Syst. Zool., 33(3):302–317, 1984

A COMPARISON OF FOURIER METHODS FOR THE DESCRIPTION OF WING SHAPE IN MOSQUITOES (DIPTERA: CULICIDAE)

F. JAMES ROHLF[1] AND JAMES W. ARCHIE[2]

[1]*Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794; and*
[2]*Department of Zoology, University of Hawaii, Honolulu, Hawaii 96822*

Abstract.—Outlines of wings from 127 species of North American mosquitoes were digitized. Comparisons were made among several different methods of reducing the information in the resulting coordinates to a series of descriptors that could be used in multivariate analysis. Methods included Fourier analysis of both radii and tangent angle change functions. In addition, the relatively new method of elliptic Fourier analysis was tried. Cluster and ordination analyses based on the various sets of descriptors summarized well the pattern of similarities and differences in wing shapes, but clusters of similar wings do not agree well with traditional taxonomic groupings. The use of elliptic Fourier descriptors appears to be especially promising for future work. [Fourier analysis; morphometrics; mosquitoes; image analysis; feature extraction.]

The present study is concerned with methods for quantitative analysis of similarities and differences in wing shape among 127 species of mosquitoes from America north of Mexico (see Fig. 1 for examples of the shapes of wings found among these species). There are many methods for quantifying outline shapes (a feature extraction problem in image analysis; Rohlf and Ferson, 1983) so that their variation can be analyzed using techniques of multivariate analysis. The conventional morphometric approach would be to define a series of characters—such as length to width ratios, and more complex ad hoc indices—to describe the ways in which wings differ in shape. This is less likely to be successful in cases like the present one where the diversity of shapes is not large and, hence, the morphometric trends are apt to be rather subtle. We are interested here in methods that allow one to capture the entire shape outline in a systematic manner and with a desired degree of precision. We have investigated the relative usefulness of five different feature extraction methods based on Fourier analyses of various functions of the coordinates of points around the outline of each wing.

Kaesler and Waters' (1972) analysis was one of the first applications of Fourier descriptors to study morphological shapes in systematics. Since that time there have been many such applications (some of them are listed in Rohlf and Ferson, 1983). The present study represents an attempt to define and evaluate new types of taxonomic characters that can be recorded automatically using optical scanners on digital computers and, thus, make larger quantitative studies more feasible. Methods based on Fourier coefficients are not the only ones that can be used. Another approach is to use moment invariants (Hu, 1962). In this method an image is treated as if it were a bivariate density function (image darkness treated as proportional to probability density) and then the x and y moments are computed. Functions of these moments are computed that are invariant to differences in image location, size, rotation, reflection, and contrast. One limitation of this approach is that one cannot easily reconstruct an image from the descriptors. Empirical evaluations of their usefulness in biological image analysis will be given by Ferson et al. (in prep.) and Rohlf and Ferson (in prep.). Rohlf and Sokal (1967) described another method in which the presence or absence of a visible feature at a series of randomly or system-

atically selected locations were used as characters. This method requires careful alignment of images and does not permit an accurate reconstruction of the image. While this method seems to "work," it is not very satisfying intellectually.

There have been few numerical taxonomic studies of mosquitoes. Rohlf (1963) studied 48 species of *Aedes* based upon characters from both the adults and larvae, while Steward (1968) reported on a similar analysis using 82 characters for 42 Canadian species of this genus. Hendrickson and Sokal (1968) studied 29 species from the genera *Psorophora* and *Aedes* based on 158 adult morphological characters. Rohlf (1967) reported on a classification of 45 species in several genera using only pupal characters. The images of these same pupae were also studied in Rohlf and Sokal (1967). Rohlf (1977) analyzed data on 63 species of *Aedes* using 14 adult thoracic setal characters. Moss et al. (1979) studied 25 species of the genus *Toxorhynchites* from the Orient using 79 adult and 26 pupal characters. A follow-up study (which included a numerical cladistic analysis) by Simon et al. (1982) included 32 species of *Toxorhynchites* based on 100 characters from the immature stages. While general agreement has been found with previous classifications, there have also been a number of important differences (Crovello, 1969; Nielsen, 1969). Much further work is clearly needed both at the higher and lower taxonomic levels.

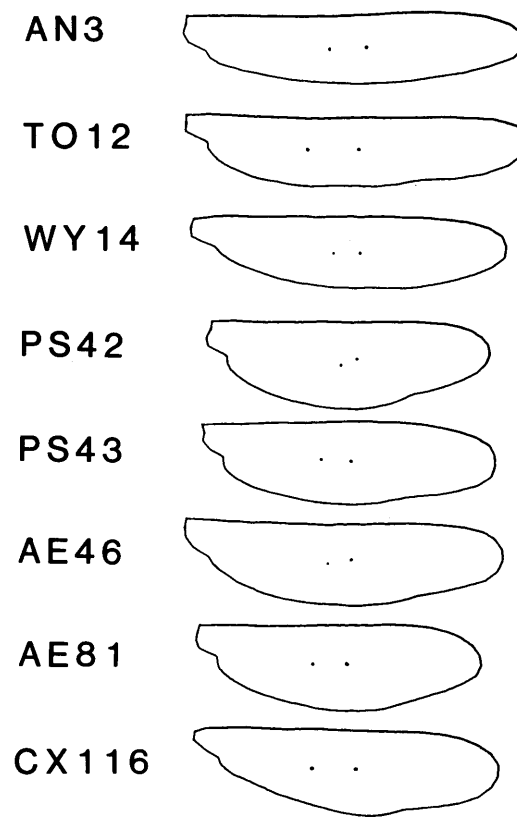

FIG. 1. Examples of mosquito wing shapes based on the *RP* data set. The central "dot" corresponds to the centroid of each wing and the "dot" to its left corresponds to the point at which the fifth longitudinal vein branches. Codes explained in Table 1.

