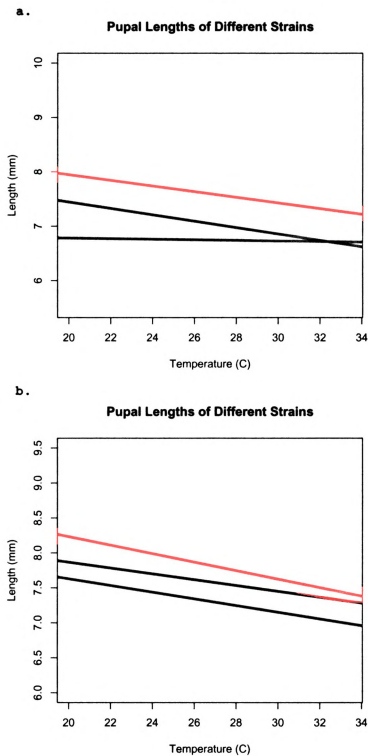


**Figure 8.4. Boxplots of average weights compared among the strains at 33.5°C in 2006. On average, CA was larger than MI, which was larger than WV.**



**Figure 8.5. Boxplots of average weights compared among strains at 20°C in 2006.** On average, CA was larger than MI, which was larger than WV.

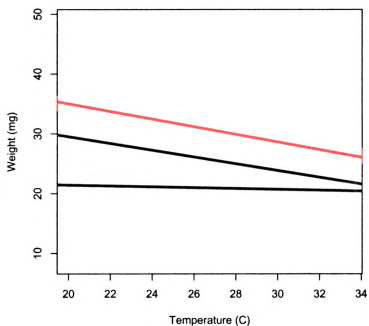




**Figure 8.6. A comparison of average lengths for each strain in 2005 (a) and in 2006 (b).** CA is red, MI is blue, WV is green. The CA mean length was the largest in both years. WV was the smallest in 2006, but MI was smallest at 33.5°C in 2005. Lengths were greater for 20°C treatments.

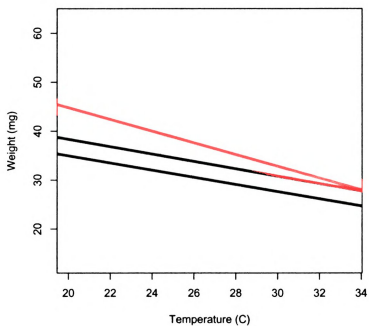
a.

### Pupal Weights of Different Strains



b.

### Pupal Weights of Different Strains



**Figure 8.7. A comparison of average weights for each strain in 2005 (a) and in 2006 (b).** CA is red, MI is blue, WV is green. The CA mean weight was the largest and WV was the smallest. Weights were greater for 20°C treatments.

In both years, ANOVA demonstrated the significance of the observed differences among strain and between temperatures on body size and minimum development time. A comparison of pupal sizes and minimum development times from 2005, though based on only two replicates per temperature per strain, resulted in a statistically significant effect of strain on length ( $P=0.034$ ) and a borderline significant effect of temperature on length ( $P=0.085$ ). Results were similar when weight was assessed, with a significant effect of strain on weight ( $P=0.03$ ) and a borderline significant effect of temperature ( $P=0.08$ ). However, when minimum development time was assessed, there were significant strain ( $P=0.001$ ) and temperature ( $P=0.03$ ) effects on the trait, as well as a significant interaction between temperature and strain ( $P=0.0071$ ).

The comparisons of the 2006 strains showed statistically significant effects of strain and temperature on all three traits assessed. Length was significantly affected by strain ( $P<0.0001$ ) and temperature ( $P<0.0001$ ). Likewise, weight was affected by strain ( $P=0.0079$ ) and temperature ( $P=0.0001$ ). Minimum development times, though they were significantly affected by strain ( $P=0.049$ ) and temperature ( $P<0.0001$ ), were not significantly affected by the interaction between the two variables ( $P=0.2$ ).

## *Discussion*

### **Temperature Effects on Body Size.**

The observation of high temperatures resulting in smaller body size is consistent with a pattern that has been recorded for approximately 80% of insect species (Sibly and Atkinson 1994). These observations present a paradox, since faster growth promotes larger body sizes when development is accelerated by improved nutrition (Berrigan and

Charnov 1994). Many attempts have been made to model development to understand why insect sizes decrease in higher temperatures, seeking to explain body size as a function of factors such as larval mortality, limited lifespans, gonad development, increased molecular metabolism, and spatial/temporal heterogeneity in the environment, (Berrigan and Charnov 1994, Sibly and Atkinson 1994, Perrin 1995, Atkinson and Sibly 1996). Two factors, explored in these analyses could potentially explain the phenomenon in *L. sericata*. First, development is dependent on biochemical metabolism and the mechanical acquisition of food. If temperature increased the rate of feeding at a slower rate than the catalysis of food, flies may be smaller due to a proportionally larger increase in metabolism. This theory has been supported by observations of a slower increase in fish feeding as compared to metabolic activity in warmer conditions (Perrin 1995), and modeling research on growth (Perrin 1995, Atkinson and Sibly 1996). In *L. sericata*, there is a proportional increase in the rate of growth for high temperature treatments as a function of development percent (Figure 3.3), indicating that a disproportionately large increase in molecular metabolism may affect blow fly size. However, higher mortality at high temperatures has also been proposed as a mechanism for promoting the paradoxical growth observations associated with higher temperatures (Berrigan and Charnov 1994, Sibly and Atkinson 1994). Essentially, if larvae are more likely to die in warmer conditions than in cooler temperatures, due to factors such as heat stress or anoxia, then there may be selection to complete development quickly at the expense of body size. The observation of higher mortality at 33.5°C, especially in the smallest strain (WV), supports this theory. Given these results, it is likely that one or both of the mechanisms is driving the evolution of small body sizes at high temperatures.

### **The Evolution of Age and Size at Metamorphosis.**

The research presented here demonstrates several interesting features that help to understand the evolution of *L. sericata* age and size at metamorphosis. First, body size was affected by both genetic strain and temperature. The CA strain was consistently the largest, while WV was smallest. Likewise, lower temperatures resulted in larger pupae. Second, there was no straightforward relationship between body size and development time. At 20°C, the smallest strain developed for the longest period of time and the largest matured, on average, in the least amount of time. Fast development times of larger strains indicate the presence of selection on developmental thresholds as proposed by Day and Rowe (2001). However, at 33.5°C, MI developed fastest, while CA took the longest to mature. The change in rank order of minimum development times, which occurred in the same pattern for both years, suggests that the nature of selection on the threshold, or thresholds, influencing development time/body size in *L. sericata* is different among strains. The selection for fast development times in one temperature and not the other is evidence that the ecosystems in which the studied strains live may be pushing them toward different development times. Variable selection may also be pushing the strains toward alternative optimal temperatures for body size, as seen in *P. rapae* (Kingsolver 2000, Kingsolver 2001). Differential selective pressure is not hard to imagine given the dissimilarities between the ecoregions that each strain originated from (Figure 1.3).

