

Delimitation of a continuous morphological character with unknown prior membership: application of a finite mixture model to classify scapular setae of *Abacarus panticis*

Tsung-Jen Shen · Chi-Chien Kuo · Chin-Fah Wang · Kun-Wei Huang

Received: 7 June 2013 / Accepted: 28 February 2014 / Published online: 20 March 2014
© Springer International Publishing Switzerland 2014

Abstract Unambiguous classification is a prerequisite for the study of polymorphism, but accurate delimitation of continuous morphological characters can be challenging. Finite mixture modeling is a rigorous and flexible approach for delimiting continuous variables with unknown prior membership, but its application to morphological studies remains limited. In this study, the lengths of scapular setae of the eriophyoid mite *Abacarus panticis* Keifer collected from 18 sites in Taiwan were used as an example to evaluate the eligibility of finite mixture models. We then tested the hypothesis that longer scapular setae can facilitate dispersal. Lastly, we investigate morphological variation in various seta morphs by geometric morphometric techniques. Finite mixture models can satisfactorily classify scapular setae of *A. panticis* into long and short seta morphs. *Abacarus panticis* of the long morph only occurred in five sites whereas the short seta morph existed in all study sites. Geometric morphometric analyses revealed a more elongated coxal area in individuals of long morph than in those of short morph. Because the short morph is more widespread in geographical distribution than the long morph, longer scapular setae seem unlikely a specialized adaptation for dispersal. Further studies should capitalize on the finite mixture model in the delimitation of continuous morphological characters.

Tsung-Jen Shen and Chi-Chien Kuo have contributed equally in this study.

T.-J. Shen

Department of Applied Mathematics and Institute of Statistics, National Chung Hsing University, Taichung 402, Taiwan

C.-C. Kuo · K.-W. Huang (✉)

Department of Biology, National Museum of Natural Science, Taichung 404, Taiwan
e-mail: eri@mail.nmns.edu.tw

C.-C. Kuo

Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei 106, Taiwan

C.-F. Wang

General Education Center, National Chiayi University, Chiayi 600, Taiwan

Keywords Polymorphism · Eriophyoid mites · Intraspecific variability · Geometric morphometrics

Introduction

Polymorphism has long been a central issue in evolutionary biology studies (Ford 1945). The investigation of occurrence and mechanism underlying the maintenance of intra-specific polymorphism has helped reveal the adaptive significance of different morphs, thus the potential scenario of speciation (Skulason and Smith 1995; Roulin 2004). However, an important prerequisite for studying polymorphism is that morphs are correctly classified; otherwise, the association between morphs and fitness can become ambiguous or even misleading.

Categorization of continuous data is especially challenging because of the difficulties in determining character boundaries unequivocally (Reid and Sidwell 2002). Tixier (2012) recommended using mean and 95 % confidence interval to delimit seta lengths of Phytoseiidae mites and other similarly continuous morphological characters, but the analyzed data sets need to be normally distributed and not highly skewed. Besides, it is debatable that an arbitrary 5 % cut-off criterion is adequate for delimiting morphs for all organisms (Ezard et al. 2010). Alternatively, a visual inspection of character frequency distribution might reveal gaps and thus assist the classification of character states (Almeida and Bisby 1984), but such non-statistical method could lead to equivocal results when a clear-cut boundary does not exist. Recently, Ezard et al. (2010) developed a clustering method based on a combination of morphological traits using principal component analysis, but to our knowledge, a more sophisticated statistical approach for classifying a single continuous morphological character with unknown prior membership is still lacking.

Finite mixture modeling is one method that can objectively classify continuous variables without information on prior membership. The finite mixture model springs from Pearson (1894) about 100 years ago (McLachlan and Peel 2005). The flexibility of this method benefits by taking a lot of parameters into account, but at the same time it is computationally intensive. Thanks to the great progress in computer capability along with a recent development of efficient numerical algorithms, however, the technique has been well greeted and has been applied to many fields, including medical research (McLachlan and Chang 2004), botany (Zhang et al. 2001; Jaworski and Podlaski 2012), fishery (Macdonald and Pitcher 1979; Millar 1987), and avian ecology (Kéry et al. 2005). Compared with conventional cluster analysis, the primary strength of finite mixture model is its flexibility in fitting data sets with various candidate distributions without prior knowledge of the number of groups or categories. Such flexibility would be very useful for handling a variety of data when we typically have little information at the beginning of data analysis. Since it is a model-based method, its performance fully depends on the pattern and size of data (observations). Many criteria are available for assessing the feasibility of selected model, and using these quantitative measures can objectively determine a model that fits the working data well. The approach we applied here is firstly to find a probability model that can fit the surveyed dataset well and subsequently determine a threshold for delimiting boundaries of morphological characters. The model-based method offers some advantages when comparing with traditional approaches: for data of interest, it is able to depict the shape of each component; it gives estimates associated with means and standard deviation

for each component; its statistical fitness can be evaluated by several theoretical measures (Fonseca and Cardoso 2007). Notwithstanding these advantages, finite mixture model has rarely been applied to delimit continuous morphological characters.

In this study, we use eriophyoid mites (superfamily Eriophyoidea) as an example to demonstrate the strength and feasibility of finite mixture model. Eriophyoid mites, also known as gall, blister, erineum, bud, or rust mites, include more than 300 genera and about 3,700 species (De Lillo and Amrine 2003). They differ from other mites by having only two pairs of legs at anterior part and by their obligate phytophagous habits. The vermiform body is minute in size (80–250 μm), with most of the setae reduced (Amrine et al. 2003). When examining eriophyoid mites collected on Yushan cane *Yushania nitakayamensis* ((Hayata) Keng f.; Poaceae) (Wang and Huang, 2011), we observed that the length of scapular setae of an eriophyoid mite, *Abacarus panticis* Keifer, appears to be bimodally distributed (Fig. 1). Although eriophyoid mites can be passively dispersed through translocation of plant parts or they can actively disperse by walking out or by adhering to animals of better mobility, aerial dispersal, especially dispersal by wind, is considered to be the primary avenue (see reviews by Michalska et al. 2010 and Galvão et al. 2012). Krantz (1973) suggested that the scapular setae could facilitate wind dispersal; nevertheless, probably due to their minute size and their inclination to hide inside plants (Lindquist and Oldfield 1996), difficult direct observation makes such evaluation challenging. The likely differential scapular seta morphs of *A. panticis* thus offer a good opportunity to test whether longer scapular setae could facilitate their dispersal when combined with information on distribution of different seta morphs. The continuous variation in the length of setae renders a threshold in differentiating morphs not easily determined by conventional methods, but this can be rigorously solved by using finite mixture model.

