

The use of transcriptional profiles to predict adult mosquito age under field conditions

Peter E. Cook*, Leon E. Hugo*[†], Iñaki Iturbe-Ormaetxe*, Craig R. Williams*[‡], Stephen F. Chenoweth*, Scott A. Ritchie*^{†¶}, Peter A. Ryan[†], Brian H. Kay[†], Mark W. Blows*, and Scott L. O'Neill*[¶]

*School of Integrative Biology, University of Queensland, Brisbane, Queensland 4072, Australia; [†]Queensland Institute of Medical Research, Herston, Queensland 4029, Australia; [‡]Anton Breinl Centre for Public Health and Tropical Medicine, James Cook University, Cairns, Queensland 4870, Australia; and [¶]Tropical Population Health Unit, Queensland Health, Cairns, Queensland 4870, Australia

Edited by Barry J. Beaty, Colorado State University, Fort Collins, CO, and approved October 5, 2006 (received for review June 10, 2006)

Age is a critical determinant of an adult female mosquito's ability to transmit a range of human pathogens. Despite its central importance, relatively few methods exist with which to accurately determine chronological age of field-caught mosquitoes. This fact is a major constraint on our ability to fully understand the relative importance of vector longevity to disease transmission in different ecological contexts. It also limits our ability to evaluate novel disease control strategies that specifically target mosquito longevity. We report the development of a transcriptional profiling approach to determine age of adult female *Aedes aegypti* under field conditions. We demonstrate that this approach surpasses current cuticular hydrocarbon methods for both accuracy of predicted age as well as the upper limits at which age can be reliably predicted. The method is based on genes that display age-dependent expression in a range of dipteran insects and, as such, is likely to be broadly applicable to other disease vectors.

Aedes aegypti | age-grading | gene expression | cuticular hydrocarbons | multivariate calibration

Mosquito-transmitted diseases, such as malaria and dengue, continue to be major causes of human morbidity and mortality throughout the world. The pathogens that cause these diseases require a period to replicate and disseminate in the mosquito's body before transmission to naïve hosts can occur. The period between when a mosquito blood feeds on an infected host and when it is able to transmit the infection is termed the extrinsic incubation period (EIP). The EIP is estimated to be between 10 and 14 days for dengue virus (1, 2) and 9 and 14 days for malaria (3). Not all female mosquitoes start feeding immediately after they eclose, therefore, a vector must live longer than the duration of the initial nonfeeding period (usually <2 days) plus the EIP of the pathogen to contribute to disease transmission. This combined period (>12 days) is quite long relative to adult mosquito lifespan and, consequently, mosquito population age structure is a critical determinant of a vector population's capacity for pathogen transmission (4).

Despite the central importance of insect population age structure to the epidemiology of these diseases, relatively few methods have been developed that can directly and accurately measure insect age in the field (5–9). Existing methods for age-grading insects, such as scoring modifications to ovarian anatomy, are laborious and discriminate only very young adult age classes. Pteridine concentration has been shown to predict age in various dipteran species (10–12), and although its initial application to predicting mosquito age showed promise (13), it has proven to be generally unreliable for wild mosquitoes because pteridines occur in limited quantities and their concentration fluctuates with blood feeding (9). In the case of the primary dengue vector, *Aedes aegypti*, the most recently developed age-grading method utilizes changes in proportions of cuticular hydrocarbons (CHC) from legs to predict adult age (6, 14). This method has been shown to be capable of estimating adult *Ae. aegypti* age up to 15 days in field evaluations (6). Given that many individuals live >15 days in the field and that dengue

transmission primarily occurs in mosquitoes 12 days and older, this approach has limited utility in epidemiological investigations. Considering that a number of research groups are proposing new control strategies for these diseases that explicitly target mosquito lifespan (15–17), there is a growing need for tools to reliably measure mosquito age under field conditions.