Many ecological factors can potentially influence fly development, but one, latitude, is known to affect body size and development time in *D. melanogaster* (Oudman et al. 1991, Hallas et al. 2002) and *Rhagoletis* (Feder et al. 2003) respectively. The

effects of latitude differences on size and age at metamorphosis in fly species indicate that it may be possible to determine a series of guiding principles for predicting minimum development time/body size based on the ecological factors (including latitude).

However, such effects are likely to be specific to blow flies. For example, body size is known to increase with latitude in *Drosophila* (e.g. Hallas et al. 2002), but the largest (CA) and smallest (WV) strains originated from similar latitudes, indicating a different mode of selection for body size in *L. sericata*. Accordingly, comparisons of *Lucilia* populations to *Drosophila* populations will likely lead to a fuller understanding of how the environment and genotype influence Dipteran phenotypes.

#### **Year-to-Year Variation.**

One interesting feature of the research presented here was the general similarity between the 2005 and 2006 studies. CA was consistently the largest and, with the exception of length comparisons in 2005, WV was consistently the smallest strain. Likewise, the rank order of mean development times was the same in both years.

However, despite these comparative continuities, there were differences in average minimum development times and sizes among strains collected in different years. There are several potential explanations for this observation. The easiest explanation is that the small sample size (2 replicates per temperature per strain) in the 2005 experiment was randomly biased as compared to the 2006 experiment, which had 10 replicates per temperature per strain. Another possibility is that the destructive sampling in the 2005 study contributed to slower pupal development. This idea is supported by the observation that destructive sampling delays the appearance of adults (Tarone and Foran 2006). However, the fact that the 33.5°C treatments were more similar to each other than the

20°C treatments and that the 2005 study took sampling into account (Chapter 3), indicates that destructive sampling did not substantially contribute to the slower 20°C development observed in the 2005 study. Alternatively, the differences could be due to year-to-year genetic variation within each population. This is supported by the observation of short-term and long-term changes in chromosomal inversion frequencies in a population of *Drosophila subobscura* (Rodriguez-Trelles et al. 1996), indicating the possibility that the differences observed herein were the result of a temporal shift in local development times and body sizes. However, the three strains were collected in different US states, but exhibited similar changes of approximately 50 hours in mean development times at 20°C (Figure 8.3) between 2005 and 2006. If temporal variation explains the differences among strains, it seems unlikely that three different strains experienced similar shifts in development times unless a major climatic effect influenced each.

### **Forensic Entomology.**

The research presented here demonstrated several important features of *L. sericata* developmental variation, which are useful for understanding blow fly body size and maturation times in the context of forensic entomology. First, significant effects of both temperature and genetic strain on the traits were observed, which have clear ramifications for their use in the forensic sciences. The subtle differences among the three strains and two temperatures for body size were previously shown to have little effect on predictions of minimum development percents (Chapters 3 and 5), but the best estimates of age, when using body size measures, may result from populations with known average body sizes. More importantly, it is clear that the predictions of minimum development percent will be more accurate if they are calibrated to the strain of interest.





One can never know the development times of all populations in every year, but it may be possible to understand the rules governing the evolution of this trait. In determining those rules it will be possible to decrease errors in age predictions resulting from populations that differ from laboratory-based tables of development times.

The variation in minimum development times among strains may be seen as detracting from the significance of the improved precision in age predictions garnered through gene expression analyses (Chapter 5, Chapter 6), but these are, in reality, two different problems with the PMI prediction process. The first problem in predicting a PMI with blow fly evidence is an issue of precision rooted in developmental biology and stems from the need to understand how mature a blow fly happens to be. For example, if a blow fly develops in 500 hours, an investigator would prefer that an estimate of 70% through development be accompanied by a window of 325–375 hours as opposed to 275–425 hours. If accurate, a shorter window of time around a PMI estimate allows the investigator to rule out more potential suspects and potentially focus on the correct suspect in a death investigation. The precision of PMI estimates with blow fly evidence has been improved with recent work that utilizes gene expression data to predict age (Chapter 5, Chapter 6) and this problem in estimating blow fly age is likely to decrease as molecular and statistical tools develop.

However, the second problem with predicting a PMI with blow fly evidence is an issue of accuracy that is rooted in population/quantitative genetics and stems from the fact that phenotypes can vary among blow fly strains, which can lead to an inaccurate PMI estimate if the development of a specific population is not in accordance with the data used to make an age prediction. As an example, the ability to precisely predict a

development percent of 70% does not mean that the predicted development percent is the same for two different populations. At a given temperature, if one population develops in 500 hours, while another develops in 470 hours, an estimated development percent of 70% would correspond to an age of 350 hours and 329 hours (respectively), potentially resulting in an inaccurate estimate of age. The occurrence of strain effects on minimum development time in this research indicates a need for population-specific data to avoid inaccuracy in PMI estimates.

Population-specific predictions of age may be aided by evolutionary ecology theory, which seeks to explain phenotypic variation in terms of ecological effects on those traits (reviewed in Conner and Hartl 2004). By studying a series of populations, the guiding principles behind the evolution of development times in *L. sericata* (and other blow flies) can be determined, allowing for more accurate PMI predictions. Such rules can be used to calibrate estimates with published data to the population studied. One possible means of developing a strain-specific PMI would be to combine predicted development times with published data, which would be used as a prior in Bayesian analyses (Scheiner and Gurevitch 2001), thereby developing a posterior blow fly age estimate tailored to the population of interest. By determining the factors that affect blow fly development times, more accurate PMI estimates can be determined with entomological data.

## CHAPTER 9: DISCUSSION

### *Improved Predictions of Blow Fly Age*

#### **Introduction.**

The goal of the research detailed herein was to improve PMI predictions that are based on blow fly evidence. There are several challenges facing traditional forensic entomological PMI predictions that can be overcome through the application of knowledge gained from the results of this dissertation. First, both the accuracy and precision of blow fly age estimates can be increased by employing molecular and quantitative genetic methods. Any research that contributes to a fuller understanding of factors that affect development times in *L. sericata* also reveals information that helps to make more accurate age predictions, as an entomologist will be better equipped to estimate the minimum development time of a blow fly in a variety of conditions and for different populations. Similarly, research that can help to decrease the window of time provided with a PMI prediction will increase precision. Second, two parts of the Daubert requirements for scientific evidence, standard operating procedures and a statistical understanding of methods, have not been fully considered in forensic entomology. Both of these requirements were addressed by the research conducted herein. The research evaluating environmental and genetic effects on variation in blow fly development demonstrated a need for standard operating procedures and will be useful in developing them, as the results presented can be used to design standard procedures for raising blow flies in the laboratory. Likewise, the use of GAMs enabled a statistical understanding of age estimates made with blow fly data. Finally, the methods employed in the dissertation

provide logistical improvements over other forensic entomology approaches that make their application appealing. The widespread use of nucleic acid technologies in forensic science laboratories means that most can employ the molecular techniques used for aging blow flies and since they have the potential to increase laboratory productivity, they will be viewed by investigators as an attractive option to pursue.

