BIOTROPIA No. 13, 1999: 37-48

MORPHOMETRIC STUDY FOR IDENTIFICATION OF THE *BACTROCERA DORSALIS* COMPLEX (DIPTERA : TEPHRITIDAE) USING WING IMAGE ANALYSIS

A. ADSAVAKULCHAI[1], V. BAIMAI[1], W. PRACHYABRUED[2], PAUL J. GROTE[1] and S, LERTLUM[3]
[1] Department of Biology and[2] Department of Physics, Faculty of Science,
Mahidol University, Rama IV Road, Bangkok, 10400, Thailand
e-mail address: toung@ait.ac.th
[3] School of Biology, Institute of Science, Suranaree University of Technology, Thailand
[4] Faculty of Computer Science Program, Chulachomklao Military Academy, Thailand

ABSTRACT

The *Bactrocera dorsalis* complex (Diptera: Tephritidae) used in this study included *B. dorsalis, B. arecae,* B. *propinqua, B. pyrifoliae, B. verbascifoliae,* and three new species complexes are species E, species K and species P. *Bactrocera tau* was used as an out-group. A total of 424 adults, which emerged from pupae collected from natural populations in Thailand, were prepared for wing measurements. Morphometric analysis was performed on measurements of wing vein characters. Wing images were captured in digital format and taken through digital image processing to calculate the Euclidean distance between wing vein junctions. Discriminant and cluster analyses were used for dichotomy of classification processes. All 424 wing specimens were classified to species in terms of the percentage of "grouped" cases which yielded about 89.6% accurate identification compared with the formal description of these species. After clustering, the percentage of "grouped"cases yielded 100.0%, 98.9%, 98.1%, 95.2% and 84.6% accurate identification between the *B. dorsalis* complex and *B. tau; B. arecae* and Species E; *B. dorsalis* and *B. verbascifoliae; B. propinqua* and *B. pyrifoliae;* and species K and species P, respectively. This method of numerical taxonomy may be useful for practical identification of other groups of agricultural pests.

Key words: *Bactrocera dorsalis* complex/wing image processing/morphometric/discriminant and cluster analyses.

INTRODUCTION

The Oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae) group of the subgenus *Bactrocera,* has been considered one of the most important groups of agricultural pest in Southeast Asia because some of these species attack seed bearing organs of plants, including soft fruits and flowers (McPheron and Steck 1996). The *B. dorsalis* group comprises about 52 closely related species in Asia, mostly Southeast and South Asia, with additional species in the south Pacific region (Drew and Hancock 1994). Some of these species are mophologically similar. Drew (1989) postulated that the Dacinae fruit flies originated in the Papua New Guinea area and speciated prolifically throughout the region. Drew and Hancock (1994) listed fourteen closely related species of the *B. dorsalis* complex from Thailand on the basis of morphological characters. These species are *B. arecae, B. carambolae, B. dorsalis, B. irvingiae, B. kanchanaburi, B. melastomatos, B. osbeckiae, B. papayae, B. propinqua, B. pyrifoliae, B. raiensis, B. thailandica, B. unimaculata* and *B. verbascifoliae.* Recently, we provided population genetic data of some members of the *B. dorsalis* complex (Baimai *et al.* 1995, 1998 unpublished data; Satayalai

1996). Nonetheless, most species of the *B. dorsalis* complex have limited distribution within the tropical and subtropical regions (Drew 1989). The limitation of the distribution range of these species is due in part to physical, climatic and gross vegetation factors. However, it is more likely that the distribution range is correlated with the specificity on fruits of particular host plants, yet little information is available on the range of host plants of these species.