Recent gene expression studies in *Drosophila* (18–21), anopheline mosquitoes (22), nematodes (23), mice (24), and monkeys (25) have demonstrated consistent age-dependent changes in global gene transcription profiles. *Drosophila melanogaster* microarray studies have shown that expression of 2–9% of the genome changes significantly with age (19–21). Results from similar studies in *Anopheles gambiae* estimate that transcription of ≈5% of the genome varies with age (22). These changes in transcription have the potential to be used as markers for determining insect age in the field. Using *D. melanogaster* transcriptional profiles (18), we identified a set of genes showing significant changes in expression across adult female life stages. Orthologues of these genes were isolated from *Ae. aegypti* and a quantitative reverse transcriptase PCR (qRT-PCR) assay developed to determine chronological age of female mosquitoes under field conditions. Comparison of this approach to the existing CHC method indicates that it provides greater accuracy and precision over a broader range of adult age classes.

Results and Discussion

Transcriptional profiling studies in *D. melanogaster* have demonstrated that expression levels of a number of genes change significantly with adult age (18, 20, 21). On examination of these data sets, we selected a set of genes for analysis in mosquitoes based on the following criteria: (i) displayed large variation in gene expression across adult age classes, (ii) expression was unlikely to be influenced by digestive or reproductive status, and (iii) orthologues were known to exist in *Aedes* mosquitoes. Once a suitable set of candidate genes was identified, PCR primers for *Ae. aegypti* orthologues of these genes were designed for qRT-PCR assays (Table 1).

Initial evaluations of the suitability of these genes for age grading was made on laboratory-reared female mosquitoes. Adult female mosquitoes aged up to 29 days were used to determine whether transcription of the selected orthologues was consistent with that seen for *D. melanogaster* (18). Transcrip-

Author contributions: P.E.C., L.E.H., I.I.-O., S.A.R., and S.L.O. designed research; P.E.C., L.E.H., C.R.W., and S.A.R. performed research; P.E.C., L.E.H., S.F.C., M.W.B., and S.L.O. analyzed data; and P.E.C., L.E.H., I.I.-O., S.F.C., S.A.R., P.A.R., B.H.K., M.W.B., and S.L.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: CHC, cuticular hydrocarbons; Ct, cycle threshold; GE, gene expression; qRT-PCR, quantitative reverse transcriptase PCR.

[¶]Present address: Sansom Institute, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, SA 5000, Australia.

[¶]To whom correspondence should be addressed. E-mail: scott.oneill@uq.edu.au.

© 2006 by The National Academy of Sciences of the USA

validate reference genes to ensure that they are not influenced by experimental treatment. We designed primers for several potential reference genes from *Ae. aegypti*; *actin-1* (*Aaeact-1*, U20287), *β -tubulin* (TIGR GI TC66155) and *ribosomal protein S17* (*Ae-RpS17*; AY927787) and quantified transcript abundance across adult age in laboratory-reared mosquitoes. PCR products were amplified from *Ae. aegypti* cDNA, gel purified and cloned into pGEM-Teasy (Promega), and confirmed by sequencing. qRT-PCR was performed on a Lightcycler (Roche) by using Platinum SYBR Green I Supermix (Invitrogen) and the following cycling conditions: 50°C for 2 min; 95°C for 2 min, then 95°C for 5 s; 60°C for 5 s; 72°C for 10 s; for 50 cycles; with fluorescence acquisition at the end of the extension step. Product specificity was confirmed by melting curve analysis. Linearized plasmids containing inserts of each reference gene were amplified in triplicate and used to construct standard curves with 10⁵-fold dynamic range. Absolute quantification of reference gene transcripts were made from their respective standard curve by using Lightcycler software (version 3.5; Roche).

Determining Transcript Abundance with qRT-PCR. Transcript abundance of the eight candidate genes and *Ae-RpS17* was determined for individual mosquitoes. qRT-PCRs were performed on the Lightcycler (Roche) by using Platinum SYBR Green I Supermix (Invitrogen) according to the manufacturer's instructions for glass capillary thermal cyclers. Cycling conditions for all templates were as reported above for the reference gene, and PCR product specificity was determined by melting curve analysis. Triplicate reactions were run for each primer set (Table 1) across all templates. The Ct was determined for each reaction by using the second-derivative maximum method (Roche). Mean Ct values were calculated, and relative changes in transcript abundances between individuals were normalized to *Ae-RpS17* by log contrasts (35).