MATERIALS AND METHODS

Since this study is concerned with the possible usefulness of the wing outline to determine broad relationships within the family as a whole (rather than to distinguish between similar species), the data were obtained from plates 1 to 127 in Carpenter and LaCasse (1955). These correspond to the 127 species of mosquitoes from America north of Mexico included in that volume. The use of these published illustrations made it possible for this study to be comprehensive without the complications of mixing images from different authors or from microscope slides prepared using different procedures. (Studies concerned with the problem of distinguishing between pairs of similar species are underway using actual wings from many replicate specimens digitized automatically from TV images. These will be reported separately.) In various diagrams in this paper, the species are identified by a two-letter designation for each genus (see Table 1) followed by a number that corresponds to the plate number in Carpenter and LaCasse (1955).

The original data were obtained using a coordinate digitizer connected to an HP9830 desk-top computer that was controlled by a program written in BASIC. The digitizer is accurate to 0.01 inches with a

TABLE 1. List of genera with codes and species code numbers.

Genus	Code	Numbers
Anopheles	AN	1-11
Toxorhynchites	TO	12
Wyeomyia	WY	13-16
Uranotaenia	UR	17-20
Culiseta	CA	21-27
Orthopodomyia	OR	28
Mansonia	MA	29-31
Psorophora	PS	32-43
Aedes	AE	44-100
Culex	CX	101-125
Deinocerites	DE	126-127

field of 17 by 17 inches (effectively a 1,700 by 1,700 pixel grid). The outline of a right wing of each species was traced using a cursor and the x- and y-coordinates were read "continuously" by the computer (i.e., the digitizer furnished coordinates as rapidly as the computer could request them). The wings were oriented more or less horizontally and the point at which the fifth longitudinal vein branches was taken as the origin (this landmark was used, since it is centrally located on most wings). These initial coordinates were converted to polar-coordinate form as the points were digitized. Angles were measured in a clockwise direction starting from the point at which the leading edge of the wing joins the thorax. In order to eliminate the effect of unequal sampling due to the variation in speed at which different parts of the wing were traced, coordinates were computed for the end points of 100 equally-spaced radii (trigonometric interpolation was used to give equal angles). Thus, the basic data set was obtained in radius-function form (it consisted of 100 distances from the branching point of the fifth horizontal vein to the edge of the wing, every 3.6°) for each of 127 species of mosquitoes (see Fig. 2A). This was repeated five times for each wing (to reduce operator error in tracing) and the averaged lengths of the radii were transmitted to the campus' UNIVAC 11/82 computer and the department's MODCOMP IV/25 and IV/35 minicomputers for further processing. These radii (see Fig. 2B) were analyzed by Fourier analysis. The coefficients for the first 15 harmonics were computed for each of the 127 wings and used as descriptors. Since the "zeroth" harmonic is just a function of the size of the wing, all coefficients were scaled so that the zeroth harmonic was equal to 1.0. The resulting matrix of descriptors for harmonics 1 through 15 is referred to below as the "*RP*" (raw polar) data set.

A second data set was constructed mathematically from the first by translating the origin of the system of polar coordinates to the centroid (center of gravity) of each wing (see Fig. 2C). The location of the centroid was determined by a numerical integration. Anstey and Delmet (1973) indicated that this choice of an origin was preferable to that of a morphological landmark since its use increases the relative amount of information in the harmonics above the first. Actually, the use of the centroid simply ensures that the results are just a function of the wing shape and not of changes in the location of the landmark within a wing (this simple adjustment seems to have been overlooked by Roberts et al., 1983:379). This proved to be important since species were found that had very similar wing outlines but differed in the location of the branching point of the fifth horizontal vein. A new set of points were computed by interpolation so that the radii would be equally-spaced (every 3.6° around the new origin).

A matrix of Fourier coefficients for harmonics 0 through 15 was calculated from these data and is denoted below as the "*CP*" (centered polar) data set. One of the major advantages of these various types of Fourier coefficients as descriptors is that one can reconstruct the outlines of the wings both from the various data matrices and from cluster's centroids and points along principal component axes. This is useful as a check on the numerical results (and as an aid to interpretation). Figure 3A shows the separate contribution of harmonics 0 through 8 to the average mosquito wing (the wing one obtains if one uses the average Fourier coefficients based

