Instar Determination of *Blaptica dubia* (Blattodea: Blaberidae) Using Gaussian Mixture Models

HAO WU,[1] ARTHUR G. APPEL, AND XING PING HU

Department of Entomology and Plant Pathology, Auburn University, 301 Funchess Hall, Auburn, AL 36849–5413

Ann. Entomol. Soc. Am. 106(3): 323–328 (2013); DOI: http://dx.doi.org/10.1603/AN12131

ABSTRACT Instar determination is fundamental to both basic entomological research and its application. The cockroach, *Blaptica dubia* Serville (Blattodea: Blaberidae), is a popular pet and an excellent feeder insect for many reptiles and amphibians. A new method using Gaussian mixture models to determine the number of instars in this species is developed. Application of the method is illustrated by analysis of data collected on *B. dubia*. The analysis indicates that there are seven instars in *B. dubia* and that the growth ratio follows the Brooks–Dyar rule. The growth ratio of pronotal length, pronotal width, and head width are 1.26, 1.24, and 1.19, respectively. Because *B. dubia* shares a similar growth pattern with other paurometabolous insects, this method may be applicable to other species as well.

KEY WORDS instar, Brooks–Dyar rule, model based clustering

Instar determination is fundamental to both basic entomological research and application (Logan et al. 1998). Determining instar distributions from a given population is required for life table analysis, key factor analysis, and other important ecological investigations. The number of instars is useful in the development of phenology models, or in the refinement of existing models (Godin et al. 2002). In paleobiology, understanding the growth, evolution, and ecology of extinct species often depends on calibrating ontogeny with respect to either absolute time or relative developmental age (Hunt and Chapman 2001). In forensic studies, instar determination helps to estimate the minimum postmortem interval (Velásquez and Viloria 2010).

The number of instars varies widely across insect species. Instar number is frequently considered to be invariable within species, although intraspecific variability in the number of instars is not an exceptional phenomenon (Esperk et al. 2007). In many species of Arthropoda, heavily sclerotized structures, such as the head capsule, remain approximately the same size during a stadium (Daly 1985). Measurable changes in size occur after molts. These sclerotized characters, particularly head capsule, have been used to differentiate instars.

For instar determination, direct observation may be most accurate, but in many cases, observations are not possible or it may take too much time to obtain the results (In our observations, it takes ≈6 mo at 30 ± 2°C for this species). There is a long history of entomologists quantifying changes in size of particular structures for a wide range of insects. It has long been supposed that the ratio of measurements of well-sclerotized structures in any instar to the equivalent measurement in the preceding instar tends to be constant throughout the life history (Hutchinson and Tongring 1984). Many different techniques have been proposed for estimating either the growth stage (instar) or the age of insects (Daly 1985). The original observations on arthropod sizes and developmental stages were made by W. K. Brooks on stomatopod larvae collected by the H.M.S. Challenger (Brooks 1886). In Brooks's observation, the length of an instar is 1.25 times the length of its predecessor. Dyar later independently came to the same conclusion of equivalent size ratios from observations on the caterpillars of 26 species of Lepidoptera in 1890. Dyar's measurements showed that the width of the head capsule increased by a ratio, ranging from 1.3 to 1.7, that was constant for a given species (Dyar 1890):

$$\frac{postmolt\ size}{premolt\ size} = constant$$

However, the sizes of different instars may overlap, making instar determination more difficult. Morphological measurements can be viewed as normal distributions in each instar. Thus, samples from a given population of insects are a mixture of several normal distributions, one for each instar. The probability density function of a mixture of G normal distributions can be written as:

$$f_{Mix}(x) = \sum_{k=1}^{G} p_k f_k(x)$$

[1] Corresponding author, e-mail: hzw0019@auburn.edu.