### **Improving Accuracy and Precision in Forensic Entomology.**

The results of the research presented here show that it is possible to achieve more accurate and precise PMI estimates with entomological data. The use of gene expression levels and GAMs to predict blow fly age led to a clear improvement in GCV and PDE statistics, particularly for postfeeding third instars and pupae. The statistics provide information that can be used to compare models predicting a response of interest (in this case, minimum development percent of *L. sericata*) and show that blow fly age estimates that include gene expression analyses are more precise than those that do not. The use of GAMs also enables a graphical assessment of the increase in precision garnered by gene expression data, as seen by a comparison of the fitted vs. response plots (Chapter 5). The accuracy of PMI predictions was also aided by molecular analyses. In assessing gene expression differences among feeding, postfeeding, and “Peter Pan” third instar larvae, the ability to distinguish amongst them was enhanced, thereby decreasing errors in stage identification that would result in inaccurate predictions.

The quantitative genetic assessment of blow fly development will also lead to increases in the accuracy of PMI predictions. The research investigating rearing conditions that affect development time (Chapter 2) and the effects of genotype and the environment on development time and body size (Chapters 3 and 8) produced results that

will aid in correctly identifying the underlying factors that contribute to ambiguities between fly growth seen in practice and growth as it appears in the literature. Any information on factors (both genetic and environmental) that influence the two phenotypes of greatest interest to a forensic entomologist, body size and development time, will help investigators produce better theories on the evolution of blow fly development that can be used to achieve more accurate PMI estimates. Likewise, the ability to recognize environmental conditions that influence development may help investigators in determining the most relevant developmental data for a case.

### **Addressing the Daubert Standards in Forensic Entomology.**

The results of the research presented here also helped address two of the requirements of the Daubert ruling, standard operating procedures and an understanding of error. The reasons for, and results of, the research in Chapter 2 demonstrate the need for standard operating procedures in rearing protocols used to produce development tables for forensically informative blow flies. Clearly, there is plasticity in fly development, as it was possible to achieve a large range of replicable development rates depending on the rearing treatments used to raise *L. sericata*. The comparison of development of flies on rats to flies raised on liver indicates that only a small subset of liver-based rearing protocols, in particular ones that keep the food source moist, produce growth that is comparable to growth under forensically relevant conditions (growth on a body). This demonstrates that the a standard operating procedure for the rearing of flies in the laboratory should produce growth rates that mimic the development of those flies under the most forensically relevant conditions—on carrion. Likewise, as more

quantitative genetic influences on blow fly development are revealed, they can be incorporated into the protocols for predicting a PMI with entomological data.

The application of molecular biology techniques has also made the analysis of blow fly data more amenable to standard operating procedures. The molecular biology community has detailed guidelines for the methods used herein. When DNA identifications of individuals were a new tool in the forensic sciences, a National Research Council report was compiled by a group of experts in the fields relevant to the new technique and its application. The report outlined a series of recommendations that resulted in a set of standards for forensic DNA identifications (NRC II). Consequently, DNA analysis is considered the “gold standard” in forensic science (Saks and Koehler 2004). The meticulous requirements of the community may also lead to a similar benefit for forensic entomology. Accordingly, the wealth of information on, and recommended uses of, gene expression technologies, including quantitative PCR (e.g. Mocellin et al. 2003, Bustin et al. 2005, and Puskas et al. 2006) and microarrays (e.g. Puskas et al. 2006), can be used to develop a set of guidelines in forensic entomology. In this manner, input from the scientific community, which contributed to the optimal use of DNA evidence, can be used to guide the development of molecular protocols in forensic entomological predictions of blow fly age.

The statistical requirements of Daubert have also been more fully addressed in this dissertation. Using GAMs, it was possible to graphically demonstrate expected error through fitted vs. response plots as well as statistically demonstrate expected error with the PDE score. The use of GCV scores can also help investigators determine the most likely model to use/trust for a PMI estimate. Finally, the validation of GAM predictions

of blow fly age with traditional forensic entomological data and gene expression data has ensured that the technique may, one day, be used in court as its ability to reliably predict age has been characterized in blind studies. These statistical results are particularly important, since research of this type is rare in forensic entomology.

### **Logistical Improvements Associated with Molecular Predictions of Blow Fly Age.**

Gene expression analyses provide some logistical improvements that make PMI predictions based on such data appealing. First, the molecular methods studied herein can be conducted by more scientists, as there are a greater number of trained molecular biologists than trained forensic entomologists. Crime laboratories generally seek the advice of the few outside entomologists in the country, while the molecular methods used in the research presented here can be easily applied in most molecular biology laboratories and should expand the accessibility of forensic entomology to more cases by enabling any laboratories equipped to do DNA research to analyze blow fly samples. Though it is unlikely that such analyses will be employed within crime laboratories, as they do not encounter insect evidence frequently enough to justify in-house analyses, the number of potential laboratories that can provide investigators with data for predicting blow fly age will increase if molecular aging techniques are used.

Second, molecular biology research is amenable both to the use of robotics and miniaturization on microarrays, which allow thousands of genes to be analyzed simultaneously. These features should enable more rapid casework and help to decrease the costs of applying gene expression analyses for estimating a PMI. During the research presented here, the production of a gene expression profile cost approximately \$15 per individual larva or pupa. With roughly 50% of samples providing full expression

profiles, the analysis of 100 individuals would cost approximately \$1500 and yield 50 full profiles, which is sufficient to provide a statistically valid sample size. Given the large expenses involved a homicide case, a \$1500 price tag is not prohibitive and with miniaturization the cost will likely decrease while increasing the number of loci assessed (Dahl et al. 2007).

Finally, the use of GAMs and gene expression data can help investigators deal with the types of evidence they receive. Forensic scientists must often work with non-ideal evidence. Sample sizes can be small, requiring a limited analysis so evidence can be reevaluated at a later date. Any technique that accommodates this need will be preferred by investigators. Also, some entomological samples may be damaged in storage, with resultant weight, length, or stage determinations impossible to assess. Through the use of gene expression analyses, samples do not need to be completely used, allowing for future analyses. Likewise, if certain data (e.g. weight measurements) are unavailable, a GAM can be developed that allows for a statistical understanding of the model specific to the available data. This will enable investigators and triers of fact to evaluate how much weight can be assigned to a PMI prediction based on entomological evidence.