The aim of this study is firstly to test the feasibility of a number of finite mixture models in fitting the frequency distribution of lengths of scapular setae of *A. panticis*. We then tested the hypothesis that longer scapular setae could assist in wind dispersal by comparing the geographic distribution of two identified morphs, that is, long versus short seta morphs. We expect long seta morph to be more widespread if long scapular setae do facilitate dispersal. Lastly, we evaluated whether different seta morphs also vary in shape and size by using geometric morphometric analyses to cast light on the potential mechanism underlying divergence in seta morphism. Relative to conventional morphometric methods that simply focus on a few basic linear measurements (such as body size, length of seta etc.), geometric morphometric techniques retain much wealthier information on shape and size that is valuable for further analyses (Adams et al. 2004; Zelditch et al. 2004). Although this study focuses on length of setae of an eriophyoid mite, finite mixture models we apply here should be applicable to delimit similarly continuous morphological characters of other taxa as well.

Materials and methods

Specimens and survey sites

Between 1993 and 2008, leaves of *Y. nitakayamensis* suspected with eriophyoid mites were collected in the mountains of Taiwan at 18 sites in Taiwan (Fig. 2). We follow Kuo et al. (2013) in the preparation of eriophyoid mites for measurement. Leaves were later preserved in 75 % ethanol with saturated sugar for at least 1 week. Eriophyoid mites were removed from leaves by spinning the preserved liquid in a centrifuge at 1,500 rpm for three

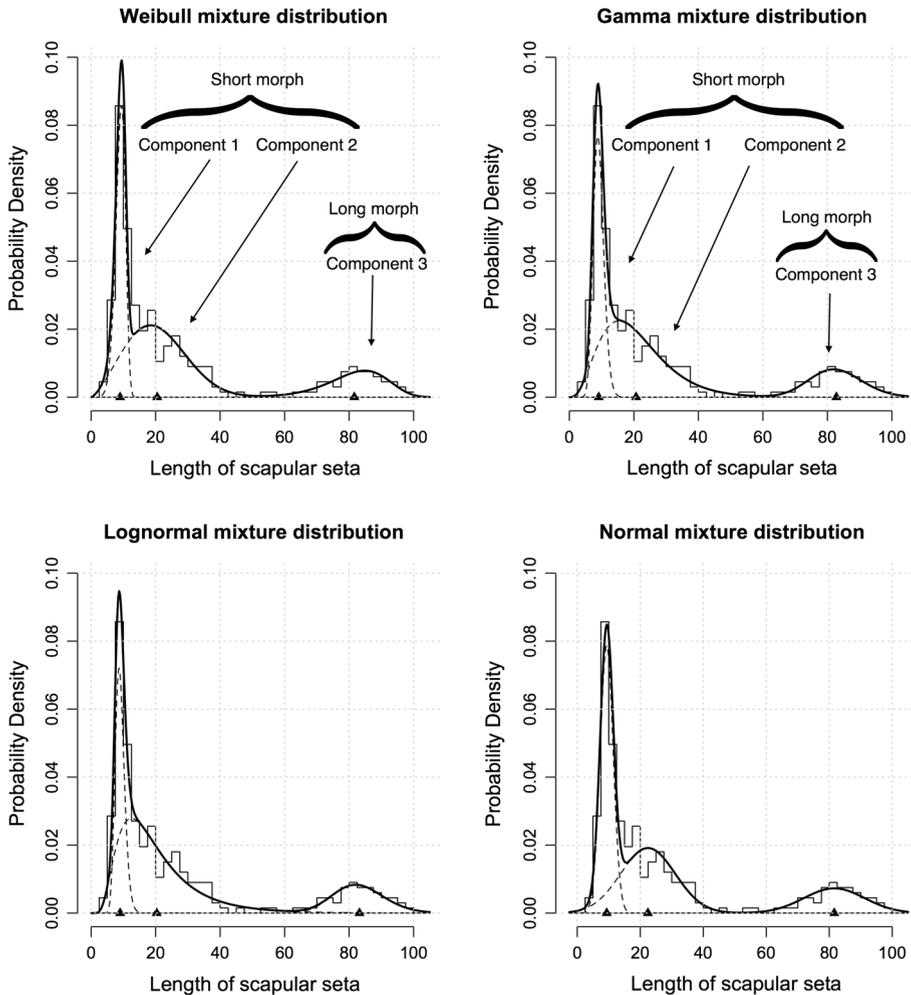
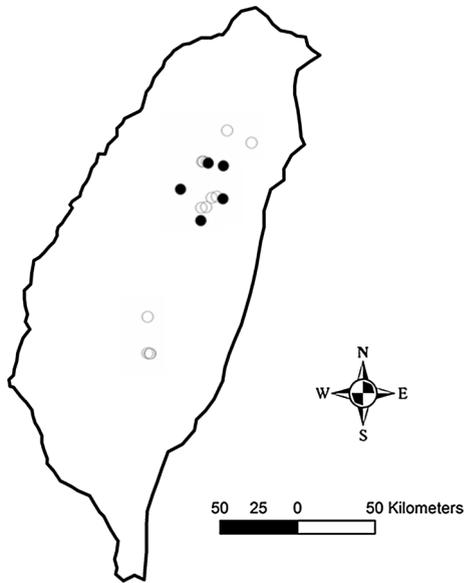


Fig. 1 Frequency distribution of length of scapular setae (histogram) imposed by correspondingly fitted distribution for each of the four finite mixture models with a *solid line*. The *dashed line* stands for the pdf of each component distribution for the fitted distribution. The *three solid triangles* indicate the three components means; the corresponding values are explicitly showed in Table 2

minutes and were then transferred to the preparatory medium in a concavity slide. The concavity slide was placed on a hotplate heated to 100 °C and the preparatory medium was stirred continuously using a probe until turning dark brown. Eriophyoid mites were later transferred to another concavity slide filled with wash medium and were left in the wash medium for at least 1 week to clean the mites. Finally, mites were singly slide-mounted in the final medium (formula recipes for the preparatory, wash and, final medium are acquired from H.H. Keifer) and were examined under a phase-contrast microscope (Leitz Laborlux S, Leica, German; objective: Plan Phaco 2, 40/0.65; eyepiece: 12.5X/18 M). Sex was determined based on the existence of genital coverflap on the females. Length of scapular setae (μm) of *A. panticis* was measured with an objective micrometer.