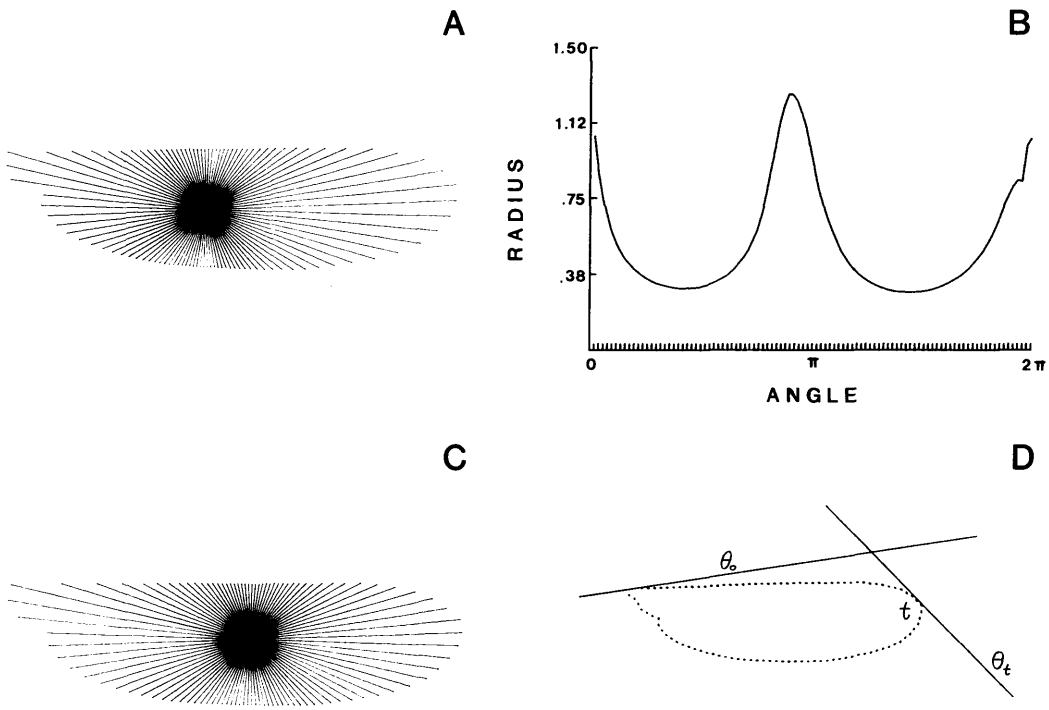

FIG. 2. Representations of the coordinate systems used for the data sets in the present study. (A) Polar coordinates relative to a landmark. (B) The radius from *A* shown as a function of the angle. (C) Polar coordinates relative to the centroid of the wing. (D) Angle of a tangent vector as a function of distance along the perimeter of the wing. The elliptic Fourier coefficients are based simply on the *x*- and *y*-coordinates of points along the wing outline and are not shown.

on all 127 species), and Figure 3B provides the successive reconstructions of the wing outline based on harmonics 0 through 8 and 15. Since the zeroth harmonic is again just a function of the size of the wing, all coefficients were scaled by dividing them by the square root of the area of the wing.

While it is not important when considering wing outlines (which are simple convex images), the two methods described above are limited to images in which the radius functions are single-valued. These methods cannot be used for shapes that have many concave curves (such as maple leaves). Zahn and Roskies (1972) suggested a more general way to describe the shape of an object based on the cumulative change in angle of a vector tangent to the outline of the object as a function of distance around the periphery of the wing (see Fig. 2D). The wing is first scaled so that its perimeter is 2π. Then the function

$$\phi(t) = \theta(t) - \theta(0) - t \qquad (1)$$

is computed for 100 equally-spaced values of t. In this function, t is the distance along the periphery of the wing from the starting point (ranges from 0 to 2π radians), $\theta(0)$ is the angle of a tangent at the starting point, and $\theta(t)$ is the angle of a tangent vector at a distance t from the starting point. Bookstein et al. (1982) referred to this as an intrinsic representation. The third, "*TA*" (tangent angle), data set consists of the coefficients for harmonics 0 through 15 based on a Fourier analysis of the function $\phi(t)$ described above. Figure 4 shows the contributions of the first eight harmonics to the description of the aver-

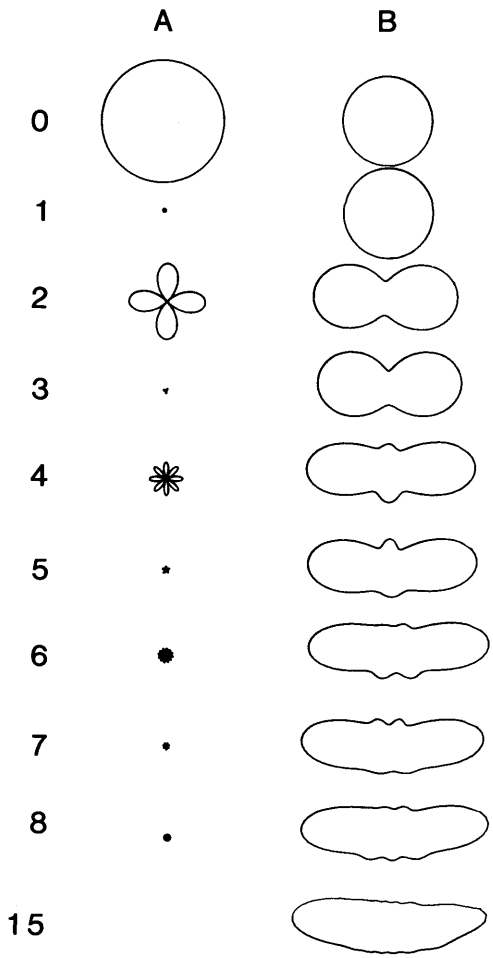

FIG. 3. Fourier harmonics for the average mosquito wing based on polar coordinates relative to the centroid of the wing. (A) Contributions of the first eight harmonics. The relative sizes of the figures reflect the magnitudes of their individual contributions. (B) Reconstructions of the wing outline based on the cumulative contributions of the first eight harmonics and for the first 15 harmonics.