#### **Future Research in Forensic Entomology.**

A solid foundation has been established for future studies in forensic entomology. One of the most obvious directions to proceed is to profile gene expression levels in the many other forensically informative fly species (e.g. Hall 1948, Kamal 1958, Catts and Haskell 1990, Anderson 2000, Greenberg and Kunich 2002). Comparisons of gene regulation among cyclorrhaphan flies indicate that gene expression is likely to be similar



among species (e.g. determination of the embryonic anterior axis by *bcd*; McGregor 2005), but continuity in gene expression profiles across species is not guaranteed (e.g. differential timing in the expression of proneural genes between the *Calliphora* and *Phormia* genera; Skaer 2002). Each species will require profiling, just as development tables are needed for each (Catts and Haskell 1990). As gene expression research in blow flies develops, a series of loci should be settled on for all species. The identification of individuals with DNA evidence is based on standard loci (Chakraborty et al. 1999), which enables a detailed understanding of the error resulting from these well-studied short tandem repeats. This example should be followed in forensic entomology to achieve the same reliability and standardization that has been reached with DNA identifications. Accordingly, the best loci for estimating blow fly age should vary throughout the development of most or all of the common forensically informative blow fly species and the number of genes necessary for assessing blow fly age should be determined by the precision that can be gained by adding loci to an age estimate.

Additionally, the 9 genes studied herein are not necessarily the best for predicting a PMI. The genes studied in this research were chosen from a short list of available *Lucilia* sequences, based on *a priori* information indicating that their expression levels would change during development, thus enabling improved PMI estimates. However, there are thousands of genes in cyclorrhaphan genomes (Yeates and Wiegmann 2005) and, as demonstrated by genomic work in *Drosophila* (Arbeitman et al. 2002, Beckstead et al. 2005), thousands of unique gene expression profiles. Clearly, many more loci could be assessed to determine an optimal set for predicting fly age. The most important reason for adding genes to the PMI prediction process can be seen in the persistence of error in

pupal estimates (Chapter 5), wherein the PDE scores for gene expression based estimates was 78%, which, while a vast improvement over the 16% score associated with estimates that did not use gene expression, can be improved by the addition of more loci that are regulated during pupation. Candidate genes may be found by searching *Drosophila* and other dipteran gene expression data in the literature. Genes that exhibit changes in regulation at any point during development are all *a priori* candidates for research. The *Amalgam* and CG17814 loci studied by Arbeitman et al. (2002) are two examples of expression changes in genes that may increase precision in pupal age estimates. As more genes are included, arrays of gene sequences (such as the ones used in Arbeitman et al. 2002) could be used to determine age with hundreds or thousands of gene expression levels, which will enable more precise estimates of age.

There were also temperature effects on gene expression that indicate transcript level analysis may be a useful means of identifying blow fly development in different environments. Doing so could potentially enable investigators in determining molecular markers for conditions (like temperature differences) that contribute to plasticity in blow fly development, which will help them avoid errors in entomologically based PMI estimates. The *ace* gene demonstrated this point best, as it was expressed at a high level at low temperatures, while high temperatures caused down-regulation of the gene during the third instar. Consequently, *ace* expression levels (and *hsp90*, which behaved similarly) should help confirm or refute that evidentiary flies developed at the temperature assumed by investigators. Any gene expression changes that are indicators of environmental effects that could influence developmental plasticity (e.g. starvation, disease, diapause) have the potential to be useful in identifying when flies were exposed

to those conditions, allowing for more accurate PMI estimates. The research on “Peter Pan” larvae shows that arrested development in third instars can be identified by expression differences in the *hsp90*, *sll*, and *usp* genes. Identification of misleading individuals will help to prevent errors associated with their inclusion in a PMI estimate. In fact, all of the states of the third instar (feeding, postfeeding, non-pupator) are distinguishable by assessing gene expression differences, which is useful for improving PMI estimates with third instar larvae as it can be difficult to distinguish among them with physical evidence alone (Anderson 2000, Chapter 5, Chapter 7). However, to avoid circular logic in the PMI prediction framework, any gene expression level employed in the identification of a developmental stage should not be included in subsequent development percent predictions. Stage exerts the largest influence on predictions (Chapter 3), so a gene used to assign stage would have a disproportionately large influence on a PMI estimate if it were also included in the within-stage prediction. Accordingly, an expression profile used to recognize third instar states should be independent of the set of genes predicting age, necessitating research to determine the best loci for third instar identifications versus loci most useful for development percent predictions.

Further work to improve the statistical understanding of PMI estimates made with gene expression data is also necessary. There are other statistical options for gene expression analyses besides GAMs that are also capable of predicting age with multivariate, non-linear data, which may be valid techniques or even exhibit superior performances in predicting blow fly age. Cluster analyses (Eisen et al. 1998) and classification and regression trees (Brieman et al. 1984) are common statistical tools that

allow for a hierarchical organization of complex data that can indicate very specific subsets within a data set by splitting it into iteratively smaller subsets. The two *Drosophila* studies on genome-wide expression that were used as theoretical support for this experiment relied on cluster analyses (Arbeitman et al. 2002, Beckstead et al. 2005), indicating that these statistics are useful for assessing gene expression data. Neural networks can also make predictions with complex data (Venables and Ripley 2002, Ripley et al. 2004). These statistics calculate probability in a manner that is similar to biological neural functions, by passing input data through nodes, which then approximate output data. The ability of both methods to make predictions with complex data sets indicates that neural networks and cluster analyses may be improvements over the GAM statistic used herein. However, descriptions of neural networks are likely to be overly complex for a jury to understand and the statistical support (such as model comparisons) for cluster analyses and neural nets are not as straightforward as they are in GAMs as they rely on thousands of randomizations of the data. It should also be noted that both approaches are capable of making predictions that are too specific to the data set used to design them and must be extensively trained against new data to avoid that problem (Brieman et al. 1984, Ripley 2004). Additionally, Bayesian analyses may be useful for developing population-specific predictions of minimum development times (Scheiner and Gurevitch 2001). These methods merge previously recorded data, or “priors”, with data that are specific to the population of interest to make posterior predictions. In the context of forensic entomology, posterior predictions may be superior estimators of PMI, as published data for development times and gene expression levels could be combined with a small data set from a population of interest (such as the strain-specific development

times in Chapter 8), providing more accurate estimates of age. All of these statistical methods remain unexplored by the forensic entomology community and warrant further consideration.

Finally, the genetic and environmental variation in body size and development time among *L. sericata* strains should be more extensively characterized. Though the effects of temperature and strain on these traits were statistically significant, there were only three population comparisons made. This low level of replication makes it difficult to interpret the roles that local adaptation due to ecological differences played in the evolution of these traits among the strains studied. The next step in understanding the major underlying factors that influence body size and development time evolution in *L. sericata* and other blow flies is to study more populations and potential ecological influences that could be driving the observed differences among traits. Any ecological components that differ among the environments where a blow fly species is found could be the subjects of future studies, though efforts should be directed at identifying factors that explain most variation in development traits. The goal in determining ecological factors that affect blow fly development is to articulate a unifying theory for how the traits evolve among populations that could be used to predict development times/body sizes of unknown populations. An improved theoretical understanding of these traits could enable the use of correction factors and/or Bayesian analyses to achieve more accurate and population-specific predictions of age.