Fig. 2 Sites of collection of eriophyoid mite *Abacarus panticis* in Taiwan. *Open circle* sites with short scapular seta morph only; *filled circle* sites with coexistence of short and long scapular seta morphs



Finite mixture model

The frequency distribution of the lengths of scapular setae of *A. panticis* reveals a bimodal shape (Fig. 1). From a Bayesian viewpoint, finite mixture models are very flexible in fitting such datasets and can be used to categorize and define morphs of the scapular setae. Assume that there are κ components in a finite mixture model, where κ is fixed. Define the random variable Y as the length of seta and assume that given $X = i$ (component i), Y follows a distribution with the conditional probability density function (pdf) $f_{Y|X}(y|i)$, where the random variable X represents the component indicator with the probability mass function

$$P(X = i) = \pi_i, \quad i = 1, 2, \dots, \kappa,$$

where $0 < \pi_i \leq 1, i = 1, 2, \dots, \kappa$ and $\sum_{i=1}^{\kappa} \pi_i = 1$. However, we have no information about the observed lengths (Y) of setae belonging to which component (X) from our dataset. Thus, we can just fit the surveyed data to the unconditional distribution of Y , whose probability density function can be derived as

$$f_Y(y; \theta) = \sum_{i=1}^{\kappa} \pi_i f_{Y|X}(y|i). \tag{1}$$

Note that θ is an unknown parameter vector associated with the conditional models $f_{Y|X}(y|i)$ for all i and the distribution of X . Though the candidate distributions of $f_{Y|X}(y|i)$ in Eq. (1) can be different between distinct components, here we consider the same distribution over different components but with parameters varying with component. Based on a well-fitted model, we can estimate the posterior probability

$$P(X = i|Y = y) = \frac{\pi_i f(y|i)}{\pi_1 f(y|1) + \dots + \pi_{\kappa} f(y|\kappa)}, \tag{2}$$

i.e., the probability of the observation y belonging to component $i, i = 1, 2, \dots, \kappa$. Thus,

each individual can be categorized into a unique component according to the estimated posterior probabilities for different components. Four promising models—Weibull, gamma, lognormal, and normal distributions—are employed to be candidate distributions of the finite mixture model for categorizing or defining morphs of the scapular setae. Let μ_i and σ_i^2 denote the conditional mean $E(Y|i)$ and the conditional variance $\text{Var}(Y|i)$ of $f_{Y|X}(y|i)$, respectively. The pdf of each distribution is enumerated as below and the parameter vector is denoted by $\theta = (\mu_1, \dots, \mu_\kappa, \sigma_1^2, \dots, \sigma_\kappa^2, \pi_1, \dots, \pi_\kappa)$ with $\sum_{i=1}^\kappa \pi_i = 1$.

- Weibull distribution

$$f_{Y|X}(y|i) = \frac{\eta_i}{\lambda_i} \left(\frac{y}{\lambda_i}\right)^{\eta_i-1} \exp\left[-\left(\frac{y}{\lambda_i}\right)^{\eta_i}\right], y \geq 0,$$

where $\eta_i > 0$ and $\lambda_i > 0$ are the shape and the scale parameters of the Weibull distribution. For comparison purpose, the two parameters of the distribution are reparameterized by the mean $\mu_i = \lambda_i \Gamma(1 + 1/\eta_i)$ and the variance $\sigma_i^2 = \lambda_i^2 \Gamma(1 + 2/\eta_i) - \mu_i^2$.

- Gamma distribution

$$f_{Y|X}(y|i) = \frac{y^{\eta_i-1}}{\Gamma(\eta_i) \lambda_i^{\eta_i}} \exp\left[-\left(\frac{y}{\lambda_i}\right)\right], y \geq 0,$$

where $\eta_i > 0$ and $\lambda_i > 0$ are the shape and the scale parameters of the gamma distribution. Similar to the preceding case, the two parameters of the distribution are also reparameterized by the mean $\mu_i = \lambda_i \eta_i$ and the variance $\sigma_i^2 = \eta_i \lambda_i^2$.

- Lognormal distribution

$$f_{Y|X}(y|i) = \frac{1}{y \sqrt{2\pi \eta_i}} \exp\left[-\frac{(\ln(y) - \lambda_i)^2}{2\eta_i^2}\right], y \geq 0,$$

where $f_{Y|X}(y|i)$ is derived from the pdf of a normal distribution with mean λ_i and variance η_i^2 . The two parameters of the distribution are reparameterized by the mean $\mu_i = \exp(\lambda_i + \eta_i^2/2)$ and the variance $\sigma_i^2 = \exp(2\lambda_i + 2\eta_i^2) - \exp(2\lambda_i + \eta_i^2)$.

- Normal distribution

$$f_{Y|X}(y|i) = \frac{1}{\sqrt{2\pi \sigma_i}} \exp\left[-\frac{(y - \mu_i)^2}{2\sigma_i^2}\right], y \geq 0.$$

Model fitting and checking

Since the number of components κ of the finite mixture model is fixed here, ideally the unknown parameters in the model can be estimated by the maximum likelihood estimation (MLE), i.e., to maximize the likelihood function derived from Eq. (1) as a function of the parameter θ when observations y_1, \dots, y_n are given. Whereas the parameter θ is on a $(3\kappa - 1)$ -dimensional space, a conventional way is to apply the expectation–maximization (E–M) algorithm (Flury 1997) to find the MLE of θ . Further, Macdonald and Du (2011)

provided an R package (*mixdist*) in the R software environment (R Development Core Team 2009) for analyzing various finite mixture models. Though the acceptability of the considered models can be evaluated by several measures, four widely used measures are suggested for this study and are characterized as below.