0013-8746/13/0323–0328$04.00/0 © 2013 Entomological Society of America

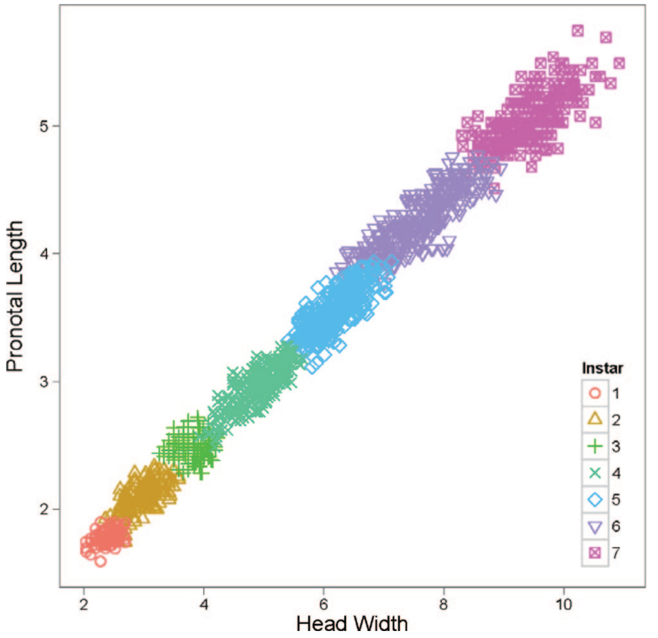

Fig. 1. The seven instars of pronotal length (mm) and head width (mm) of *B.dubia*.

where p_k is the mixing proportion, $p_k \geq 0$, $\Sigma_{k=1}^{G} p_k = 1$, $f_k(x)$ is the p-dimensional normal density function $N_p(\mu_k, \Sigma_k)$, with mean vector μ_k, and covariance matrix Σ_k. This assumption is reasonable because measurements of individuals from the same ontogenetic stage are often distributed normally (Hunt and Chapman 2001).

Distribution based methods have been adopted by many entomologists for a variety of species (Logan et al. 1998, Hunt and Chapman 2001, Hammack et al. 2003, Gullan and Cranston 2005, Delbac et al. 2010). Almost all of these methods are based on the widths of head capsules. There are some efforts to use different characters to determine instars (McClellan and Logan 1994, Velásquez and Viloria 2010), but the characters are used independently. Because different characters of insects are highly correlated, it is more

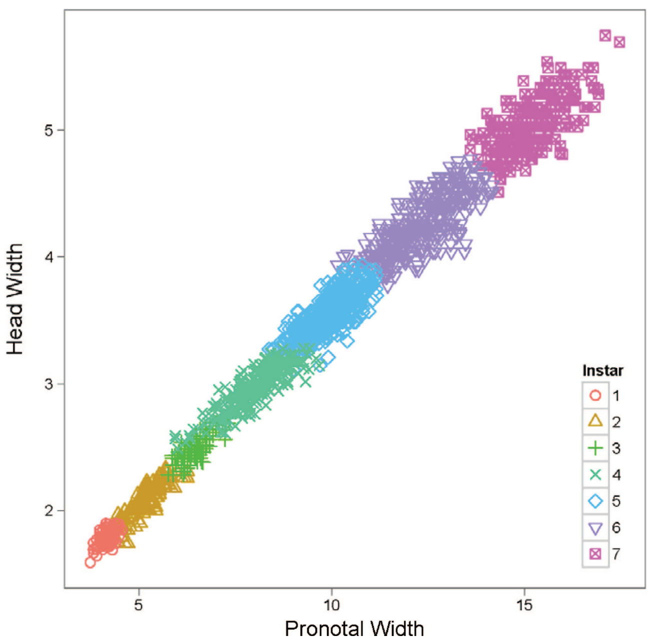

Fig. 2. The seven instars of pronotal width (mm) and head width (mm) of *B.dubia*.