### *Controlling Blow Fly Populations with Gene Expression and Plasticity Data*

The results of this dissertation point to several means of improving the control of blow fly pests. The gene expression analyses conducted herein (Chapter 5) identified expression level differences among stages of *L. sericata* development in four loci that encode insecticide targets (*cs*, *ace*, *ecr*, *usp*), as well as variation in the expression of a gene that can decrease the toxicity of organophosphates (*rop-1*). Such information can be used to determine stages that are likely to be more susceptible to specific insecticides, which may allow for the most efficient use of chemicals to kill specific blow fly developmental stages.

Additionally, the research provided ample information regarding environmental effects on development time (Chapters 2, 7, 8), body size (Chapters 3, 8), and gene expression levels (Chapter 5) throughout the development of *L. sericata*. Low temperatures were shown to increase *ace* expression in larvae (Chapter 5), indicating that temperature differences could potentially affect organophosphate resistance. Likewise, body size and/or development times varied in response to food moisture (Chapter 2), fresh pupation substrate (Chapter 2), and temperature (Chapters 3, 8). Finally, arrested development in *L. sericata* postfeeding third instars was reliably associated with differences in the expression levels of *hsp90*, *sll*, and *usp* (Chapter 7), which may provide clues as to how to environmentally or chemically induce a delay in development. Any means of interrupting blow fly development or decreasing the size of blow fly pests (thereby preventing or decreasing the damage caused during myiasis) may be useful in developing management plans to ameliorate the damages caused by blow fly pests.

## *Conclusions*

Several conclusions can be drawn from the research presented here. First, gene expression analyses improve the accuracy and precision of PMI estimates, especially in postfeeding third instars and pupae, which currently provide the least precise age estimates with traditional approaches. Transcript abundances were useful in distinguishing amongst feeding, postfeeding, and “Peter Pan” third instar larvae, helping to avoid errors derived from misidentified third instar states. More importantly, the developmental profiles produced enabled significantly superior statistical descriptions of development, allowing for a mathematically defined understanding of error associated with age estimates.

Second, blow fly size, development time, and gene expression levels were shown through ANOVA and GAM to be affected by both genotype and the environment. Rearing conditions influenced body sizes and development times. The CA, MI, and WV strains developed at different rates and had different average sizes. Finally, gene expression levels significantly varied according to genetic strain and/or temperature treatments.

Third, studying blow fly developmental variation can contribute to the understanding of the evolution of age and size at metamorphosis. Such knowledge will help to provide a framework for developing standard operating procedures for the rearing of flies in the laboratory, addressing a key requirement of the Daubert standard for scientific evidence. Additionally, information on strain-specific development can be used to achieve more accurate PMI estimates by tailoring them to local populations.

Fourth, GAMs are useful for predicting blow fly age. They were used herein for estimating age with gene expression data, as well as with more traditional developmental stage and body size measures. Likewise, with this statistical approach, it was possible to demonstrate that, though significant, the effects of strain and temperature on *L. sericata* body size may be ignored with minimal effects on the precision of development percent predictions. The ability to introduce and remove variables from these models, and understand the effects on resultant predictions, will also be of value to investigators both in allowing them to work with all available data and enabling them to more fully address the statistical requirements of the Daubert standards for scientific evidence.

Last, the study of *L. sericata* development can improve scientific understanding in multiple fields of biological research. The data presented have provided preliminary information on potential influences of the environment on the evolution of *L. sericata* body size and development time traits. The molecular data provided a picture of the effects of strain and temperature associated with gene expression profiles that may aid future research in comparative genomics and information on the distributions associated with profiles of developmentally regulated genes. In addition, data on plasticity and the expression of specific genes can contribute to future research on the control of insect pests. Finally, and most importantly, the accuracy and precision of PMI estimates made with blow fly evidence was significantly improved, as was the ability to address the Daubert standards for scientific evidence.



## APPENDIX

**Appendix 2.1. Treatments and duration of the immature lifecycle and individual stages from all cohorts of *Lucilia sericata*.**

Cohort	Meat	Destructive	Transfer	Substrate Under	Substrate	Paper Towel	Egg
1	NFMD	No	No	No	Vermiculite	No	21.5
2	NFMD	Yes	No	No	Vermiculite	No	22.5
3	FMD	Yes	No	No	Vermiculite	No	20
4	FMD	No	No	No	Vermiculite	No	24
5	FMD	No	No	Yes	Vermiculite	No	21
6	FMD	No	Yes	No	Vermiculite	No	24
7	NFMD	No	No	No	Vermiculite	No	19
8	NFMD	Yes	No	No	Vermiculite	No	24
9	FMD	Yes	No	No	Vermiculite	No	22.5
10	FMD	No	No	No	Vermiculite	No	25
11	FMD	No	Yes	No	Vermiculite	No	25
12	FMD	Yes	No	No	Vermiculite	No	19.5

Cohort	Meat	Destructive	Transfer	Substrate Under	Substrate	Paper Towel	Egg
13	FMD	No	No	No	Vermiculite	No	18.5
14	FMD	No	Yes	No	Vermiculite	No	18.5
15	FMD	No	No	Yes	Vermiculite	No	21
16	FMD	No	No	No	Sand	No	15
17	FMD	No	Yes	No	Sand	No	15
18	FMD	No	No	No	Sand	Yes	18.5
19	FMD	No	Yes	No	Sand	Yes	18.5
20	FMD	No	Yes	No	Sand	Yes	24
21	FMD	No	No	No	Sand	Yes	24
22	FMD	No	No	Yes	Sand	No	19
23	FMD	No	No	Yes	Sand	No	22
24	FMD	No	No	No	Sand	No	24
25	FMD	No	Yes	No	Sand	No	24
26	FMD	No	No	No	Vermiculite	Yes	22

Cohort	Meat	Destructive	Transfer	Substrate Under	Substrate	Paper Towel	Egg
27	FMD	No	Yes	No	Vermiculite	Yes	22
28	FMD	No	Yes	No	Vermiculite	No	21.75
29	FMD	No	Yes	No	Sand	No	23
30	FMD	No	Yes	No	Sand	Yes	21.5
31	FMD	No	No	No	Vermiculite	No	21.75
32	FMD	No	No	No	Sand	No	23
33	FMD	No	No	No	Sand	Yes	21.5
34	FMD	No	Yes	No	Vermiculite	Yes	24.5
35	FMD	No	Yes	No	Vermiculite	Yes	24.5
36	FMD	No	No	No	Vermiculite	Yes	24.5
37	FMD	No	No	No	Vermiculite	Yes	24.5
38	Rat	No	Yes	No	Sand	No	20.5
39	Rat	No	Yes	No	Sand	No	20
40	Rat	No	Yes	No	Sand	No	18.5