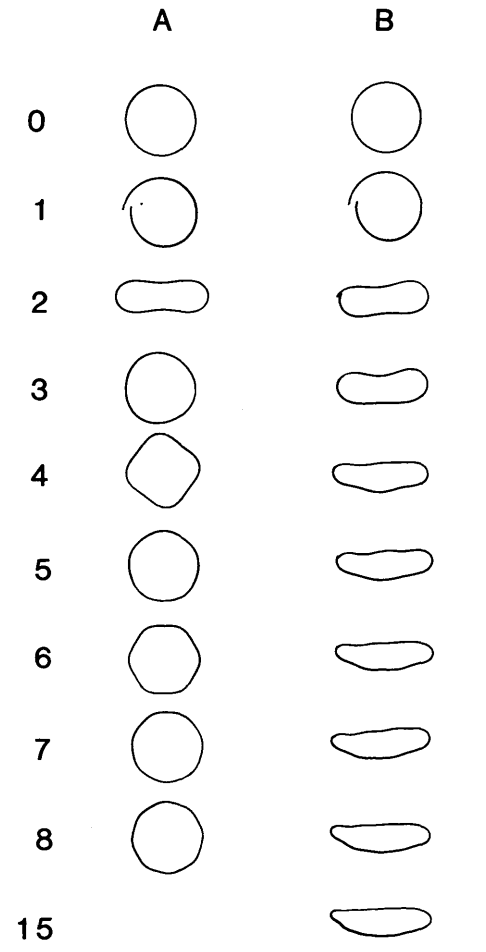

FIG. 4. Fourier harmonics for the average mosquito wing based on angle of a tangent vector as a function of distance along the perimeter of the wing. (A) Contributions of the first eight harmonics. Note that the contribution for the first harmonic does not result in a closed curve. (B) Reconstructions of the wing outline based on the cumulative contributions of the first eight harmonics and for the first 15 harmonics. The curves do not close due to the effects of the first harmonic.

age wing (defined as above) and its reconstruction using the sums of the contributions of various harmonics.

The Fourier analyses applied above can be thought of simply as multiple regression analyses in which the best fitting values (in a least squares sense) of the coefficients a_i and b_i are found in the following trigonometric regression equation,

$$Y = a_0 + \sum_{i=1}^{k} (a_i \cos[it] + b_i \sin[it]). \quad (2)$$

In this equation the angle t varies from 0 to 2π. The zeroth harmonic describes the contribution of a centered circle, the first harmonic an offset circle, the second a

"figure 8," the third a trefoil, the fourth a quatrefoil, etc. (see Fig. 3A). The contributions are less obvious for the *TA* data set (see Fig. 4A), since the function being fitted, $\phi(t)$, is not just a simple radius. The coefficients can be computed by a variety of computational algorithms. The algorithm given in Ralston (1965) was used for data sets *RP*, *CP*, and *TA* described above (this algorithm takes advantage of the fact that the data points were equally-spaced). Since there are relatively few data points, there is no great advantage in using the fast Fourier transformation type algorithms for any of the data sets we used.

The fourth, fifth, and sixth data sets were constructed using the x- and y-coordinates (rectilinear) of each data point in the original basic data matrix. The *"EF"* (elliptic Fourier) data matrix consists of coefficients of the elliptic Fourier series of the generalized chain-encoded outline of a figure as suggested by Kuhl and Giardina (1982). Their algorithm does not require the points to be spaced equally and can fit an arbitrary closed contour—given enough harmonics. The coefficients of the nth harmonic in a Fourier series expansion for the x-projections of an outline are

$$A_n = \frac{T}{2n^2\pi^2}\sum_{p=2}^{k}\frac{\Delta x_p}{\Delta t_p}\cdot(\cos[2n\pi t_p/T] - \cos[2n\pi t_p - 1/T]) \quad (3)$$

and

$$B_n = \frac{T}{2n^2\pi^2}\sum_{p=2}^{k}\frac{\Delta x_p}{\Delta t_p}\cdot(\sin[2n\pi t_p/T] - \sin[2n\pi t_p - 1/T]), \quad (4)$$

where k is the number of steps in the trace around the outline (indexed by p), and Δx_p is the displacement along the x-axis of the contour between steps $p - 1$ and p. Δt_p is the length of the linear segment between these steps, t_p is the accumulated length of such segments, and $T = t_k$ is the total length of the contour as approximated by the trace polygon. The coefficients for the y-coordinates, C_n and D_n, are found in the same way using the incremental changes in the y-direction. The resulting coefficients are not dependent on the x or y displacement of the wing, and the original wings were aligned so as to have the same orientation and starting point for the trace of their outline.

A related method called "dual-axis Fourier shape analysis," DAFSA, has been proposed by Moellering and Rayner (1981, 1982). As presented by them, this technique requires that the outline must be a closed plane curve that is sampled at uniform arc-length intervals. The x- and y-coordinates are treated as pairs of complex numbers $(x + iy)$ expressed as a function of the arc-length distance of each point from an arbitrary reference point. This function is then fitted by Fourier analysis. Kincaid and Schneider (1983) seem to have used the same method.

The *"EF0"* data matrix is the *"EF"* matrix with the terms for the zeroth harmonics deleted to remove the effects of these terms (the starting position for the digitization trace along the outline). Figure 5 shows the individual and the cumulative contributions of harmonics one through eight to the reconstruction of the average mosquito wing.