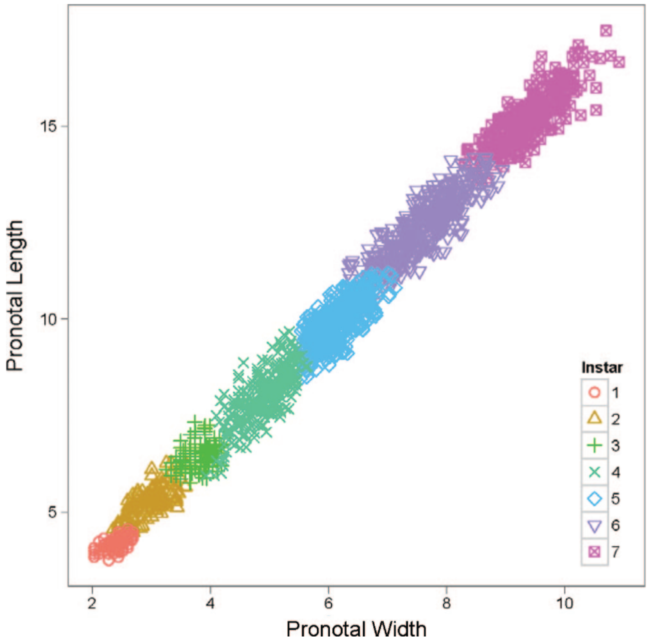

Fig. 3. The seven instars of pronotal width (mm) and pronotal length (mm) of *B.dubia*.

statistically appropriate to adopt a multivariate approach.

Blaptica dubia Serville (Blattodea: Blaberidae), (commonly known as the Dubia cockroach, Guyana spotted cockroach, or orange-spotted cockroach) is a large (up to ≈4.5 cm in length), sexually dimorphic blaberid cockroach that is often kept as a pet. More importantly, this species is used as a food source for many reptiles and amphibians. Compared with other food sources such as crickets, *B. dubia* contains a higher percentage of protein, is easier to maintain, and produces little odor. Consequently, they have become an increasingly popular food among amphibian and reptile enthusiasts. In *B. dubia*, the head capsule width, pronotal width, and body length are the most convenient characters to measure in all nymphal instars. These three characters are used together in the analysis to determine instars. Because the three characters can be seen as a multivariate normal distribution in each instar, they form a Gaussian mixture when combined (Hastie et al. 2003). As a result, a new method using Gaussian mixture models to determine the number of instars in this species is developed. The results are assessed by the Brooks–Dyar rule and verified by direct observations on the development of *B. dubia* in the laboratory and corroborated with results of a previous study (Hintze-Podufal and Nierling 1986).

Materials and Methods

Insect Rearing. *B. dubia* were reared in clear gallon glass jars (3.8 liters, 15.24 cm wide by 25.4 cm high) with cardboard harborage at 30 ± 2°C. They were exposed to a photoperiod of 12:12 (L:D) h and supplied with water and dry dog food (Purina Dog Chow, Ralston Purina, St. Louis, MO) ad libitum.

Measurements. In total 1,925 nymphs of *B. dubia* were measured from January–April 2012. Head capsule width (the widest distance between the two compound eyes) was measured to 0.01 mm using a microscope (Leica MZ 6, Solms, Germany) with an ocular micrometer. The length and width of the pronotum were measured to 0.01 mm using an electronic digital caliper (model: 62379-531, Control Company, Friendswood, TX).

Statistical Model. In model-based clustering, it is assumed that the data are generated by a mixture of underlying probability distributions f_k (x),in which each component k represents a different group or cluster (Johnson and Wichern 2007). If there are G clusters, the observation of variables is modeled as arising from the mixture distribution:

$$f_{Mix}(x) = \sum_{k=1}^{G} p_k f_k(x)$$

where p_k is the mixing proportion, $p_k \geq 0$, $\Sigma_{k=1}^{G} p_k = 1$, $f_k(x)$ is the p-dimensional normal density function $N_p(\mu_k, \Sigma_k)$ with mean vector μ_k, and covariance matrix Σ_k. The Gaussian mixture model for one observation with G components is:

$$f_{Mix}(x|\mu_1, \Sigma_1, \ldots, \mu_G, \Sigma_G)$$

$$= \sum_{k=1}^{G} p_k \frac{\exp\left\{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right\}}{(2\pi)^{\frac{p}{2}}|\Sigma_k|^{\frac{1}{2}}}$$

Table 1. The means and 95% simultaneous T^2-intervals of three characters in each instar (mm) of *B. dubia*