<u>1st Instar</u>	<u>2nd Instar</u>	<u>3rd Instar (Feeding)</u>	<u>Postfeeding</u>	<u>Pupa</u>	<u>Total Hours</u>	<u>Total Days</u>
26	47	69.5	100.5	162.5	427	17.79
24.5	25.25	48	98.25	224	442.5	18.44
30	21.75	47.25	94.5	284	497.5	20.73
27	28	45	98	162.25	384.25	16.01
29.5	22.75	49	99.25	187.5	409	17.04
27	28	45	48	163.25	335.25	13.97
26	47.5	47.5	119.25	161.5	420.75	17.53
24.5	26	70	76	285	505.5	21.06
26.5	23.75	52	91.5	189	405.25	16.89
25.25	23.75	51	118	187	430	17.92
26.25	23.75	51	94	192	412	17.17
27.75	46.75	47	100.75	234	475.75	19.82

<u>1st Instar</u>	<u>2nd Instar</u>	<u>3rd Instar (Feeding)</u>	<u>Postfeeding</u>	<u>Pupa</u>	<u>Total Hours</u>	<u>Total Days</u>
28	23.5	52.25	95.25	140	357.5	14.9
28	23.5	52.25	43.75	163	329	13.71
25.5	22.5	53	93	160	375	15.63
30.75	23.75	48	95.5	163	376	15.67
30.75	23.75	48	71.75	187	376.25	15.68
29.5	25.5	25	69	164	331.5	13.81
29.5	25.5	25	48	185	331.5	13.81
26.75	28	21.5	72.25	162.5	335	13.96
26.75	28	21.5	72.25	162.5	335	13.96
27.75	24.75	47.75	93.75	188	401	16.71
28	25	47	72.25	162.75	357	14.88
27.25	24	48.25	65	168.5	357	14.88
27.25	24	48.25	65	168.5	357	14.88
24	29.5	23.5	71.75	162.25	333	13.88

<u>1st Instar</u>	<u>2nd Instar</u>	<u>3rd Instar (Feeding)</u>	<u>Postfeeding</u>	<u>Pupa</u>	<u>Total Hours</u>	<u>Total Days</u>
24	29.5	23.5	71.75	162.25	333	13.88
24	28.25	50.25	47	163.25	334.5	13.94
23.5	28.25	50.75	46.75	167.75	340	14.17
24.25	29.75	25.75	70.5	188.5	360.25	15.01
24	28.25	50.25	65.25	166.5	356	14.83
23.5	28.25	50.75	94.25	164.25	384	16
24.25	29.75	25.75	70.5	212.5	384.25	16.01
25	22	23.5	50.25	185.75	331	13.79
25	22	24	49.5	186	331	13.79
25	22	23.5	95.5	168	358.5	14.94
25	22	24	75.5	167.5	338.5	14.10
29	22.5	24	68.5	172.5	337	14.04
31	25.5	24	67	166	333.5	13.9
32	22.5	25	46.5	188.5	333	13.88

The minimum development times of each stage and the minimum total development time for cohorts of *Lucilia sericata*. Also listed are the combinations of variables that each cohort experienced. All times are reported in hours (to the closest quarter hour) except the total development time, which is reported in hours and days. Minimum development times ranged from 329 to 505.5 hours for liver-fed cohorts (Cohorts 1–37) and from 333 to 337 hours for rat-fed cohorts (Cohorts 38–40). Labels are as in Table 2.1.



## LITERATURE CITED

**Ali RA, Mellenthin K, Fahmy K, Da Rocha S, and Baumgartner S. 2005.** Structural conservation of the salivary gland-specific slalom gene in the blowfly *Lucilia sericata*. *Dev Genes Evol.* 215:537-43.

**Ames C and Turner B. 2003.** Low temperature episodes in development of blow flies: implications for postmortem interval estimation. *Med Vet Entomol.* 17:178–86.

**Anderson GS. 2000.** Minimum and maximum development rates of some forensically important Calliphoridae (Diptera). *J Forensic Sci.* 45:824–32.

**Arbeitman MN, Furlong EEM, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, and White KP. 2002.** Gene expression during the lifecycle of *Drosophila melanogaster*. *Science.* 297:2270–5.

**Atkinson D and Sibly RM. 1996.** On the solutions to a major life-history puzzle. *Oikos.* 77:359–65.

**Beckstead RB, Lam G, and Thummel CS. 2005.** The genomic response to 20-hydroxyecdysone at the onset of *Drosophila* metamorphosis. *Genome Biology.* R99.

**Berrigan D and Charnov EL. 1994.** Reaction norms for age and size at maturity in response to temperature: a puzzle for life historians. *Oikos.* 70:474–8.

**Bialecki M, Shilton A, Fichtenberg C, Segraves WA, and Thummel CS. 2002.** Loss of the ecdysteroid-inducible E75A orphan nuclear receptor uncouples molting from metamorphosis in *Drosophila*. *Dev Cell.* 3:209–20.

**Blanckenhorn W. 2002.** The consistency of quantitative genetic estimates in field and laboratory in the yellow dung fly. *Genetica.* 114:171–82.

**Brieman L, Friedman J, Stone CJ, and Olshen RA. 1984.** Classification and regression trees. CRC Press LLC: Boca Raton, FL.

**Bustin SA, Benes V, Nolan T, and Pfaffl MW. 2005.** Quantitative real-time RT-PCR--a perspective. *J Mol Endocrinol.* 34:597–601.

**Byrd JH and Allen JC. 2001.** The development of the black blow fly, *Phormia regina* (Meigen). *Forensic Sci Int.* 120:79–88.

**Catts EP and Haskell NH. 1990.** *Entomology and Death: A Procedural Guide.* Joyce's Print Shop, Inc.: Clemson, SC.

**Chakraborty R, Stivers DN, Su B, Zhong Y, and Budowle B. 1999.** The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. *Electrophoresis.* 20:1682–96.

**Chaudhury MF, Welch JB, and Alvarez AL. 2002.** Responses of fertile and sterile screwworm (Diptera: Calliphoridae) flies to bovine blood inoculated with bacteria originating from screwworm-infested animal wounds. *J Med Entomol.* 39:130–4.

**Chen Z, Newcomb R, Forbes E, McKenzie J, and Batterham P. 2001.** The acetylcholinesterase gene and organophosphorus resistance in the Australian sheep blowfly, *Lucilia cuprina*. *Insect Biochem Mol Biol.* 31:805–16.

**Clark K, Evans L and Wall R. 2006.** Growth rates of the blowfly, *Lucilia sericata*, on different body tissues. *For Sci Int.* 156:145–9.

**Collins FS, Patrinos A, Jordan E, Chakravarti A, and Gesteland R. 1998.** New Goals for the U.S. Human Genome Project: 1998-2003. *Science.* 5389:682–9.

**Conner JK and Hartl DL. 2004.** *A primer of ecological genetics.* Sinauer Associates, Inc.: Sunderland, Massachusetts.

**Cook PE, Hugo LE, Iturbe-Ormaetxe I, Williams CR, Chenoweth SF, Ritchie SA, Ryan PA, Kay BH, Blows MW, and O'Neill SL. 2006.** The use of transcriptional profiles to predict adult mosquito age under field conditions. *PNAS.* 103:18060-5.

**Crystal MM and Ramirez R. 1975.** Screwworm flies for sterile-male release: laboratory tests of the quality of candidate strains. *J Med Ent.* 12:418-422.

