Aging Blow Fly Eggs Using Gene Expression: A Feasibility Study

ABSTRACT: Forensic entomology can aid death investigations by using predictable developmental changes to estimate the ages of flies associated with a body. In developmental stages that do not increase in size however, including the egg and pupa, it can be difficult to objectively refine an age estimate beyond the limits of the stage duration. Gene transcript levels, changing throughout development, represent a potential data source useful for objectively identify smaller units of developmental time. The expression of three genes (bcd, sll, cs) was profiled throughout the maturation of blow fly eggs to determine the feasibility of predicting age, identifying significant linear trends in expression during their development. Models estimating egg age made predictions within 2 h of true age when all expression data were available, while the presence/absence of cs transcripts identified two age classes, together indicating that gene expression can be used to more precisely predict blow fly age.

KEYWORDS: forensic science, quantitative polymerase chain reaction, development, Calliphoridae, postmortem interval, forensic entomology, Lucilia sericata, bicoid, slalom, chitin synthase

Insects found on human remains can be useful in estimating a postmortem interval (PMI) during death investigations (1). 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., 2–5). 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 (6), and references therein]. Flies have been particularly well studied in this regard (7–13), 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 (7), 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 (10), including the Drosophilidae and Calliphoridae. sll affects dorsal-ventral patterning (12), and is also highly expressed in the salivary glands of D. melanogaster and L. sericata larvae (13). cs was profiled as chitin formation is required for the proper assembly of the larval cuticle (14). 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
Lucilia sericata was collected in East Lansing, Michigan, and was identified visually and genetically as previously described (15). A fly cage at room temperature was presented with beef liver and examined every 15 min until females were observed laying eggs, which was allowed to continue for 1 h. Egg masses (comprised of c. 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 1-h collection span. The first eggs hatched between 9 and 10 h, thus the 8–9 h time span was the oldest analyzed.
Gene Sequencing and Primer Design

Expression levels of bed, 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 bed, sll, and rp49 on the National Center for Biotechnology Information website (16), 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 (16). PCR consisted of 35 cycles of denaturing at 95°C for 30 sec, annealing primers at 55°C for 30 sec, and extending amplification 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 300 μL of lysis solution without use of a prefiltration plate. RNA was eluted from plates with 100 μL elution solution and incubated at 37°C for 1 h with 70 units of DNase-I buffer (Applied Biosystems). The enzyme was inactivated at 75°C for 10 min and the RNA was precipitated using 110 μL of isopropanol followed by centrifugation at maximum speed for 30 min at 4°C. Two 70% ethanol washes followed, using the same centrifuge settings. RNA samples were allowed to air dry for 15 min, at which point 32 μL 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 30 μL of RNA in a final volume of 120 μL. Gene expression levels were assessed by quantitative PCR on an Applied Biosystems 7900HT using SYBRgreen PCR mastermix in 15 μL reactions on a 384-well plate. Each reaction received 1.5 μL cDNA, 7.5 μL SYBRgreen PCR mastermix, and 1 μL 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 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 (17). 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 (18,19), which produce a statistic, percent deviance explained (similar to R2), 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

Lucilia sericata sequences for β tubulin 56D and cs are listed under the National Center for Biotechnology Information accession number.
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, R² = 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 < 2 h old never producing the transcript, while those 6 h and older always expressed the gene, therefore cs expression state could be plotted during egg development (Fig. 1). The presence of the transcript was a statistically significant predictor of age (p < 0.0001, R² = 0.59).

Each of the genes had a different quantitative expression pattern (Figs. 2–4; note that the displayed CT values are inversely related to gene expression levels). Though only expressed after 2 h, cs transcripts significantly increased during egg development (p = 0.0004, R² = 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, R² = 0.19 and p = 0.0023, R² = 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 (Fig. 5), estimated ages followed the True = Predicted line, with 30 of 33 predictions within 2 h of the true age.

**Discussion**

The goal of this research was to examine the feasibility of using gene expression to more precisely age immature flies of forensic interest, consequently generating more accurate estimates of PMI. The loci examined demonstrated statistically significant, though noisy, trends in expression levels throughout egg development. Additionally, egg masses <2 h old did not express cs, while egg
masses older than 6 h always expressed the gene. Following on
efforts to predict adult mosquito age using gene expression and
multiple regression (20), generalized additive models were used to
predict egg ages. When CT scores were available for all loci, 91% of
predictions 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 develop-
mental 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 expression commencing in the developing salivary
glands (13). Finally, cs, which is required only for production of the
larval cuticle (14), followed a different and predicted transcriptional course, wherein the earliest portion of egg develop-
ment 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 consider-
ation of any new forensic protocol. The methods are amenable to
microarrays (7) and robotics, potentially producing simplified and
high throughput blow fly aging analyses. Finally, DNA-based meth-
odologies 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. Natu-
really, 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.

Acknowledgments
This project was supported by Grant Number 2004-DN-BXK005, awarded by the National Institute of Justice, Office of
Justice Programs, U.S. Department of Justice. Thanks to Jeff
Tomberlin of Texas A&M and Ryan Kimbrauskas of Michigan State
University for confirming visual species identifications. The advice of Jeff Wells from West Virginia University was
greatly appreciated, as was the assistance of Annette Thelen in
the Michigan State University Research Technology Support
Facility. This body of work was a portion of the Ph.D. disserta-
tion research of AT.

Disclaimer
Points of view in this document are those of the authors and
do not necessarily represent the official position or policies of
the U.S. Department of Justice.
References


Additional information and reprint requests:
David R. Foran, Ph.D.
School of Criminal Justice and Department of Zoology
560 Baker Hall
Michigan State University
East Lansing, MI 48824
E-mail: foran@msu.edu