	Instar 1	Instar 2	Instar 3	Instar 4	Instar 5	Instar 6	Instar 7	Ratio
N	127	178	146	280	394	408	285	
Pronotal length	2.41 (2.37–2.45)	2.97 (2.91–3.03)	3.82 (3.77–3.87)	4.91 (4.84–4.98)	6.17 (6.11–6.23)	7.54 (7.46–7.63)	9.39 (9.30–9.47)	1.26
Pronotal width	4.18 (4.14–4.22)	5.21 (5.13–5.30)	6.47 (6.40–6.54)	8.01 (7.89–8.13)	9.95 (9.86–10.04)	12.28 (12.15–12.41)	15.17 (15.04–15.30)	1.24
Head width	1.79 (1.78–1.80)	2.08 (2.06–2.11)	2.49 (2.47–2.51)	2.97 (2.94–3.00)	3.55 (3.53–3.58)	4.24 (4.21–4.28)	5.04 (5.00–5.07)	1.19

Inference is based on the likelihood, which for N objects and a fixed number of clusters G is (Johnson and Wichern 2007):

$$L(p_1,\ldots,p_k;\ \mu_1, \Sigma_1, \ldots, \mu_G, \Sigma_G) = \prod_{j=1}^{N} f_{mix}(x_j|\mu_1, \Sigma_1, \ldots, \mu_G, \Sigma_G)$$

The Bayesian Information Criterion (BIC),

$$BIC = 2ln(L_{max}) - 2ln(N)\left(G\frac{1}{2}(p+1)(p+2) - 1\right)$$

where $L_{\max}$ is the maximum of likelihood function, N is the number of observations, G is number of components, and p is number of variables. It has been used to determine the number of clusters (Fraley and Raftery 1998). In instar determination, the BIC gives a quantitative method to determine the number of instars for a given insect.

Fraley and Raftery (2012) developed the (R Development Core Team 2012) package "mclust" for model-based clustering based on a parameterized Gaussian mixture model (Fraley and Raftery 1998). The software package "mclust" was used in the data analysis.

Results and Discussion

Model Selection. After comparing the BIC of every cluster (sub-model) and all the parameterized models, the unconstrained model with seven clusters was selected because it has the largest BIC value, −6,232.165 (range, −25,282.88 ≈ −6,232.165). Therefore, there are seven instars in *B. dubia* when reared as described above.

Cluster Analysis. Gaussian mixture model based clustering was conducted based on the model selection using the BIC. Figures 1, 2 and 3 show the cluster results of three characters. From the scatter plots, there are seven instars in *B. dubia* and the characters of each instar cluster together. The means and 95% simultaneous T^2-intervals (similar to confidence intervals, but for multivariate data) of the three characters in each instar are shown in Table 1. The pronotal length is in the range of 2.41–9.37 mm; the pronotal width is in the range of 4.18–15.17 mm; and the range of head width is 1.79–5.04 mm; for first and seventh instars, respectively (Table 1). Based on the result in the cluster analysis, predictions can be made on new observations, using the function provided in the mclust package.

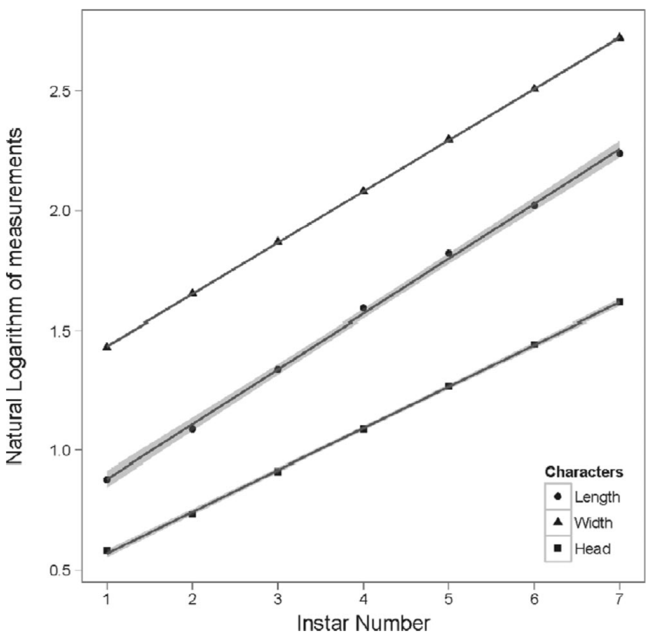

Fig. 4. The linear regression of three characters with 95% confidence interval region in each instar. Pronotal length (Length), Pronotal width (Width), Head width (Head).