**Dahl A, Sultan M, Jung A, Schwartz R, Lange M, Steinwand M, Livak KJ, Lehrach H, and Nyarsik L. 2007.** Quantitative PCR based expression analysis on a nanoliter scale using polymer nano-well chips. *Biomed Microdevices*. [Epub ahead of print].

**Day T and Rowe L. 2002.** Developmental Thresholds and the Evolution of Reaction Norms for Age and Size at Life-History Transitions. *Am Nat*. 159:338–50.

**East IJ and Eisemann CH. 1993.** Vaccination against *Lucilia cuprina*: the causative agent of sheep blowfly strike. *Immunol Cell Biol*. 71:453–62.

**Eisen MB, Spellman PT, Brown PO, and Botstein D. 1998.** Cluster analysis and display of genome-wide expression patterns. *PNAS*. 95:14863–8.

**Faulde M, Sobe D, Burghardt H, and Wermter R. 2001.** Hospital infestation by the cluster fly, *Pollenia rudis* sensu stricto Fabricius 1794 (Diptera: Calliphoridae), and its possible role in transmission of bacterial pathogens in Germany. *Int J Hyg Environ Health*. 203: 201-204.

**Feder JL, Berlocher SH, Roethele JB, Dambroski H, Smith JJ, Perry WL, Gavrilovic V, Filchak KE, Rull J, and Aluja M. 2003.** Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. *PNAS*. 100:10314–9.

**Flatt T, Tu MP, and Tatar M. 2005.** Hormonal pleiotropy and the juvenile hormone regulation of *Drosophila* development and life history. *Bioessays*. 27:999–1010.

**Gibson G and Weir B. 2005.** The quantitative genetics of transcription. *Trends Genet*. 21:616–23.

**Goff ML. 2007.** Insects and time since death: What do we really estimate? American Academy of Forensic Sciences Annual Meeting. San Antonio, Texas.

**Gomez-Mestre I and Buchholz DR. 2006.** Developmental plasticity mirrors differences among taxa in spadefoot toads linking plasticity and diversity. *PNAS* 103:19021–6.

- Goodbrod JR and Goff ML. 1990.** Effects of larval population density on rates of development and interactions between two species of *Chrysomya* (Diptera: Calliphoridae) in laboratory culture. J Med Entomol. 27:339–43.
- Gorham JR. 1987.** Insect and mite pests in food: an illustrated key. US Department of Agriculture, Agriculture Handbook Number 655.
- Grassberger M. and Reiter C. 2001.** Effect of temperature on *Lucilia sericata* (Diptera: Calliphoridae) development with special reference to the isomegalen- and isomorphen-diagram. For Sci Int. 120:32–6.
- Greenberg B. 1991.** Flies as forensic indicators. J Med Entomol. 20:565–77.
- Greenberg B and Kunich JC. 2002.** Entomology and the Law Flies as Forensic Indicators. Cambridge University Press: Cambridge, UK.
- Hallas R, Schiffer M, and Hoffmann AA. 2002.** Clinal variation in *Drosophila serrata* for stress resistance and body size. Genet Res. 79: 141–8.
- Hall DG. 1948.** The blow flies of North America. Monumental Printing Co.: Baltimore, MD.
- Hall MJ, Hutchinson RA, Farkas R, Adams ZJ, Wyatt NP. 2003.** A comparison of Lucitraps and sticky targets for sampling the blowfly *Lucilia sericata*. Med Vet Entomol. 17:280–87.
- Hargrove WW and Hoffman FM. 2004.** Potential of multivariate quantitative methods for delineation and visualization of ecoregions. Environ Manage. 34:S39–60.
- Hastie TJ and Tibshirani RJ. 1990.** Generalized additive models. Chapman and Hall/CRC.: Boca Raton, FL
- Henrich VC and Brown NE. 1995.** Insect nuclear receptors: a developmental and comparative perspective. Insect Biochem Molec Biol. 25:881–97.

**Hoffmann AA and Harshman LG. 1999.** Desiccation and starvation resistance in *Drosophila*: patterns of variation at the species, population and intrapopulation levels. *Heredity*. 83:637–43.

**Johnson FM and Schaffer HE. 1973.** Isozyme variability in species of the genus *Drosophila*. VII. Genotype-environment relationships in populations of *D. melanogaster* from the Eastern United States. *Biochem Genet*. 10:149–63.

**Kalthoff K.** Analysis of biological development. 2001. McGraw-Hill: New York, NY.

**Kamal AS. 1958.** Comparative study of thirteen species of sarcosaprophagous Calliphoridae and Sarcophagidae (Diptera) I. Bionomics. *Ann Entomol Soc Am*. 51:261–70.

**Kaneshraiah G and B. Turner B. 2004.** *Calliphora vicina* larvae grow at different rates on different body tissues. *Int J Legal Med*. 118:242–4.

**Khaitovich P, Enard W, Lachmann M, and Paabo S. 2006.** Evolution of primate gene expression. *Nat Rev Genet*. 7:693–702.

**Kingsolver JG. 2000.** Feeding, Growth, and the Thermal Environment of Cabbage White Caterpillars, *Pieris rapae* L. *Physiol Biochem Zool*. 73:621–8.

**Kingsolver JG, Gomulkiewicz R, and Carter PA. 2001.** Variation, selection and evolution of function-valued traits. *Genetica*. 112–113: 87–104.

**Kulathinal RJ and Hartl DL. 2005.** The latest buzz in comparative genomics. *Genome Biol*. 6:201.

**Liu Z, Gong P, Wu K, Sun J, and Li D. 2006.** A true summer diapause induced by high temperatures in the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae). *J Insect Physiol*. 52:1012–20.

**Luders F, Segawa H, Stein D, Selva EM, Perrimon N, Turco SJ, and Hacker U. 2003.** Slalom encodes an adenosine 3'-phosphate 5'-phosphosulfate transporter essential for development in *Drosophila*. EMBO J. 22:3635–44.

**Ludwig MZ, Patel NH, and Kreitman M. 1998.** Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. Development. 125:949–58.

**Mackay TFC. 2001.** The genetic architecture of quantitative traits. Annu Rev Genet. 35:303–39.

**Mansson RA, Frey JG, Essex JW, and Welsh AH. 2005.** Prediction of properties from simulations: a re-examination with modern statistical methods. J Chem Inf Model. 45:1791–803.

**McAlpine JF. 1989.** Manual of Nearctic Diptera. Canadian Government Publishing Center.

**McGregor AP. 2005.** How to get ahead: the origin, function, and evolution of *bicoid*. BioEssays. 27:904–13.

**Mellenthin K, Fahmy K, Ali RA, Hunding A, Da Rocha S, and Baumgartner S. 2006.** Wingless signaling in a large insect, the blowfly *Lucilia sericata*: a beautiful example of evolutionary developmental biology. Dev Dyn. Feb;235(2):347–60.