The coefficients for the *"EF"* data set can be mathematically adjusted (i.e., normalized) to be invariant to size, rotation, and starting position of the outline trace. The use of the following matrix transformation yields the *"NEF"* (normalized elliptic Fourier) data set.

$$\begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} = \frac{1}{E^*}\begin{pmatrix} \cos\psi & \sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \cdot \begin{pmatrix} A_n & B_n \\ C_n & D_n \end{pmatrix} \cdot \begin{pmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{pmatrix}, \quad (5)$$

where n is the harmonic order of the four coefficients (a, b, c, and d), and

$$E^* = (a^{*2} + c^{*2})^{0.5} \quad (6)$$

is the magnitude of the semimajor axis of the best-fitting ellipse. The angle of rotation (between 0 and 2π radians) of this ellipse is

$$\psi = \arctan(c^*/a^*), \quad (7)$$

and the angle of rotation (from 0 to π radians) of the starting point from the new starting point at the end of the ellipse (standardized arbitrarily to one end for a given set of images) is

$$\theta = 0.5 \arctan\left([2A_1B_1 + C_1D_1]/[A_1^2 + C_1^2 - B_1^2 - D_1^2]\right). \quad (8)$$

The values of a^* and c^* are given by the matrix equation

$$\begin{pmatrix} a^* \\ c^* \end{pmatrix} = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix} \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}. \quad (9)$$

After this transformation, three coefficients are constant ($a_1 = 1$, $b_1 = 0$, $c_1 = 0$) and can be ignored. The last coefficient for the first harmonic, d_1, gives the eccentricity of the ellipse.

Each of the matrices of Fourier coefficients were then used as input to several methods of multivariate analysis. The effect of standardization of the characters (coefficients) was investigated by using both standardized and unstandardized data (the variances among species for each harmonic are shown in Fig. 6). The use of unstandardized data is of interest here since the original coordinates were scaled so that the wings all had the same area and the coefficients of the various harmonics were all in comparable units. An unweighted pair group (UPGMA) cluster analysis based on a matrix of average taxonomic distances among the 127 species of mosquitoes was performed and cophenetic correlation coefficients were computed as measures of goodness of fit (for a general account, see Sneath and Sokal, 1973). Next a principal components analysis (PCA) was performed on the variance—covariance matrix for the characters and the 127 species were then projected onto the resulting axes. The configuration of points obtained from the data sets were compared visually and by Gower's (1971) M^2 criterion which measures the sum of the squared distances between corresponding points after one PCA matrix has been rotated to superimpose as well as possible on another matrix. Smaller M^2 values correspond to better agreement between two PCA solutions.

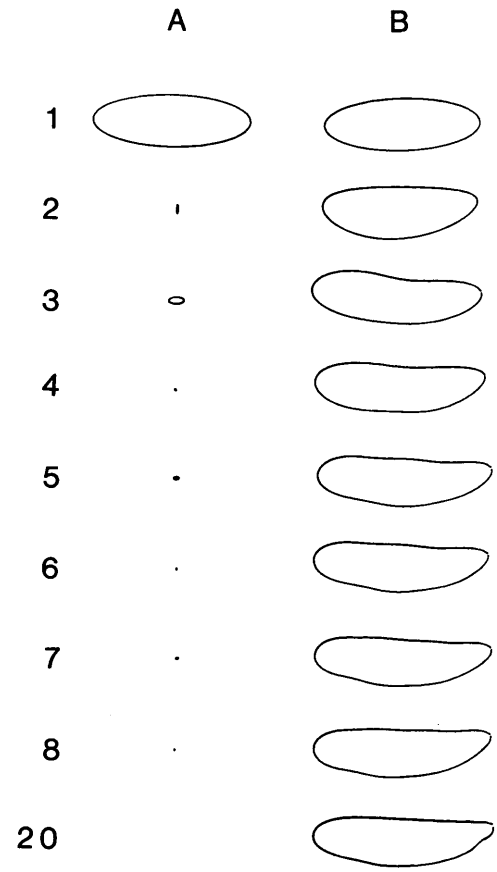

FIG. 5. Elliptic Fourier harmonics for the average mosquito wing based on x- and y-coordinates of the wing. (A) Contributions of harmonics one through eight. (B) Reconstructions of the wing outline based on the cumulative contributions of the first 8 harmonics and for the first 20 harmonics.