The *B. dorsalis* complex is systematically one of the most interesting groups of insect pest (Ibrahim and Ibrahim 1990). Because of similarity in external morphology among the members of the *B. dorsalis* complex and the geographic variation in morphology within each species, it has been very difficult to separate these species. Consequently, such morphological variation has caused taxonomic problems (Hardy 1977). Thus, the most common errors are synonyms, homonyms, misidentifications and establishment of supra-specific groups based on questionable morphological characters (White and Elson-Harris 1992). There is still a major need for more taxonomic study in correlation with population genetic investigations of the *B. dorsalis* complex to address some sibling species problems. In most countries, a complete list of reference collections for identification purposes can be found, resulting from the work of trained taxonomists. It has long been obvious that fruit flies of the subfamily Dacinae have major economic effects on society. Therefore, economic entomologists were needed to identify the various species involved. In spite of this tireless work, however, many taxonomic problems and misidentifications accrued (McPheron and Steck 1996). Some of these systematic problems can be elucidated with the aid of the recent development of taxonomic techniques such as cytotaxonomy and molecular biology as well as improved numerical taxonomy (Sneath and Sokal 1973). Yu *et al.* (1992) studied morphometric analysis of linear wing measurements for identification of ichneumonid wasps using image analysis of wings. The authors outlined the procedure to digitize and to measure various wing elements with an image analyzer and the wing specimens were assigned to species by discriminant analysis and independent univariate comparisons of wing measurements. Recently, Weeks (1996) developed Daisy (Digital Automated Identification System) based on the idea that the pattern of veins and pigments on insect wings are distinct-like fingerprints. Much of the acquisition by computer of morphological characters of insects (White and Scott 1994) has involved measurement of projected images on digitizing tablets (Howell *et al.* 1982). Use of computers efficiently provides accurate measurements and saves development time. In addition, these methods can be easily repeated and made available to any user and reworked with a minimum of effort. However, these methods are costly hi terms of software development and maintenance. Image analysis of morphological characters of wings may be the first step towards a completely automated insect identification technique *Cfuetal.* 1992).

In our ongoing research on the biology of fruit flies hi Thailand, we attempt to employ ecological observations in the field and genetic investigations in the laboratory coupled with morphological examination of larvae, pupae and adults to help solve the problems of identifying some cryptic or isomorphic species. Thus some new sibling species of the *B. dorsalis* complex have been found through

allozyme electrophoresis (Satayalai 1996) and cytogenetic studies (Baimai *et al.* 1995, 1998 unpublished data). In this paper we describe the methodology of image analysis to acquire and quantify morphological characters of the wing veins of eight species of the *B. dorsalis* complex in a computer compatible form for suitable identification of these species.

MATERIALS AND METHODS

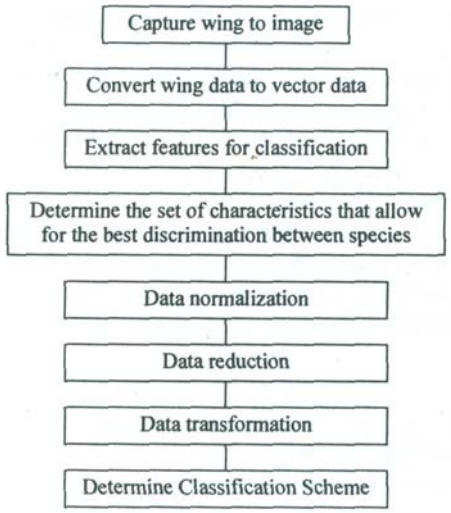
Specimen collection

Eight species of the *B. dorsalis* complex used in this study include *B. dorsalis* (Hendel), *B. arecae* (Hardy and Adachi), *B. propinqua* (Hardy and Adachi), *B. pyrifoliae* Drew and Hancock, *B. verbascifoliae* Drew and Hancock, and three new species complexes are species E, species K and species P with morphological characters different from the record of Drew and Hancock (1944). In addition, *Bactrocera tau* (Walker), of the subgenus *Zeugodacus*, was used as the out-group. Larval specimens of these members of the *B. dorsalis* complex were obtained from a wide variety of infested fruits from various parts of Thailand (Table 1). Some larvae were processed for mitotic karyotype study which provided useful information for

Table 1. Specimens of the eight species of the *Bactrocera dorsalis* complex and *Bactrocera tau* collected from different localities in Thailand