- The test statistic for the likelihood ratio is denoted by

$$\chi^2 = -2 \sum_{i=1}^b O_i \log\left(\frac{E_i}{O_i}\right), \tag{3}$$

which asymptotically follows a Chi square distribution with $b - (3\kappa - 1) - 1$ degrees of freedom, where b stands for the number of bins/intervals for the analyzed data, O_i and E_i are respectively the observed and the expected/predicted number of setae in the i th bin, where E_i is evaluated from the considered model with parameters estimated by the MLE. Intuitively, if the discrepancy between the observed and expected numbers in each bin is negligible, the resulting statistic χ^2 will tend to be small and thus indicates a reasonable fitting for the data.

The p value is to compute the probability of that the Chi square statistic is larger than the value χ^2 computed by Eq. (3). In theory, if the data is promising from the indicated model, the p value should be larger than a significant level, say 0.05 or 0.01. Namely, the magnitude of p value is used to reveal the applicability regarding the indicated model. Meanwhile, the value χ^2 is not too large to coincide with a large p value.

- The root mean square errors (RMSE) is also a conventional way for checking the adequacy of an underlying model for a dataset and is computed by

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^b (O_i - E_i)^2}{b}}.$$

Similar to the test statistic χ^2 , the RMSE is to measure the discrepancy between the observed and expected numbers in each bins but with a different measurement. If the underlying model is reasonable to fit the data, the minor difference between the observed and expected numbers in each bins can make the computed RMSE small and be evidence of favoring the indicated model.

- The Bayesian information criterion (BIC) is defined by

$$\text{BIC} = -2 \sum_{i=1}^n \log(f_Y(y_i; \hat{\boldsymbol{\theta}})) + (3\kappa - 1) \log(n),$$

where n (sample size) is the number of individuals in the given data and $\hat{\boldsymbol{\theta}}$ is the MLE of the parameter vector $\boldsymbol{\theta}$. Note that $\hat{\boldsymbol{\theta}}$ is obtained by maximizing the likelihood function derived from Eq. (1) or equivalent to minimizing $-2 \sum_{i=1}^n \log(f_Y(y_i; \boldsymbol{\theta}))$ in the first term of the BIC formula. In contrast, the second term of the BIC can be treated as a penalty for adjusting the well fitness of an underlying model is caused by increasing the sample size n or the number of parameters $3\kappa - 1$. As a result, a favored model requires having a smaller BIC value than other competitive models.

For our dataset of the lengths of scapular setae, the four measures were computed for each model we have considered and tabulated in Table 1. Though the four candidate distributions are relevant to portray our data set at the significance level $\alpha = 0.05$

Table 1 Finite mixture models evaluated by four measures regarding the adequacy of model fitting

	Weibull	Gamma	Lognormal	Normal
χ^2	21.23	21.72	24.97	28.71
p value	0.93	0.92	0.81	0.63
RMSE	1.62	1.39	1.53	2.17
BIC	2,094.95	2,090.67	2,092.48	2,097.99

Table 2 The estimates of component means and variances for two favored models. In addition, the range of each component for a given model is used to define two scapular seta morphs

Component (i)	Weibull			Gamma		
	1	2	3	1	2	3
$\hat{\pi}_i$	0.32	0.50	0.18	0.31	0.52	0.17
$\hat{\mu}_i$	8.98	20.48	81.59	9.10	20.68	82.78
$\hat{\sigma}_i^2$	1.54	9.11	9.55	1.64	10.41	8.21
Minimum	5.88	4.20	54.60	5.88	4.20	63.70
Maximum	11.65	46.62	99.70	11.65	56.20	99.70

compared with the p values, we suggest using the Weibull or gamma distribution based on simultaneously taking the four criteria into account. Thus, the parameters of the two favored finite mixture models were estimated by means of the R package *mixdist* and enumerated in Table 2. Though our data set shows a bimodal shape, two components ($\kappa = 2$) may be a good starting. Whereas, since the p values associated with the four models are extremely smaller than $\alpha = 0.05$ when $\kappa = 2$ and $\kappa \geq 4$, we have only showed the results at $\kappa = 3$ for Tables 1 and 2. The E–M algorithm adopted in the package *mixdist* was discussed in Du (2002) in detail. First of all, the working data needs to be binned first. Since a larger binned size can result in a smaller degrees of freedom for the Chi square test and thus the test is difficult to detect the violation for the data to an underlying model. Reversely, the Chi square test will be sensitive to detecting the discrepancy between the observed and expected numbers in each bin and can be in company with a large type I error (the dataset is truly from the candidate distribution, but the Chi square test is significant and give a conclusion to abandon the candidate distribution). As a consequence, the binned size in our study was set at 2.5 (μm). In addition, the pdfs of the four fitted distributions were also imposed on the histogram of frequencies of lengths of scapular setae for reference; see the solid lines of Fig. 1. From the figure, it reveals that the four considered models are fitted satisfactorily and showed why there needs three components in the finite mixture models (Fig. 1).