**Mocellin S, Rossi CR, Pilati P, Nitti D, and Marincola FM. 2003.** Quantitative real-time PCR: a powerful ally in cancer research. Trends Mol Med. 9:189–95.

**Moe SJ, Kristoffersen AB, Smith RH, and Stenseth NC. 2005.** From patterns to processes and back: analyzing density-dependent responses to an abiotic stressor by statistical and mechanistic modeling. Proc R Soc B. 272:2133–42.

**Montooth KL, Marden JH, and Clark AG. 2003.** Mapping determinants of variation in energy metabolism, respiration and flight in *Drosophila*. Genetics. 165:623–35.

**Morey S and Reznick D. 2000.** A comparative analysis of plasticity in larval

development in three species of spadefoot toads. *Ecology*. 81: 1736–49.

**Mount SM, Gotea V, Lin CF, Hernandez K, and Makalowski W. 2007.** Spliceosomal small nuclear RNA genes in 11 insect genomes. *RNA*. 13:5–14.

**National Center for Biotechnology website.** [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)

**Newcomb RD, Gleeson DM, Yong CG, Russell RJ and Oakeshott JG. 2005.** Multiple mutations and gene duplications conferring organophosphorus insecticide resistance have been selected at the Rop-1 locus of the sheep blowfly, *Lucilia cuprina*. *J Mol Evol*. 60:207–20.

**Oudman L, Van Delden W, Kamping A, and Bijlsma R. 1991.** Polymorphism at the *Adh* and alpha *Gpdh* loci in *Drosophila melanogaster*: effects of rearing temperature on developmental rate, body weight, and some biochemical parameters. *Heredity*. 67:103–15.

**Parsch J, Russell JA, Beerman I, Hartl DL and Stephan W. 2000.** Deletion of a conserved regulatory element in the *Drosophila Adh* gene leads to increased alcohol dehydrogenase activity but also delays development. *Genetics*. 156:219–27.

**Perrin N. 1995.** About Berrigan and Charnov's life-history puzzle. *Oikos*. 73:137–9.

**Plaistow SJ, Lapsley CT, Beckerman AP and Benton TG. 2004.** Age and size at maturity: sex, environmental variability and developmental thresholds. *Proc Biol Sci*. 271: 919–24.

**Powell JR. 1997.** Progress and prospects in evolutionary biology. Oxford University Press: New York, NY.

**Puskas LG, Menesi D, Feher LZ, and Kitajka K. 2006.** High-throughput functional genomic methods to analyze the effects of dietary lipids. *Curr Pharm Biotechnol*. 7:525–9.

**R Development Core Team. 2004.** R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL <http://www.R-project.org>.

**Ripley RM, Harris AL, and Tarassenko L. 2004.** Non-linear survival analysis using neural networks. *Stat Med*. 23:825–42.

**Rogina B and Helfand SL. 1995.** Regulation of gene expression is linked to life span in adult *Drosophila*. *Genetics*. 141:1043–8.

**Rogina B, Benzer S, and Helfand SL. 1997.** *Drosophila* drop-dead mutations accelerate the time course of age-related markers. *PNAS*. 94:6303–6.

**Saks MJ and Koehler JJ. 2005.** The coming paradigm shift in forensic identification science. *Science*. 309:892–5.

**Scheiner SM and Gurevitch J. 2001.** Design and analysis of ecological experiments. 2<sup>nd</sup> Ed. Oxford University Press. New York, NY.

**Sherman RA and My-Tien Tran JM. 1995.** A simple, sterile food source for rearing the larvae of *Lucilia sericata* (Diptera: Calliphoridae). *Med Vet Entomol*. 9:393–8.

**Sherman RA. 2000.** Wound myiasis in urban and suburban United States. *Arch Intern Med*. 160: 2004–14.

**Shingleton AW. 2005.** Body-size regulation: combining genetics and physiology. *Curr Biol*. 15:R825–7.

**Shingleton AW, Das J, Vinicius L, and Stern DL. 2005.** The temporal requirements for insulin signaling during development in *Drosophila*. *PLoS Biol*. 3:e289.

**Sibly RM and Atkinson D. 1994.** How rearing temperature affects optimal adult size in ectotherms. *Func Ecol*. 8:486–93.



**Skaer N, Pistillo D, and Simpson P. 2002.** Transcriptional heterochrony of *scute* and changes in bristle pattern between two closely related species of blowfly. *Dev Bio.* 252:31–45.

**Sliter TJ and Gilbert LI. 1992.** Developmental arrest and ecdysteroid deficiency resulting from mutations at the *dre4* locus of *Drosophila*. *Genetics.* 130:555–68.

**Sullivan AA and Thummel CS. 2003.** Temporal profiles of nuclear receptor gene expression reveal coordinate transcriptional responses during *Drosophila* development. *Mol Endocrinol.* 17:2125–37.

**Tachibana S, Numata H, and Goto SG. 2005.** Gene expression of heat-shock proteins (Hsp23, Hsp70 and Hsp90) during and after larval diapause in the blow fly *Lucilia sericata*. *J Insect Physiol.* 51:641–7.

**Tachibana S and Numata H. 2004.** Effects of temperature and photoperiod on the termination of larval diapause in *Lucilia sericata* (Diptera: Calliphoridae). *Zoolog Sci.* 21:197–202.

**Tarone AM and Foran DR. 2006.** Components of plasticity in the development of a Michigan population of *Lucilia sericata*. *J Med Ent.* 43:1023–33.

**Tarone AM, Nasser YM, and Nuzhdin SV. 2005.** Genetic variation for expression of the sex determination pathway genes in *Drosophila melanogaster*. *Genet Res.* 86:31–40.

**Tellam RL, Vuocolo T, Johnson SE, Jarmey J, and Pearson RD. 2000.** Insect chitin synthase cDNA sequence, gene organization and expression. *Eur J Biochem.* 267:6025–43.

**Varki A and Altheide TK. 2005.** Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* 15:1746–58.

**Venables WN and Ripley BD. 2002.** Modern Applied Statistics with S (4th edn). Springer: New York, NY.

**Wall R, Pitts KM and Smith KE. 2001.** Pre-adult mortality in the blowfly *Lucilia sericata*. Med Vet Entomol. 15:328–334.

**Ware WG and Whitacre DM. 2004.** An Introduction to Insecticides (4th edition) Extracted from The Pesticide Book, 6th ed. MeisterPro Information Resources: Willoughby, Ohio.

**Wells JD and Kurahashi H. 1994.** *Chrysomya megacephala* (Fabricius) (Diptera: Calliphoridae) development: rate, variation, and the implications for forensic entomology. Jpn J Sanit Zool. 45:303–9.

**Wells JD and Lamotte LR. 1995.** Estimating maggot age from weight using inverse prediction. J Forensic Sci. 40:585–90.

**Wood SN. 2006.** Generalized additive models: an introduction in R. Chapman and Hall/CRC Taylor and Francis Group: Boca Raton, FL.

**Yeates DK and Wiegmann BM. 2005.** The evolutionary biology of flies. Columbia University Press: New York, NY.