Species	Host plant species	Wings of females	Wings of males	Locality	Date of collection
B. dorsalis	*Syzygium malaccense* (L.) Merr. and L.M. Perry	21	29	Bangkok	10.10.1996
B. arecae	*Areca catechu* L.	9	10	Suratthani	05.03.1994
	Areca catechu L.	10	11	Suratthani	04.03.1994
	Areca catechu L.	11	9	Narathiwas	17.06.1994
	Areca catechu L.	10	10	Songkhla	18.06.1994
B. propinqua	*Garcinia* sp. (Guttiferae)	27	25	Ranong	16.04.1996
B. pyrifoliae	*Artocarpus chaplasha* Roxb.	22	28	Chiangmai	23.11.1995
B. verbascifoliae	*Solanum erianthum* D.Don	12	10	Suratthani	05.03.1994
	Solanum erianthum D.Don	8	9	Narathiwas	17.06.1994
	Solanum erianthum D.Don	6	9	Songkhla	18.07.1994
Species E	*Nauclea* sp.	19	10	Saraburi	05.10.1994
Species K	*Payena* sp. (Sapotaceae)	21	29	Ranong	19.04.1996
Species P	*Nauclea brunnea* Craib	9	13	Phang Nga	14.08.1994
Bactrocera tau	*Trichosanthes cordata* Roxb.	8	12	Ranong	23.11.1995
	Cucurbita moschata Decne	6	8	Chiangmai	25.11.1995
	Psidium guajava L.	6	10	Phetchaboon	22.01.1996

species identification. Most of the larvae were reared either in the field or in the laboratory allowing them to pupate and finally emerge as adults. Some adults were processed for electrophoretic study to confirm the genetic species as determined by mitotic chromosome markers. Adults from each collection were examined morphologically for species identification in correlation with chromosomal evidence and electromorphic allozyme patterns. Some adults were kept for wing specimens preparation. The framework of the research is shown in Figure 1.



Wing preparation

Wings of individual adults were detached from the thorax and they were placed on a microscope slide. The wings were secured under a coverslip with Canada Balsam.

Image processing

Wing image processing procedure is shown in Figure 2. The microscope slide with wing samples was positioned on a Nikon SMZ-2T stereomicroscope with a low objective lens (Ix). The vertical tube has a control light path switchover which allows the diversion of the right eye image to the camera. A Nikon E2s Digital Still Camera was attached with a CF projection lens (4x) that captured the whig image on the memory card. Digital imaging with high-resolution of 1.3 million pixels was then transferred to application handling JPEG (Joint Photographic Experts Group) files (Fig. 3).

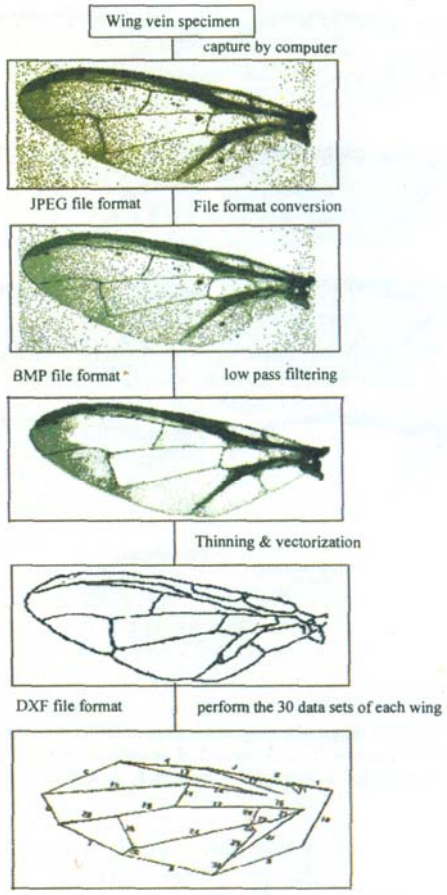


Figure 2. Wing image processing.

The original color image (JPEG : raster format) was transformed into gray scale in the form of BMP (BitMaP) raster file format that allows efficient localized image processing. From BMP, the image was pre-processed by low-pass filtering with average of 3x3 to create a smooth image (Gonzalez and Woods 1992); then the image was transformed by applying a linear function to enhance the image by calculating the ratios from pixel values of the original image divided by pixel values of the smooth image. Then, the image was vectorized to create vector file format in DXF (Drawing exchange File) file format and manually adjusted until suitable for measuring the 30 wing vein distances (Fig. 4). The vector file was used to create