Geometric morphometric analyses

The difference in shape and size between short and long seta morphs was analyzed by geometric morphometric techniques. Eight pairs of ventral setal tubercles locating in opisthosoma and coxigenital region were selected as landmarks for further analyses (Fig. 3). Images of body of *A. panticis* were obtained using a camera system consisting of a phase contrast optical microscope attached to a computer. These 16 landmarks were

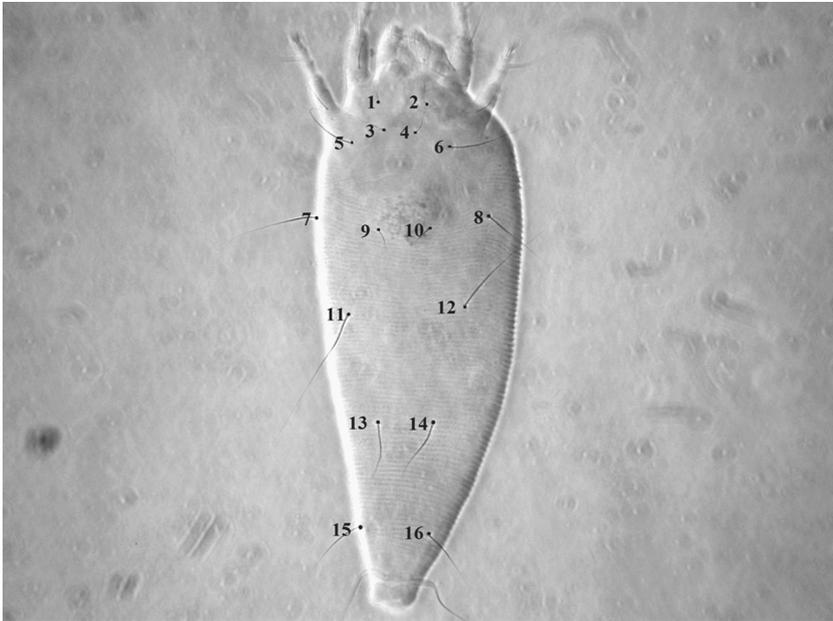


Fig. 3 Sixteen landmarks (8 pairs) on ventral region of eriophyoid mite *Abacarus panticis* for geometric morphometric studies

recorded by tpsDig software (Rohlf 2010) and were expressed as Cartesian coordinates. Shape differences between short and long seta morphs were investigated using generalized least-square (GLS) Procrustes superimposition methods (Rohlf and Slice 1990) to separate shape information from orientation, position, and size (Zelditch et al. 2004). Relative warp analysis with $\alpha = 0$ was used to compare shape difference between the two morphs. The partial warp scores were then calculated by the thin-plate spline. The uniform and non-uniform shape components were used for a Hotelling's T^2 test to examine the differences in shape between the two seta morphs. Discriminant analysis was used to test the successful classification rate. Centroid size, the square root of the sum of the squared distance between the centroid and all landmarks (Bookstein 1991), was used to evaluation differences in size between short and long seta morphs and is tested by t test. We performed a multivariate regression of shape variables on centroid size to detect any allometric effects for two seta morphs separately. Centroid size was logarithmically transformed before analysis. All the geometric morphometric analyses were implemented in PAST ver. 2.07 (Hammer et al. 2001).

Results

The definition of two scapular seta morphs

Length of scapular setae of 414 *A. panticis* was measured, including 266 females, 127 males and 21 nymphs. Because of the difference in length of scapular setae among females, males, and nymphs (Kruskal–Wallis test, $H = 20.9$, $p < 0.001$) and the small sample size

of males and nymphs, only females were analyzed. Based on the two favored models (Weibull and gamma distribution), we calculated the posterior probabilities for the data set using the formula in Eq. (2) with the estimated parameters enumerated in Table 2. The observation y is categorized to the i th component if $P(X = i|Y = y)$ is the maximum over $P(X = x|Y = y)$, $x = 1, 2, 3$. Though there are three components in the two selected mixture distributions, the lengths of scapular setae are only categorized into two morphs due to the overlapped length ranges for components 1 and 2. From Fig. 1, it is observed that the first two dashed lines (the pdfs of the first two components) are greatly overlapped. Similarly, Table 2 also shows that the range of the first component is a subset of that of the second component, i.e., the interval [5.88, 11.65] of the first component is included in [4.20, 46.62] of the second component. Moreover, it is reasonable to characterize the data as a bimodal pattern based on the corresponding pdfs of the two favored mixture models. Consequently, the short morph is defined as the combination of the first two components and the long morph as the third component. Based on the finite Weibull mixture model, lengths of scapular setae of short morph range from 4.2 to 46.62 μm and long morph from 54.6 to 99.7 μm . A total of 219 and 47 females belong to the short and long morph, respectively. For the finite gamma mixture model, short seta morph ranges from 4.2 to 56.2 μm and long seta morph from 63.7 to 99.7 μm . Short and long morph include 221 and 45 females, respectively. The long and short morphs co-occur at five sites, while the other 13 sites only contain the short morph (Fig. 2).

Geometric morphometric analyses

A total of 124 female *A. panticis*, including 88 of short seta morph and 36 of long seta morph, were mounted in dorsoventral position and are selected for further evaluation of shape/size variation between the two seta morph types. There was a significant difference in shape between the short and long morphs (MANOVA, Wilks' $\lambda = 0.69$; $F_{25, 161} = 2.9$; $p < 0.01$), but centroid size was similar between the two morphs (t test, $t = 0.5$; $p = 0.6$; mean CS for short morph = 743.7; mean CS for long morph = 700.5). The overall successful classification rate is 81.5 %, with 80.6 and 83.3 % for short and long morph, respectively. Variation in shape was correlated with centroid size in short morph (Wilks' $\lambda = 0.44$, $F_{28, 2408} = 5.5$, $p < 0.0001$), but was unrelated to centroid size in long morph (Wilks' $\lambda = 0.13$, $F_{28, 952} = 0.74$, $p = 0.82$). The first two relative warps (RW) accounted for 42.3 % of variation in shape, with 23.6 % for RW1 and 18.7 % for RW2, respectively. In general, *A. panticis* of long morph was more elongated than those of short morphs in coxal area but less elongated in the caudal part of opisthosoma (Fig. 4).