Table 2. Linear regression models for three characters in the seven instars of *B. dubia*

Character	Slope ± SE	Ratio[a]	Intercept ± SE	R^2	F (df = 1,5)	P (> t)
Pronotal length	0.2278 ± 0.0039	1.26	0.6573 ± 0.0176	0.9988	4,082	<0.0001
Pronotal width	0.2136 ± 0.0011	1.24	1.2235 ± 0.0047	0.9999	108,300	<0.0001
Head width	0.1734 ± 0.0014	1.19	0.3979 ± 0.0064	0.9996	12,990	<0.0001

[a] Ratio is e^{slope}.

Result Assessment. According to the Brooks–Dyar rule (Dyar 1890):

$$\frac{I_n}{I_{n-1}} = constant$$

where I_n = *postmolt size* and I_{n-1} = *premolt size*, this equation can also be written as:

$$y = ae^{bx}$$

which is equivalent to:

$$lny = lna + bx$$

,

a linear function. We can, therefore, assess the cluster results using linear regression models. The constant growth ratio is e^b. From Fig. 4 and Table 2 the linear relationships of the three characters agree with the Brooks–Dyar rule. In our analysis, the growth ratios of pronotal length, pronotal width, and head width are 1.26, 1.24, and 1.19, respectively (Table 1). The growth rates of pronotal length and width are different and the shape of the pronotum changes slightly during development. According to Cole (1980), the median growth ratio for hemimetabolous insects is 1.27, and the mean is 1.27 ± 0.011 (n = 50, α = 0.05). For holometabolous insects the median is 1.52, and the mean is 1.55 ± 0.033 (n = 55, α = 0.05). This indicates that all instars are represented in the samples and that the model-based approach gives reasonable results.

In another study, the development of *B. dubia* was monitored at the same conditions as stated above [30 ± 2°C, and a photoperiod of 12:12 (L:D) h]. We directly observed seven instars in *B. dubia* over a period of 6 mo. In addition, Hintze-Podufal and Nierling (1986) also indicated that there were seven instars in *B. dubia* reared at 28 ± 2°C.

Although the results of our analyses support the Brooks–Dyar rule, they do not rely on it. Model-based clustering can also be used for other species even though their growth may not follow the Brooks–Dyar rule (Bliss and Beard 1954, Klingenberg and Zimmermann 1992), as long as there were size differences between different instars.

All of the studies about instar determination that use statistical techniques are based on the normal distribution of the given characters in each instar (Logan et al. 1998, Hunt and Chapman 2001, Godin et al. 2002,Gullan and Cranston 2005, Delbac et al. 2010). Our approach uses the multinormal distribution of characters in each instar and this is the only assumption in our analysis. Distributions were examined visually using the bivariate plot of the characters.

Compared with the widely adopted method of McClellan and Logan (1994), model-based clustering provides a better approach to solve the instar determination problem. Instar determination using Gaussian mixture models successfully solved the problem of determining the number of instars using the BIC. In the McClellan and Logan (1994) method, the number of instars is based on direct observations and the initial value of the mean and variance of each instar are based on this observation. This approach may produce some bias because different observers may have various initial values.

The clustering method based on Gaussian mixture models described above illustrates that it can be successfully used to determine instars in *B. dubia.* Because no characters in our analysis are specific to *B. dubia,* the same cluster methods should be applicable to other species as well.

Acknowledgments

We thank Peng Zeng, Department of Mathematics and Statistics, and Michael C. Wooten, Department of Biological Sciences, Auburn University for reviews of the manuscript. We also thank the Department of Entomology and Plant Pathology for partial financial support for this project.