RESULTS

The effect of standardization was examined first. The cophenetic correlation

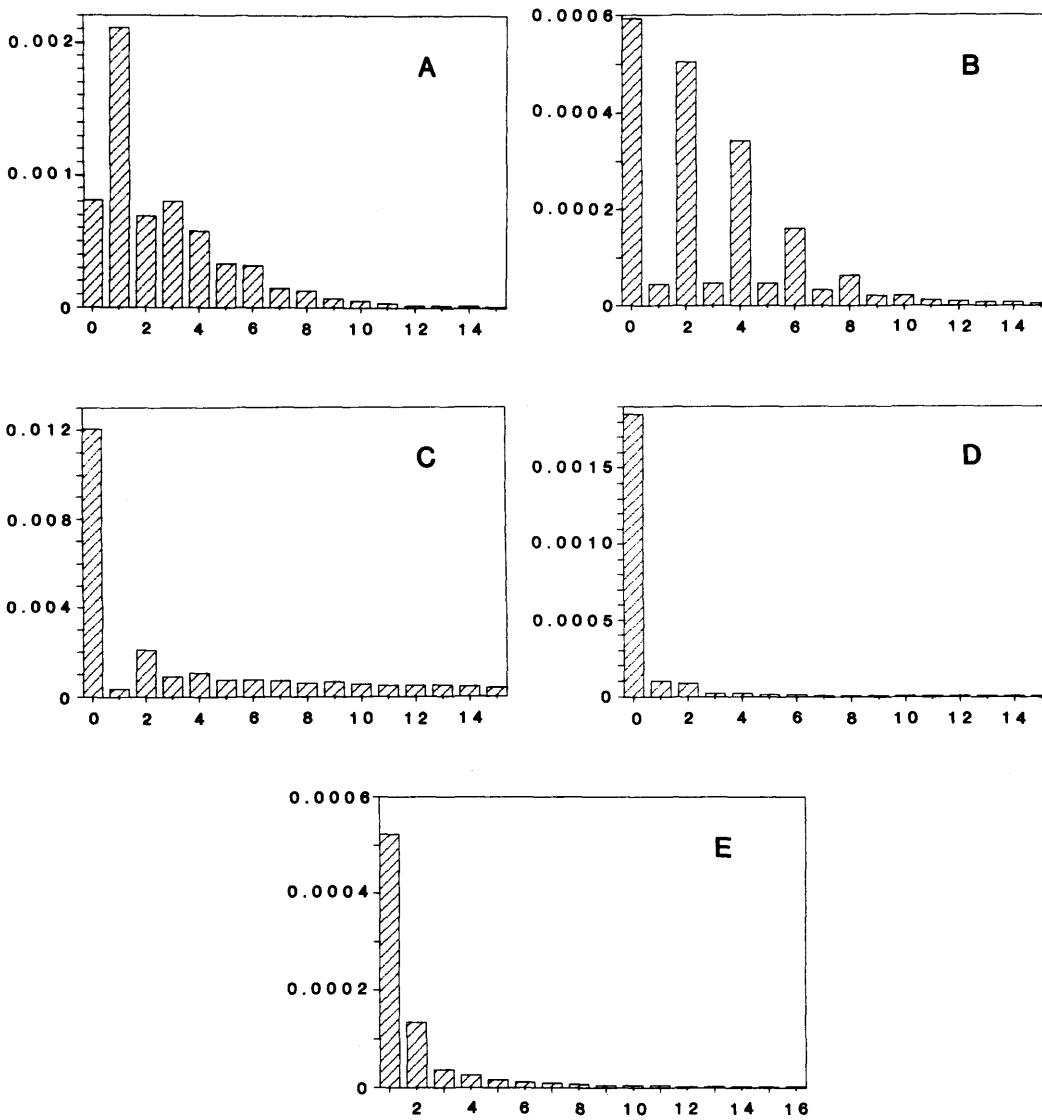

FIG. 6. Plot of the variance among wings of the 127 species of mosquitoes for the harmonics used in this study. The variances for each coefficient have been summed to give a total contribution for each harmonic. Variances for the: (A) *RP* data set; (B) *CP* data set; (C) *TA* data set; (D) *EF* data set; and (E) *NEF* data set.

coefficients from the UPGMA cluster analyses were similar whether or not standardization of the variables was performed. The average correlation for unstandardized data was 0.705 versus 0.691 for standardized data. The average matrix correlation between the distance matrices and distances based upon three-dimensional principal component analyses (see Rohlf, 1972) was higher for unstandardized data than for standardized data (0.991 versus 0.909). The percentage of variance explained by the first three principal component axes was also higher for unstandardized data than it was for standardized data. Standardization increases the rela-

TABLE 2. Correlations among taxonomic distance matrices for data sets based upon unstandardized data.

Data set[a]	1 *RP*	2 *CP*	3 *TA*	4 *EF*	5 *EF0*
2 *CP*	0.592				
3 *TA*	0.180	0.310			
4 *EF*	0.561	0.798	0.297		
5 *EF0*	0.560	0.798	0.289	1.000	
6 *NEF*	0.469	0.827	0.296	0.907	0.907

[a] Data set codes: *RP*, raw polar; *CP*, centroid translated polar; *TA*, tangent angle; *EF*, elliptic Fourier; *EF0*, *EF* with zeroth harmonic left out; and *NEF*, normalized elliptic Fourier.

tive contribution of harmonics that do not explain much of the variance (e.g., the odd harmonics in Fig. 3). This increases the effective dimensionality of the space so that the first three axes explain a smaller proportion of the now larger total variance.

The relationships indicated by unstandardized data seem to correspond more to one's subjective visual impression of similarity. This can be seen by examining particular pairs of species for which very different relationships are indicated when the data are standardized. For example, based on unstandardized data from the *RP* matrix, species PS42 and AE81 are indicated to be very similar to each other but rather different from the other species. When the matrix was standardized, species PS42 is indicated to be very different from all other species. If one superimposes the wing outlines for these two species one finds that there is a general overall agreement in the outlines. There are, however, some small differences (that would result in different values for the higher order harmonics). Species AN3, AN11, and AE55 present another interesting example of the effect of standardization. Based on standardized data AN11 and AE55 are indicated to be much more similar to each other than either is to AN3. Using unstandardized data one finds that AN3 and AE55 now are more similar than either is to AN11. These latter similarities agree closely with the visual impression one obtains by looking at the actual wing outlines.

The explanation for the unsatisfactory effects of standardization appear to be just that it gives equal weight to all harmonics and, thus, results in much greater emphasis on differences among species in the higher order harmonics (which are much more sensitive to small irregularities in the outlines and to measurement errors). This is clearly seen in the results of principal components analyses where many more of the higher order harmonics had high loadings on the first few axes. Therefore, unstandardized data and variance-covariance matrices were used in all analyses reported below.