Discussion

Continuous variation in the length of scapular setae might render an arbitrary classification of different morph types untenable. Relative to some typical methods such as discriminant analysis and cluster analysis based on deterministic methods that require several morphological variables to establish categorizing rules, our aim is firstly to determine a threshold for classifying length of scapular seta, and subsequently to investigate morphological variation between different morphs. More specifically, our classification solely relies on length of setae, so a model-driven method is employed in this study. Among several available models in statistical textbooks or in the literature, we suggest using four widely used models to more objectively classify seta morphs and find two of them, gamma

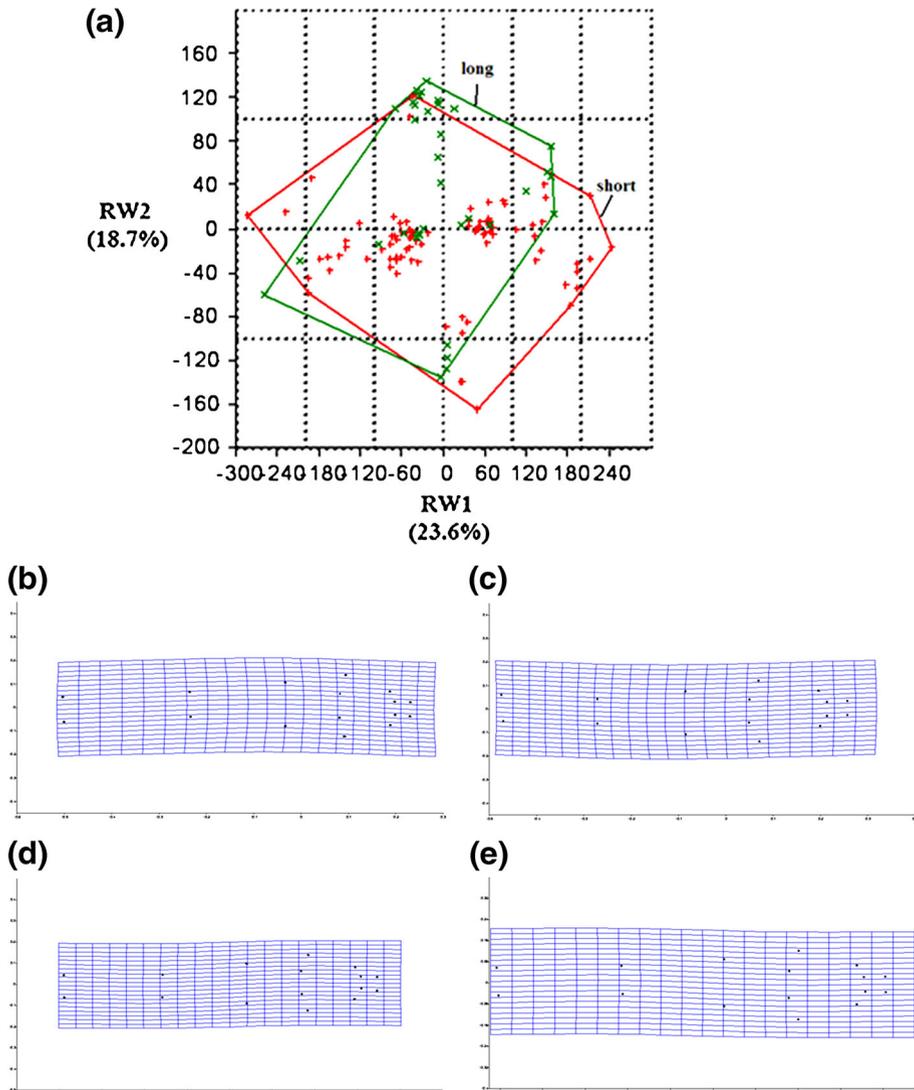
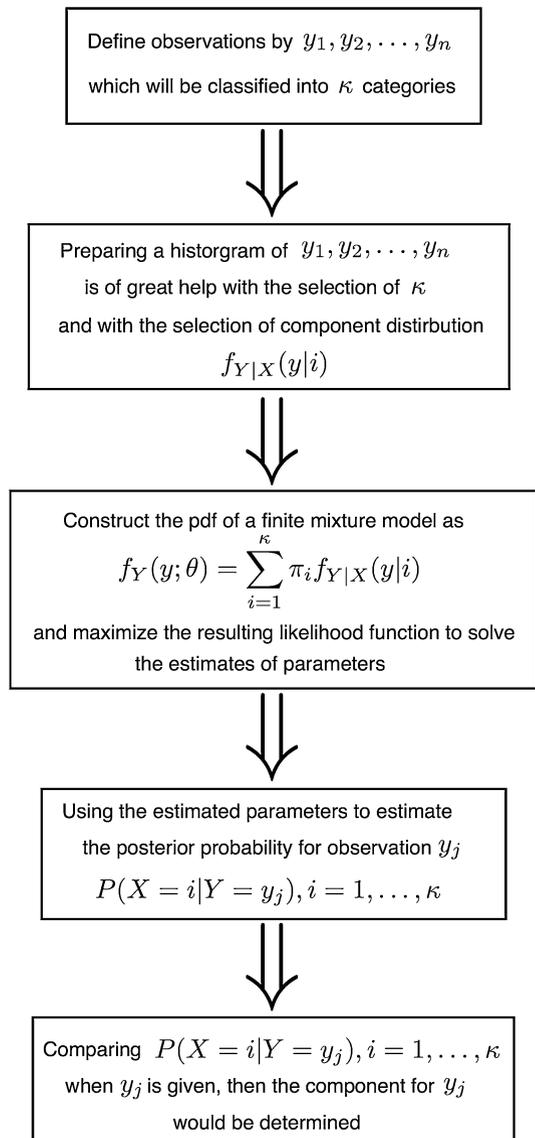


Fig. 4 Relative warps (RW) of opisthosoma and coxigenital region of *Abacarus panticis*. **a** Scatterplot of individual scores from RW analysis (short (+); long(x)); **b** positive deviation of RW1; **c** negative deviation of RW1; **d** positive deviation of RW2; **e** negative deviation of RW2

and Weibull, satisfactory in fitting our data and successful in categorizing length of scapular seta of *A. panticis* into long and short morphs. Similar approach should be applicable to objectively classify continuous morphological characters of other organisms, but in our knowledge this has still been rarely applied. To facilitate the usage of finite mixture model in other similar topics, a flowchart on data analyses has been included (Fig. 5).