References Cited

Bliss, C. I., and R. L. Beard. 1954. The growth of the head capsule in individual milkweed bugs. Ann. Entomol. Soc. Am. 47: 388–392.

Brooks, W. K. 1886. Report on the stomatopoda collected by H.M.S. Challenger during the years 1873–76. *In* the voyage of H.M.S. Challenger, report 45,vol. 16. http://www.19thcenturyscience.org/HMSC/HMSC-Reports/Zool-45/htm/doc.html.

Cole, B. J. 1980. Growth ratios in holometabolous and hemimetabolous Insects. Ann. Entomol. Soc. Am. 73: 489–491.

Daly, H. V. 1985. Insect morphometrics. Annu. Rev. Entomol. 30: 415–438.

Delbac, L., P. Lecharpentier, and D. Thiery. 2010. Larval instars determination for the European grapevine moth (Lepidoptera: Tortricidae) based on the frequency distribution of head-capsule widths. Crop Protect. 29: 623–630.

Dyar, H. G. 1890. The number of molts of lepidopterous larvae. Psyche 5: 420–422.

Esperk, T., T. Tammaru, and S. Nylin. 2007. Intraspecific variability in number of larval instars in insects. J. Econ. Entomol. 100: 627–645.

Fraley, C., and A. E. Raftery. 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput. J. 41: 578–588.

Fraley, C., and A. E. Raftery. 2012. MCLUST version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation. Technical report no. 597, Department of Statistics, University of Washington, Seattle, WA.

Godin, J., P. Maltais, and S. Gaudet. 2002. Head capsule width as an instar indicator for larvae of the cranberry fruitworm (Lepidoptera: Pyralidae) in southeastern New Brunswick. J. Econ. Entomol. 95: 1308–1313.

Gullan, P. J., and P. Cranston. 2005. The insects: an outline of entomology, 3rd ed. Wiley–Blackwell, Malden, MA.

Hammack, L., M. M. Ellsbury, R. L. Roehrdanz, and J. L. Pikul. 2003. Larval sampling and instar determination in field populations of northern and western corn rootworm (Coleoptera: Chrysomelidae). J. Econ. Entomol. 96: 1153–1159.

Hastie, T., R. Tibshirani, and J. H. Friedman. 2003. The elements of statistical learning. Springer, New York.

Hintze-Podufal, C, and U. Nierling. 1986. Der Einfluss der Nahrung auf Entwicklung, Wachstum und Prareproduktionsphase von *Blaptica dubia* Stal (Blaberoidea, Blaberidae). Mitteilungen der Schweizerischen Entomologischen Gesellschaft. 59: 177–186.

Hunt, G., and R. E. Chapman. 2001. Evaluating hypotheses of instar-grouping in arthropods: a maximum likelihood approach. Paleobiology 27: 466–484.

Hutchinson, G. E., and N. Tongring. 1984. The possible adaptive significance of the Brooks–Dyar rule. J. Theor. Biol. 106: 437–439.

Johnson, R. A., and D. W. Wichern. 2007. Applied multivariate statistical analysis, 6th ed. Prentice Hall, Upper Saddle River, NJ.

Klingenberg, C. P., and M. Zimmermann. 1992. Dyar's rule and multivariate allometric growth in nine species of waterstriders (Heteroptera: Gerridae). J. Zool. 227: 453–464.

Logan, J. A., B. J. Bentz, J. C. Vandygriff, and D. L. Turner. 1998. General program for determining instar distributions from head capsule widths: example analysis of mountain pine beetle (Coleoptera: Scolytidae) data. Environ. Entomol. 27: 555–563.

McClellan, Q. C., and J. Logan. 1994. Instar determination for the gypsy moth (Lepidoptera: Lymantriidae) based on the frequency distribution of head capsule widths. Environ. Entomol. 23: 248–253.

R Development Core Team. 2012. R: a language and environment for statistical computing. Vienna, Austria.

Velásquez, Y., and Á. L. Viloria. 2010. Instar determination of the neotropical beetle *Oxelytrum discicolle* (Coleoptera: Silphidae). J. Med. Entomol. 47: 723–726.

Received 26 October 2012; accepted 25 February 2013.