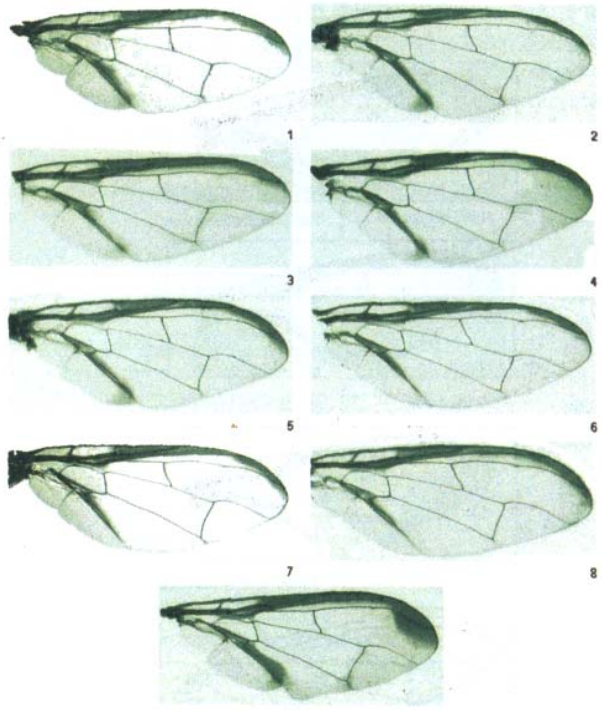


Figure 3. Showing digital image of a wing in the form of JPEG file format;
1. *Bactrocera arecae* 2. *B. dorsalis* 3. *B. propinqua* 4. *B. pyrifoliae*
5. B. verbascifoliae 6. Species E 7. Species K 8. Species P 9. *Bactrocera tau*

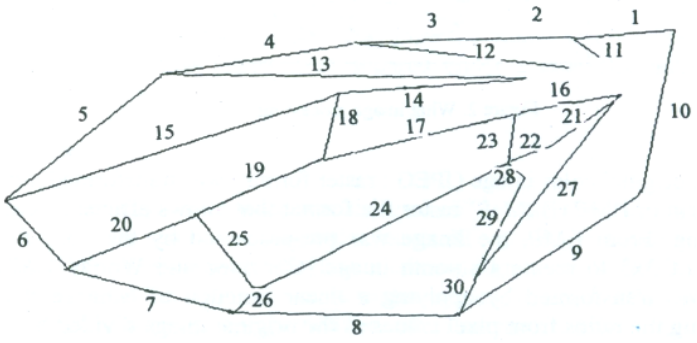


Figure 4. Diagrammatic representation of a wing of the *B. dorsalis* complex showing the 30 wing vein measurements.

automatically the coverage that contained 30 data sets of each sample in terms of Euclidean distances. Euclidean distance is the measurement between the 2 co-ordinates and is computed as the square root of the sum of the squared differences as shown in the following formula:

$$\text{Euclidean distance (D)} = [\ (x\text{-}s)2 + (y\text{-}t)2\ ]^{VI}$$

Where Euclidean distance is the distance between two points : (x,y) is co-ordinate of p; (s,t) is co-ordinate of q.

Statistical analysis

SPSS for windows version 6.0 (George and Paul 1995) was used for data analysis. Discriminant and cluster analyses were used for dichotomy of the classification processes. All 424 wing specimens were classified to species in terms of the percentage of "grouped" cases.

RESULTS AND DISCUSSION

Our data show that the length of a vein is correlated with the size of the wing. Ratios of the vein lengths provide a very effective means for recognizing trends in variation and for a quick diagnostic character. Discriminant function analysis was used to derive a function as a criterion for separation of the eight species used in this study. The percentage of "grouped" cases correctly classified with the accuracy of 89.6% is shown in Table 2. The linear discriminant function can completely discriminate members of the *B. dorsalis* complex from *B. tau* as follows:

$$Y = 2.76xj + 9.85x_2 - 16.74$$

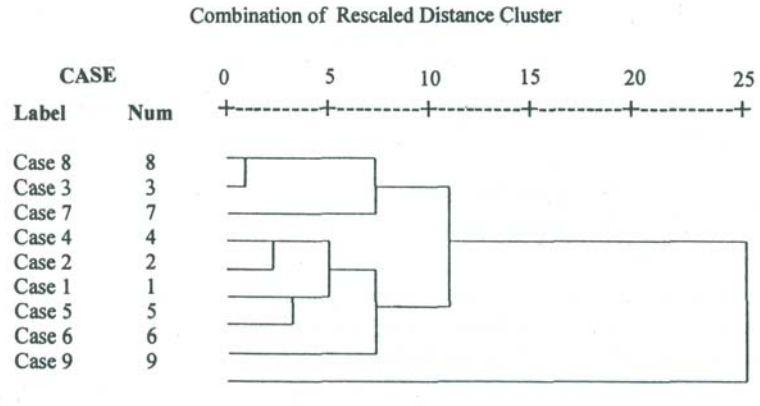
Where X! = vein2/vein22, and x_2 = vein3/vein8. Specimens are identified by the following rule :

If Y < 0 then species is *Bactrocera dorsalis* complex
If Y > 0 then species is *Bactrocera tau*

The eight species of the *B. dorsalis* complex can also be separated by using cluster analysis (Fig. 5). The results from cluster analysis show that the eight members of the *B. dorsalis* complex can be classified into 4 classes : (1) *B. arecae* and species E; (2) *B. verbascifoliae* and *B. dorsalis',* (3) *B. propinqua* and *B. pyrifoliae;* and (4) species K and species P. *Bactrocera tau* was used as the out-group. Discriminant function analysis was used to derive a function to provide maximum values for separation of these 4 classes. The "grouped" cases were correctly classified with an accuracy of 98.9%, 98.1%, 95.2% and 84.6%, respectively (Table 3). The stepwise

Table 2. Classification results of the 8 species of the *Bactrocera dorsalis* complex and *Bactrocera tau* and percent of "grouped" cases correctly classified: 89.6%.

Actual group	No. of cases	Predicted group membership								
		1	2	3	4	5	6	7	8	9
Group 1	60	56	0	0	0	1	0	3	0	0
B. arecae		**93.3%**	0.0%	0.0%	0.0%	1.7%	0.0%	5.0%	0.0%	0.0%
Group 2	50	0	44	0	0	3	3	0	0	0
B. dorsalis		0.0%	**88.0%**	0.0%	0.0%	6.0%	6.0%	0.0%	0.0%	0.0%
Group 3	53	0	0	52	0	0	0	0	1	0
B. propinqua		0.0%	0.0%	**98.1%**	0.0%	0.0%	0.0%	0.0%	1.9%	0.0%
Group 4	50	0	8	0	39	0	1	1	1	0
B. pyrifoliae		0.0%	16.0%	0.0%	**78.0%**	0.0%	2.0%	2.0%	2.0%	0.0%
Group 5	54	0	3	0	0	50	0	1	0	0
B. verbascifoliae		0.0%	5.6%	0.0%	0.0%	**92.6%**	0.0%	1.9%	0.0%	0.0%
Group 6	29	0	0	0	0	0	28	1	0	0
Species E		0.0%	0.0%	0.0%	0.0%	0.0%	**96.6%**	3.4%	0.0%	0.0%
Group 7	56	0	2	7	0	0	2	42	3	0
Species K		0.0%	3.6%	12.5%	0.0%	0.0%	0.0%	**75.0%**	5.4%	0.0%
Group 8	22	1	1	0	0	0	0	1	19	0
Species P		4.5%	4.5%	6.0%	6.0%	0.0%	6.0%	4.5%	**86.4%**	6.0%
Group 9	50	0	0	0	0	0	0	0	0	50
Bactrocera tau		0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	**100%**



1. Bactrocera arecae 2. B. dorsalis 3. *B. propinqua* 4. *B. pyrifoliae*
5. *B. verbascifoliae* 6. Species E 7. Species K. 8. Species P 9. *Bactrocera tau*

Table 3. Classification results in 4 groups of the *B. dorsalis* complex revealing percentage of "grouped" cases correctly classified as 98.9%, 98.1%, 95.2% and 84.6% for classes 1, 2, 3 and 4, respectively