Knowledge on how and why eriophyoid mites disperse is still limited despite its significance for pest management (Michalska et al. 2010). For instance, the coconut mite

Fig. 5 The procedure for finite mixture model analyses



(*Aceria guerreronis*) is one of the most devastating coconut pests worldwide and elucidates how they disperse could assist its control (Galvão et al. 2012; Navia et al. 2013). Because disperse by wind is considered to be the primary dispersal mode for eriophyoid mites (Michalska et al. 2010; Galvão et al. 2012), it is critical to evaluate whether eriophyoid mites evolve specialized morphological characters, such as longer scapular setae, to facilitate aerial dispersal. Scapular setae of *A. panticis* can be classified into long and short morphs, and both morphs coexist in some study sites (Fig. 2). If the scapular setae could facilitate wind dispersal (Krantz 1973; Lindquist and Oldfield 1996; Sabelis and Bruin 1996), it would be expected that *A. panticis* of long seta morph will be more widely distributed than the short seta morph for their higher dispersal capability. On the contrary,

we found that short morphs occurred over wider areas (Fig. 2). Accordingly, longer scapular setae seem unlikely a specialized adaptation for aerial dispersal. Whether long seta morph develops under specific environmental conditions, is maintained by neutral mutation, or is reproductively isolated from short morph cannot be assessed from the current study and related studies on eriophyoid mites are still lacking. Indeed, a study on *Nenteria pandionis* mite (Acari: Uropodina) concluded that causes for individual variation in setal morphology should be complex, and might depend on age, developmental conditions, or genetic variation (Adamski et al. 2008). Because setae in general could serve sensory functions, are potentially receptors of chemical particles (Nuzzaci and Alberti 1996), or could help orient in the habitat as in the feather mite (Acari: Astigmata) (Proctor 2003), whether longer scapular setae could facilitate *A. panticis* in locating food or escaping from predators warrant further study.

Modern geometric morphometric techniques enable the preservation of detailed shape information while at the same time allowing for sophisticated statistical analyses in comparative morphological research (Zelditch et al. 2004). Unsurprisingly, geometric morphometrics has gained much attention in recent decades and has been widely applied to several research topics and fields of studies (Rohlf 1998; Richtsmeier et al. 2002; Adams et al. 2004; Slice 2007; Mitteroecker and Gunz 2009). Among these, unraveling intra-specific variation in morphology and exploring potentially causative forces underlying the differentiation is commonly employed (Lawing and Polly 2010). For example, Idaszkin et al. (2013) revealed shape difference in carapace of one crab species between two habitat types using geometric morphometric methods. Similarly, variation in cranial shape of an Australian marsupial and their association with latitudes were also investigated with geometric morphometric techniques (Dawson and Milne 2012). Baran et al. (2011) applied similar approach to differentiate species of oppiid mites. Nevertheless, the application of geometric morphometrics to eriophyoid mites remains very rare (but see Navia et al. 2006) despite some species are important agricultural pests and investigating geographic variation in morphology is insightful for adopting more efficient pest control strategy (Navia et al. 2006). Navia et al. (2006) compared morphology of an eriophyoid mite *A. guerreronis*, a worldwide coconut pest, from America, Africa, and Asia, and found more morphometric variation among mites from America than those from Africa and Asia, suggesting its American origin and the potential to find suitable natural enemies over there. In this study, we applied 16 landmarks and found shape difference between *A. panticis* of long vs. short seta morph, with a satisfactory successful classification rate of 81.5 %. This demonstrates that geometric morphometrics can be effectively applied to differentiate closely related morph of minute creatures such as eriophyoid mites. Further studies could take advantage of this comprehensive and rigorous method in quantifying morphometric variation and in helping resolve dilemmas regarding species or subspecies classification in eriophyoid mites. Moreover, combined with other ecological information, such as distribution or habitat use, geometric morphometrics should advance our understanding of important mechanisms leading to divergence in eriophyoid mites.

Acknowledgments We are indebted to W.C. Yang for field collection and S. Huang, R. Wang, J. Jeng, S.K. Hu for organizing the data for analyses. The work was financially supported by National Science Council of Taiwan for T.J. Shen, National Taiwan University for C.C. Kuo, and National Museum of Natural Science of Taiwan for K.W. Huang.