Mosquito Cuticular Hydrocarbon Extraction and Quantification. CHCs were extracted from all legs of individual field-reared mosquitoes and quantified by GC-MS based on methods described in Hugo *et al.* (36). Peak areas of four CHC measures of interest ($C_{25}H_{52}$, $C_{26}H_{54}$, $C_{27}H_{56}$, and $C_{28}H_{58}$) and a reference measure ($C_{29}H_{60}$) were quantified. Relative changes in peak area were normalized to $C_{29}H_{60}$ by calculating log contrasts.

Multivariate Calibration and Indirect Regression Methodologies. A multivariate calibration framework was developed to estimate mosquito age from GE and CHC measures. A training data set was established, for both GE and CHC measures, by quantifying each phenotypic measure from mosquitoes of known age. Relative changes in GE and CHC measures from each individual were calculated as log contrasts (35). This approach was used to remove the unit sum constraint associated with compositional data. Log contrasts were calculated as follows:

$$\log \text{ contrast } X_i = \log_{10} \left(\frac{(X_i/X_{\text{total}})}{(X_{\text{ref}}/X_{\text{total}})} \right),$$

where X_i equals the mean Ct value for a particular gene (or area of a CHC peak), X_{ref} equals the mean Ct values of the reference gene (or peak area for reference CHC), and X_{total} equals the sum of Ct values for all genes from an individual (or total of all CHC peak areas).

Because some CHC measures were absent for certain individuals, log contrasts were calculated with the addition of 1 to eliminate missing values from the CHC data set. Training data were entered into a canonical redundancy analysis to reduce the dimensionality of the regressions between multiple GE or CHC measures and age. Linear regression of the first redundancy variate on age was used for each calibration model. Age estimates then were derived from this calibration model by inverse regression of new GE or CHC measures.

Additional samples were processed and used to evaluate the reliability of both GE and CHC calibration models. A nonparametric bootstrapping method was used to assess sampling error in the calibration model procedure (see *Supporting Text* and Table 2). Data sets were standardized ($N \approx 0,1$) before undertaking the bootstrapping procedure. Field cage samples were bootstrapped by resampling individuals within cages and age classes with replacement, thereby generating 1,000 pseudorandom data sets. A canonical redundancy analysis was performed on each data set to calculate a single redundancy variate that was a linear combination of phenotypic values that maximized the correlation with age. Each individual within each bootstrap replicate was scored for its value of the redundancy variate. The relationship between the redundancy variate and age was predicted by using simple linear regression, producing 1,000 regression equations. These models then were used to score mosquitoes that had been sampled from field cages as effective blind samples for age estimation. Phenotypes of these samples were also standardized ($N \approx 0,1$) before analysis. Each individual was scored for 1,000 bootstrap redundancy variates and ages predicted according to the 1,000 regression equations from the corresponding redundancy variates. The top and bottom 2.5% of predicted age values represent the 95% confidence intervals for age. For all numerical procedures, we used SAS (version 9.1; SAS Institute, Cary, NC); see *Supporting Text* for SAS editor syntax.

We thank the staff at Cairns Tropical Population Health Unit, especially Sharron Long for her assistance with field work; Geoff Eaglesham and Neil Holling from Queensland Health Scientific Services for technical assistance with gas chromatography mass spectroscopy analysis; and Dr. Jeremy C. Brownlie for his constructive comments on drafts of the manuscript. This research was funded by Australian Research Council Grant LP0455732 and grants from the Foundation for the National Institutes of Health through the Grand Challenges in Global Health Initiative and the Queensland Government's Smart State Ph.D. funding initiative.