Class	Actual group	No. of cases	Predicted group membership	
			1	6
	Group 1	60	59	1
I	*B. areacae*		**98.3%**	1.7%
	Group 6	29	0	29
	Species E		0.0%	**100.0%**
	Group	cases	2	5
	Group 2	50	49	1
II	*B. dorsalis*		**98.0%**	2.0%
	Group 5	54	1	53
	B. verbascifoliae		1.9%	**98.1%**
	Group	cases	3	4
	Group 3	53	53	0
III	*B. propinqua*		**100.0%**	0.0%
	Group 4	50	5	45
	B. pyrifoliae		10.0%	**90.0%**
	Group	cases	7	8
	Group 7	56	49	7
IV	Species K		**87.5%**	12.5%
	Group 8	22	5	17
	Species P		22.7%	**77.3%**

discriminant analysis procedure has certain advantages in reducing the use of a large number of variates to a small number of canonical variables. These selected variables were calculated as a ratio between variables, which were used in discriminant analysis for classification in each group afterwards. The stepwise discriminant analysis procedure was performed to take data correlation and to select variables for transformation in the next step. Finally, the best ratio was selected as an index for each species as shown in Table 4.

The results appear to be generally satisfactory in separation of these species of the *B. dorsalis* complex compared with the genetic data (Baimai *et al.* 1995, 1998 unpublished; Satayalai 1996) and classical taxonomy (Drew and Hancock 1994). Some adults were processed for electrophoretic study to confirm the genetic species as determined by mitotic chromosome markers. Adults from each collection were examined morphologically for species identification in correlation "with chromosomal evidence and electromorphic allozyme patterns. However, correct

Table 4. Index for classification of the 8 species of the *Bactrocera dorsalis* complex and *Bactrocera tau*. (SD) = standard deviation; + = 2 or more indices must be used for classification

Species	Index	Value	%accuraccy
Bactrocera tau	vein2/vein22+	2.82 (0.33)	
(out-group)	vein3/vein8	0.83 (0.06)	100.0%
B. dorsalis complex	vein2/vein22+	4.59 (0.48)	
	vein3/vein8	0.97 (0.05)	100.0%
	vein!8+	25.8 (2.5)	
	vein7/vein!4+	0.39 (0.03)	
B. arecae	veinl8/vein20	1.55(0.09)	98.9%
	vein!8+	31.0(1.7)	
	vein7/vein!4+	0.4 (0.01)	
Species E	vein!8/vein20	1.63(0.08)	98.9%
	veinS /vein 17+	0.83 (0.05)	
	vein8/vein!6+	1.75(0.07)	
B. dorsalis	vein8/vein29+	1.93 (0.26)	98.1%
	vein!3/vein24	1.15(0.04)	
	vein5/vein!7+	0.88 (0.05)	
	vein8/vein!6+	1.57(0.08)	
B. verbascifoliae	vein8/vein29+	1.85(0.26)	98.1%
	veinl3/vein24	1.22(0.04)	
	vein8+	91.0(7.2)	
	vein26+	10.28 (1.57)	
	vein7/vein!8+	3.09 (0.23)	
B. propinqua	vein20/vein24+	0.6 (0.05)	95.2%
	vein!3/vein20	2.02 (0.14)	
	vein8+	85.0 (7.9)	
	vein26	12.92(1.27)	
	vein7/veinl8+	3.03(0.12)	
B. pyrifoliae	vein20/vein24+	0.6 (0.04)	95.2%
	vein!3/vein20	2.03 (0.22)	
	vein 23/perimeter+	0.036 (0.002)	
	veinl8/vein21+	0.77 (0.06)	
Species K	veinl3/vein24	1.16(0.04)	84.6%
	vein23/perimeter+	0.034 (0.002)	
	vein!8/vein21+	0.63 (0.03)	
Species P	vein 13^6^4	1.21 (0.04)	84.6%

identification of some species is rather low, especially species K with a correct classification of only 75%, although this species is quite distinct in external morphology. In addition, "Cluster Membership of Cases using Average Linkage (Between Groups)" among the eight species of the *B. dorsalis* complex and *B. tau* showed some overlapping characters since there are mixed characteristics which lead to difficulty in classification (Fig. 3). Moreover, the dendrogram shows how the species could overlap as a result of genetic differentiation during the speciation processes (Ashlock 1979).