References

- Adams DC, Rohlf FJ, Slice DE (2004) Geometric morphometrics: ten years of progress following the 'revolution'. *Ital J Zool* 71:5–16
- Adamski Z, Bloszyk J, Gwiazdowicz DJ (2008) Individual variability of setal morphology in *Nenteria pandioni* (Acari: Mesostigmata: Uropodina): genetic variability or aging? *Biologia* 63:236–244
- Almeida MT, Bisby FA (1984) A simple method for establishing taxonomic characters from measurement data. *Taxon* 33:405–409
- Amrine JW Jr, Stasny TA, Flechtmann CHW (2003) Revised keys to world genera of Eriophyoidea (Acari: Prostigmata). Indira Publishing Houses, West Bloomfield
- Baran S, Altun A, Ayyildiz N, Kence A (2011) Morphometric analysis of oppiid mites (Acari, Oribatida) collected from Turkey. *Exp Appl Acarol* 54:411–420
- Bookstein FL (1991) Morphometric tools for landmark data: geometry and biology. Cambridge University Press, Cambridge
- Dawson R, Milne N (2012) Cranial size and shape variation in mainland and island populations of the quokka. *J Zool* 288:267–274
- De Lillo E, Amrine JW Jr (2003) Catalogue of the Eriophyoidea of the world. Version Computer of Filemaker Pro 4.0
- Du J (2002) Combined algorithms for constrained estimation of finite mixture distributions with grouped data and conditional data. Dissertation, McMaster University
- Ezard T, Pearson P, Purvis A (2010) Algorithmic approaches to aid species' delimitation in multidimensional morphospace. *BMC Evol Biol* 10:175
- Flury B (1997) A first course in multivariate statistics. Springer, New York
- Fonseca JRS, Cardoso MGMS (2007) Mixture-model cluster analysis using information theoretical criteria. *Intell Data Anal* 11:155–173
- Ford EB (1945) Polymorphism. *Biol Rev* 20:73–88
- Galvão A, Melo JS, Monteiro V, Lima D, Moraes G, Gondim M (2012) Dispersal strategies of *Aceria guerreronis* (Acari: Eriophyidae), a coconut pest. *Exp Appl Acarol* 57:1–13
- Hammer Ø, Harper DAT, Ryan PD (2001) Past: paleontological statistics software package for education and data analysis. *Palaeontol Electron* 4:9
- Idaszkin YL, Marquez F, Nocera AC (2013) Habitat-specific shape variation in the carapace of the crab *Cyrtograpsus angulatus*. *J Zool* 290:117–126
- Jaworski A, Podlaski R (2012) Modelling irregular and multimodal tree diameter distributions by finite mixture models: an approach to stand structure characterization. *J For Res* 17:79–88
- Kéry M, Royle JA, Schmid H (2005) Modeling avian abundance from replicated counts using binomial mixture models. *Ecol Appl* 15:1450–1461
- Krantz GW (1973) Observations on the morphology and behavior of the filbert rust mite, *Aculus comatus* (Prostigmata: Eriophyoidea) in Oregon. *Ann Entomol Soc Am* 66:709–717
- Kuo CC, Wang CF, Tsai JI, Feng FL, Huang KW (2013) Altitudinal variation in body size in *Abacarus panticis* Keifer (Acari: Eriophyidae). *Ecol Res* 28:1091–1097
- Lawing AM, Polly PD (2010) Geometric morphometrics: recent applications to the study of evolution and development. *J Zool* 280:1–7
- Lindquist EE, Oldfield GN (1996) Evolution of eriophyoid mites in relation to their host plants. In: Lindquist EE, Sabelis MW, Bruin J (eds) *Eriophyoid mite—their biology, natural enemies and control*. Elsevier, Amsterdam, pp 277–300
- Macdonald P, Du J (2011) Mixdist: finite mixture distribution models. R package version 0.5-4. (<http://CRAN.R-project.org/package=mixdist>)
- Macdonald PDM, Pitcher TJ (1979) Age-groups from size-frequency data: a versatile and efficient method of analysing distribution mixtures. *J Fish Res Board Can* 36:987–1001
- McLachlan GJ, Chang SU (2004) Mixture modeling for cluster analysis. *Stat Methods Med Res* 13:347–361
- McLachlan GJ, Peel D (2005) Finite mixture models. Wiley, New York
- Michalska K, Skoracka A, Navia D, Amrine J (2010) Behavioural studies on eriophyoid mites: an overview. *Exp Appl Acarol* 51:31–59
- Millar RB (1987) Maximum likelihood estimation of mixed stock fishery composition. *Can J Fish Aquat Sci* 44:583–590
- Mitteroecker P, Gunz P (2009) Advances in geometric morphometrics. *Evol Biol* 36:235–247
- Navia D, de Moraes GJ, Querino RB (2006) Geographic variation in the coconut mite, *Aceria guerreronis* Keifer (Acari: Eriophyidae): a geometric morphometric analysis. *Int J Acarol* 32:301–314
- Navia D, Gondim M Jr, Aratchige N, Moraes G (2013) A review of the status of the coconut mite, *Aceria guerreronis* (Acari: Eriophyidae), a major tropical mite pest. *Exp Appl Acarol* 59:67–94

- Nuzzaci G, Alberti G (1996) Internal anatomy and physiology. In: Lindquist EE, Sabelis MW, Bruin J (eds) Eriophyoid mites—their biology, natural enemies and control. Elsevier, Amsterdam, pp 101–150
- Pearson K (1894) Contributions to the theory of mathematical evolution. Philos Trans R Soc Lond 185:71–110
- Proctor HC (2003) Feather mites (Acari: Astigmata): ecology, behavior, and evolution. Annu Rev Entomol 48:185–209
- R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, (<http://www.r-project.org>)
- Reid G, Sidwell K (2002) Overlapping variables in botanical systematic. In: MacLeod N, Forey PL (eds) Morphology, shape and phylogeny. Taylor & Francis, London, pp 53–66
- Richtsmeier JT, Deleon VB, Lele SR (2002) The promise of geometric morphometrics. Yearb Phys Anthropol 45:63–91
- Rohlf FJ (1998) On applications of geometric morphometrics to studies of ontogeny and phylogeny. Syst Biol 47:147–158
- Rohlf FJ (2010) TpsDig software for windows version 2.16. Department of Ecology and Evolution, State University of New York, New York. (<http://life.bio.sunysb.edu/morph/>)
- Rohlf FJ, Slice D (1990) Extensions of the procrustes method for the optimal superimposition of landmarks. Syst Biol 39:40–59
- Roulin A (2004) The evolution, maintenance and adaptive function of genetic colour polymorphism in birds. Biol Rev 79:815–848
- Sabelis MW, Bruin J (1996) Evolutionary ecology: life history patterns, food plant choice and dispersal. In: Lindquist EE, Sabelis MW, Bruin J (eds) Eriophyoid mite—their biology, natural enemies and control. Elsevier, Amsterdam, pp 329–366
- Skulason S, Smith TB (1995) Resource polymorphisms in vertebrates. Trends Ecol Evol 10:366–370
- Slice DE (2007) Geometric morphometrics. Annu Rev Anthropol 36:261–281
- Tixier MS (2012) Statistical approaches to assess intraspecific variations of morphological continuous characters: the case study of the family Phytoseiidae (Acari: Mesostigmata). Cladistics 28:489–502
- Wang CF, Huang KW (2011) Taxonomy of eriophyoid mites (Acari: Trombidiformes) infesting *Yushania nitakayamensis* (Poaceae: Bambuseae) in Taiwan. Formosan Entomol 31:249–280
- Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) Geometric morphometrics for biologists: a primer. Elsevier, San Diego
- Zhang L, Gove JH, Liu C, Leak WB (2001) A finite mixture of two Weibull distributions for modeling the diameter distributions of rotated-sigmoid, uneven-aged stands. Can J For Res 31:1654–1659