- Siler JF, Hall MW, Hitchens AP (1926) *Philipp J Sci* 29:1–304.
- Watts DM, Burke DS, Harrison BA, Whitmore RE, Nisalak A (1987) *Am J Trop Med Hyg* 36:143–152.
- Gilles HM, Warrell DA (2002) *Essential Malariaology* (Arnold, London).
- Dye C (1992) *Annu Rev Entomol* 37:1–19.
- Detinova TS (1962) in *WHO Monograph No. 47* (World Health Org, Geneva).
- Gerade BB, Lee SH, Scott TW, Edman JD, Harrington LC, Kitthawee S, Jones JW, Clark JM (2004) *J Med Entomol* 41:231–238.
- Hayes EJ, Wall R (1999) *Physiol Entomol* 24:1–10.
- Tyndale-Biscoe M (1984) *Bull Entomol Res* 74:341–377.
- Penilla RP, Rodriguez MH, Lopez AD, Viader-Salvado JM, Sanchez CN (2002) *Med Vet Entomol* 16:225–234.
- Hayes EJ, Wall R (1999) *Physiol Entomol* 24:1–10.
- Lehane MJ (1985) *Parasitol Today* 1:81–85.
- Robson SK, Vickers M, Blows MW, Crozier RH (2006) *J Exp Biol* 209:3155–3163.
- Wu D, Lehane MJ (1999) *Med Vet Entomol* 13:48–52.
- Desena ML, Edman JD, Clark JM, Symington SB, Scott TW (1999) *J Med Entomol* 36:824–830.
- Blanford S, Chan BH, Jenkins N, Sim D, Turner RJ, Read AF, Thomas MB (2005) *Science* 308:1638–1641.
- Brownstein JS, Hett E, O'Neill SL (2003) *J Invertebr Pathol* 84:24–29.
- Rasgon JL, Styer LM, Scott TW (2003) *J Med Entomol* 40:125–132.
- Arbeitman MN, Furlong EEM, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP (2002) *Science* 297:2270–2275.
- Landis GN, Abdueva D, Skvortsov D, Yang J, Rabin BE, Carrick J, Tavare S, Tower J (2004) *Proc Natl Acad Sci USA* 101:7663–7668.
- Pletcher SD, Macdonald SJ, Marguerie R, Certa U, Stearns SC, Goldstein DB, Partridge L (2002) *Curr Biol* 12:712–723.
- Zou S, Meadows S, Sharp L, Jan LY, Jan YN (2000) *Proc Natl Acad Sci USA* 97:13726–13731.
- Marinotti O, Calvo E, Nguyen QK, Dissanayake S, Ribeiro JM, James AA (2006) *Insect Mol Biol* 15:1–12.

23. Lund J, Tedesco P, Duke K, Wang J, Kim SK, Johnson TE (2002) *Curr Biol* 12:1566–1573.
24. Lee CK, Klopp RG, Weindruch R, Prolla TA (1999) *Science* 285:1390–1393.
25. Kayo T, Allison DB, Weindruch R, Prolla TA (2001) *Proc Natl Acad Sci USA* 98:5093–5098.
26. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JMC, Wides R, et al. (2002) *Science* 298:129–149.
27. Sanders HR, Evans AM, Ross LS, Gill SS (2003) *Insect Biochem Mol Biol* 33:1105–1122.
28. Sundberg R (1999) *Scand J Stat* 26:161–191.
29. Rencher AC (2001) *Methods of Multivariate Analysis* (Wiley, New York).
30. Sinkins SP, Gould F (2006) *Nat Rev Genet* 7:427–435.
31. Nasci RS (1986) *J Am Mosq Control Assoc* 2:61–62.
32. Russell RC, Webb CE, Williams CR, Ritchie SA (2005) *Med Vet Entomol* 19:451–457.
33. Thellin O, Zorzi W, Lakaye B, De Borman B, Coumans B, Hennen G, Grisar T, Igout A, Heinen E (1999) *J Biotechnol* 75:291–295.
34. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) *Genome Biol* 3:research0034.1–0034.11.
35. Aitchison J (2003) *The Statistical Analysis of Compositional Data* (Blackburn, Caldwell, NJ).
36. Hugo LE, Kay BH, Eaglesham GK, Holling N, Ryan PA (2006) *Am J Trop Med Hyg* 74:462–474.