The methodology used for discrimination of members of the *B. dorsalis* complex proposed and employed in this study has some advantages over other tedious taxonomic techniques (e.g. cytotaxonomy and electrophoresis) for separation

of closely related species of insects. First, this method does not require fresh specimens. Second, it can be operated by a person who has a minimal knowledge of taxonomy or a non-taxonomist. Finally, the methodology described in this study seems to be promising for further development of on-line identification systems. It is clear that the data in the form of numerical tables can be easily stored and the computations can be rapidly made (Frampton *et al.* 1991). The methodology of morphometric analysis described here also illustrates the rapid advance in automated methods of on-line biological classification schemes which may have implications in the field of agricultural entomology, particularly in the tropical regions.

ACKNOWLEDGMENTS

We wish to thank S. Tigvatananont for providing most of the dried specimens of the fruit flies used in this study. We are grateful to Hollywood International Ltd. for providing access to use Nikon E2 Series (Nikon Digital Still Cameras). This work was supported by the Thailand Research Fund (RTA 3880008).

REFERENCES

ASHLOCK, P.O. 1979. An evolutionary systematist's view of classification. Systematic Zoology, 28: 441-50. BAIMAI, V., W. TRINACHARTVANIT, S. TIGVATTANANONY, PJ. GROTE, R. PORAMARCOM. and I). KIJCHALAO. 1995. Metaphase karyotypes of fruit flies of Thailand I. Five sibling species of the *Bactrocera dorsalis* complex. Genome, 38: 1015-1022.

DREW, R.A.I. 1989. The taxonomy and distribution of tropical and subtropical Dacinae (Diptera: Tephritidae). World Crops Pests (ed. By A.S. Robinson and G. Hooper). Amsterdam, Elsevier. p. 13-66.

DREW, R.A.I. and D.I. HANCOCK, 1994. The *Bactrocera dorsalis* complex of fruit flies (Diptera: Tephridae: Dacinae) in Asia. Bulletin of Entomological Research Supplement Series: Supplement, 2, 1-68.

FRAMPTON, E.R., J. FRY, B.P. STEPHENSON, and J.M. COWLEY, 1991. A Computerised system for data management during a fruit fly outbreak. The International Symposium on the Biology and Control of Fruit Flies.

GEORGE, D. and M. PAUL, 1995. SPSS/PC + Step by Step: A Simple Guide and Reference. CA: Duxbury Press. Belmont.

GONZALEZ, C.R. and E.R. WOODS, 1992. Digital Image Processing. Addison-Wesley Publishing Company. Inc.

HARDY, D.E. 1977. Family Tephritidae. Catalog of the Diptera of the Oriental Region (ed. By Delflnado M.D. and Hardy D.E.). Honolulu, The University Press of Hawaii. 854 p.

HOWELL, V.D., K. HOELMER, P. NORMAN and T. ALLEN, 1982. Computer-assisted measurement and identification of honey bees (Hymenoptera: Apidae). Annals of the Entomological Society of America, 75, 591-5944.

IBRAHIM, R. and G.A. IBRAHIM, 1990. Handbook on Identification of Fruit Flies in the Tropics. Universiti Pertanian Malaysia Press.

McPHERON, B.A. and G.J. STECK, 1996. The Fourth International Symposium on Fruit Flies of Economic Importance. A World Assessment of their Biology and Management. St. Lucie Press.

SATAYALAI, O. 1996. Electrophoretic study of natural populations of the *Bactrocera dorsalls* complex (Diptera: TephritidaeJ) in Thailand. Ph.D. Thesis, Bangkok. Faculty of Graduate Studies, Mahidol University.

SNEATH, P.H.A. and R.R. SOKAL. 1973. Numerical taxonomy: The principles and practice of numerical classification. W.H. Freeman and Company.

WEEKS, P. 1996. Daisy takes wing to sort out the bugs of the world. The Times. July 17.

WHITE, I.M. and M.M. ELSON-HARRIS, 1992. Fruit flies of economic significance: Their identification and bionomics. Wallingford, UK, CAB International.

WHITE, I.M. and P.R. SCOTT, 1994. Computerized information resources for pest identification: a review, p.129-137 *In:* Hawksworth, D.L. (ed.). The Identification and Characterisation of Pest Organisms. Wallingford, UK, CAB International.

Yu, D.S., E.G. KOKKO, J.R. BARRON, G.B. SCHAALJE, and B.E. GOWEN, 1992. Identification of ichneumonid wasps using image analysis of wings. Systematic Entomology, 17, 389-395.