The overall degree of similarity across the different data sets for the relationships among the 127 species is shown in Table 2. This table gives matrix correlations between pairs of average taxonomic distance matrices from the various data sets. The matrices all are positively correlated, indicating some degree of general agreement. Data set *TA* has much lower correlations with the other data sets. Leaving out the zeroth elliptical harmonic apparently had little effect in this study since *EF* and *EF0* had a correlation of 1.000 (the two matrices are not, however, identical). The *NEF* distance matrix is highly correlated with those of the *EF* and *EF0* data sets. Data sets *CP* and *EF* indicate quite similar relationships among the 127 species since the correlation between their distance matrices was 0.798. The similarity among these data sets is a measure of the consistency with which the original wings were manually aligned.

Table 3 gives some comparisons among the results of UPGMA cluster analysis applied to the various data sets. The *TA* data set resulted in the largest cophenetic correlation (due in part to the fact that one species was indicated to be isolated from the other species). Pairs of phenograms were compared by computing the CI_C consensus index from a strict consensus tree (Rohlf, 1982). This very stringent criterion indicates that the phenograms are by no means identical for the different data sets. The largest value, 0.752, indicates, for example, that the *EF* and the *EF0* phenograms have 75.2% of their clusters represented identically in both trees. The lowest values are for comparisons of data set *TA*

TABLE 3. Comparisons among UPGMA phenograms based on distance coefficients from unstandardized data. Coefficients are the CI_c consensus coefficients based on strict consensus trees. Cophenetic correlation coefficients are also given for each data set. Larger values correspond to closer agreement.

Data set	Data set					Cophenetic correlation
	1 *RP*	2 *CP*	3 *TA*	4 *EF*	5 *EF0*	
1 *RP*						0.636
2 *CP*	0.048					0.720
3 *TA*	0.088	0.008				0.786
4 *EF*	0.088	0.080	0.016			0.704
5 *EF0*	0.088	0.088	0.024	0.752		0.690
6 *NEF*	0.016	0.096	0.016	0.128	0.128	0.708

with the others. Figure 7 shows the phenograms from the *CP* and *NEF* data set (the others are not furnished in order to save space). While similar wings usually cluster together, the clusters so found do not correspond well to conventional taxonomic groupings.

Table 4 summarizes the results of comparisons of pairs of three-dimensional principal components solutions for the various data sets. The values given are Gower's (1971) M^2 coefficient. Again, data sets *EF* and *EF0* agree closely and data set *TA* gives the least agreement with the others. The percentages of variance explained, matrix correlations with the original distance matrices and between the principal component solutions are also furnished in Table 4.

Figure 8 shows the results of the principal components analysis for the *CP* and the *NEF* data sets. In order to properly appreciate the relationship among the points, it is essential that the first and second axes are plotted to the same scale (this important point is often overlooked, even in recent papers). It is apparent from these figures that, while there is little evidence of taxonomic structure in the data, similarly-shaped wings are plotted close together in the figures. There are also consistent changes in the shapes of the wings as one goes from left to right and from top to bottom across the figures. These trends can best be visualized by reconstructing wings corresponding to points exactly along the principal component axes. Figure 9 depicts wings that would be plus or minus three standard deviations from the mean along axes I and II and zero along all other axes. This shows very clearly that axis I has narrower wings at the right and wider at the left. Axis II is harder to describe but it clearly contrasts wings that are wider near the base versus near the apex. This technique is potentially a very powerful tool since it allows one to visualize what an organism would look like if it were to be located in a particular region of a principal component (or even a canonical variate) space, even if no observed points fell into that particular region. The ordination for the *TA* data set (not shown) is more similar to the results given by the other data sets than is suggested by the values given in Table 4. Most of the differences seemed to be due to the fact that species CX116 is indicated to be an extreme outlier in the plot for the *TA* data set, but not for the other data sets.

DISCUSSION

Bookstein et al. (1982) pointed out a number of limitations on the biological interpretations one can give to the coefficients of Fourier functions (but see also the rebuttal by Ehrlich et al., 1983). Bookstein et al. pointed out, for example, that the methods used for the *RP* and *CP* data sets are sensitive only to differences in shape not to differences in interpretation of homology between radii at different points along an outline (however, if one's goal is to measure shape per se, this can be considered an advantage). Another limitation is that a change in part of the shape (a "local change") may result in changes in the values of many of the coef-

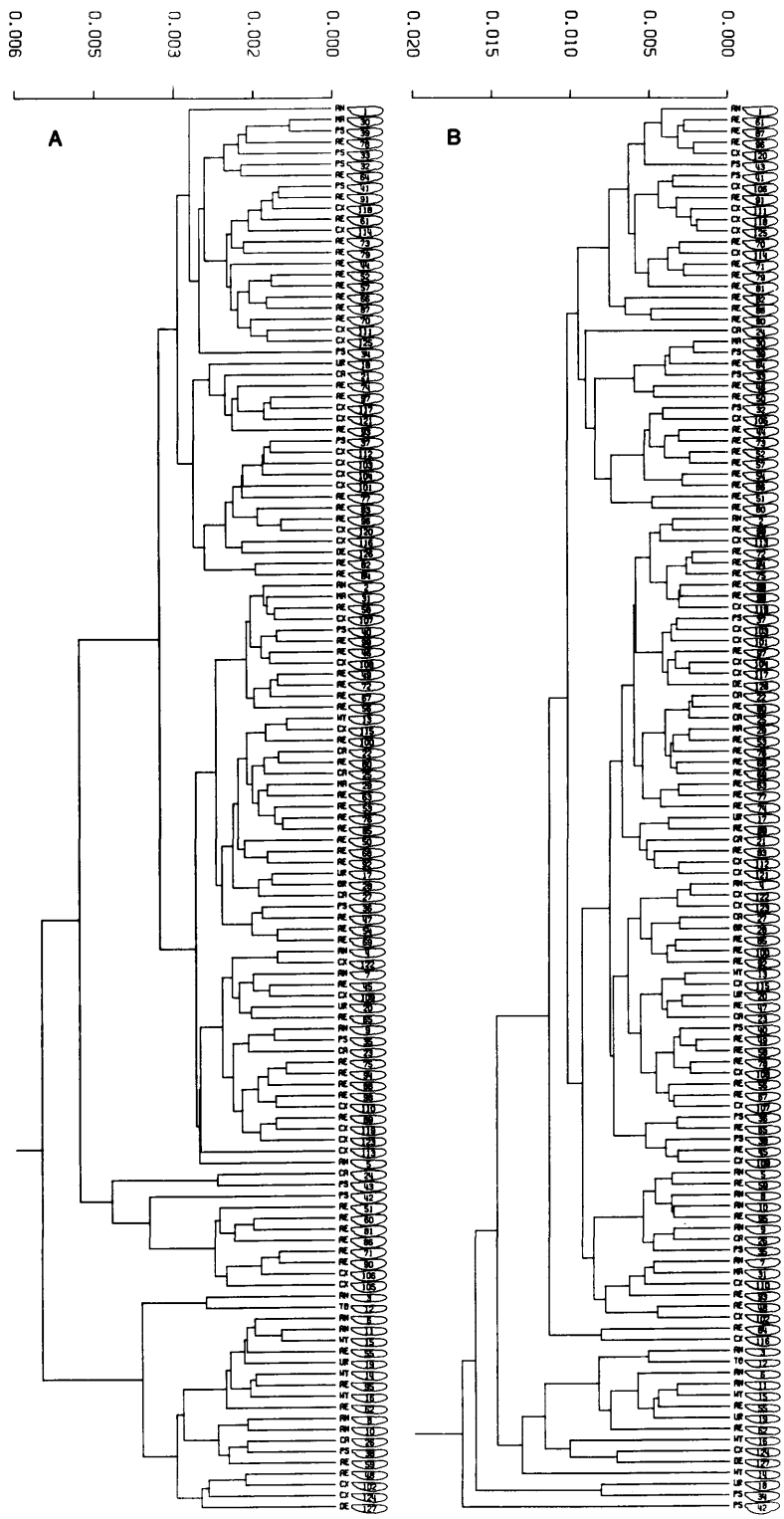

TABLE 4. Comparisons of projections of the 127 species onto the first three principal component axes based upon unstandardized data. Gower's M^2 coefficient is given for each pair of data sets compared. The percentages of variance explained and matrix correlations are also provided for each data set.

Data set	Data set					Percent explained	Matrix correlation
	1 *RP*	2 *CP*	3 *TA*	4 *EF*	5 *EF0*		
1 *RP*						85.33	0.984
2 *CP*	0.692					88.25	0.990
3 *TA*	1.460	0.989				79.32	0.985
4 *EF*	0.706	0.469	0.968			92.04	0.998
5 *EF0*	0.706	0.469	0.968	0.000		92.15	0.996
6 *NEF*	0.878	0.466	0.975	0.288	0.288	88.60	0.993

ficients (not just in coefficients corresponding to the high frequency components as one might have hoped). This makes interpretation of the coefficients more difficult. However, these problems are not unique to Fourier methods. The coefficients of linear, quadratic, cubic, etc., polynomials fitted to some data need not be biologically interpretable if the underlying biological process that generated the relationship under study corresponds to some other functional relationship (such as exponential or sinusoidal). This is an inherent problem in curve fitting when one fits functions that do not correspond to the actual underlying biological processes. Traditional morphometric approaches have an analogous problem since the usual arbitrary suites of morphometric measurements need not correspond directly to the underlying developmental, evolutionary factors either. Principal components and canonical variates analyses have the same type of problem in interpretation. The most important limitation of our approach is that we deal only with overall shape—not with changes in distances between homologous points. Bookstein (1982) considered the latter to be of more fundamental importance in morphometrics. Work in progress by one of us (FJR) uses information on the shape and position of each wing vein. The branching points of the veins and their points of intersection with the margin of the wing will provide homologous reference points.

The fact that the individual Fourier coefficients will almost never be morphologically meaningful might actually be considered an advantage—since it may help an investigator to resist the temptation to create complex interpretations of the meaning of the individual coefficients (Rohlf and Ferson, 1983). We would not expect, for example, that a relatively high contribution of the second spherical harmonic in the average mosquito wing (see Fig. 3) would lead an investigator to conclude that there was a growth field that had the form of a quatrefoil. The investigator is thereby forced to deal with overall shape differences by using suites of coefficients.

After comparing several methods, the question as to which method is "best" naturally arises. All the methods are mathematically equivalent in that given enough harmonics they can encode the shape of the wing outlines exactly. For objects with more complex outlines there may not be a single "center" that enables the radius function to be single-valued. This eliminates the two polar methods from consideration as general methods (even though they may work well for many shapes such as those of insect wings). The method

←

FIG. 7. UPGMA phenograms based on average taxonomic distance from unstandardized Fourier coefficients. (A) *CP* data set. (B) *NEF* data set.

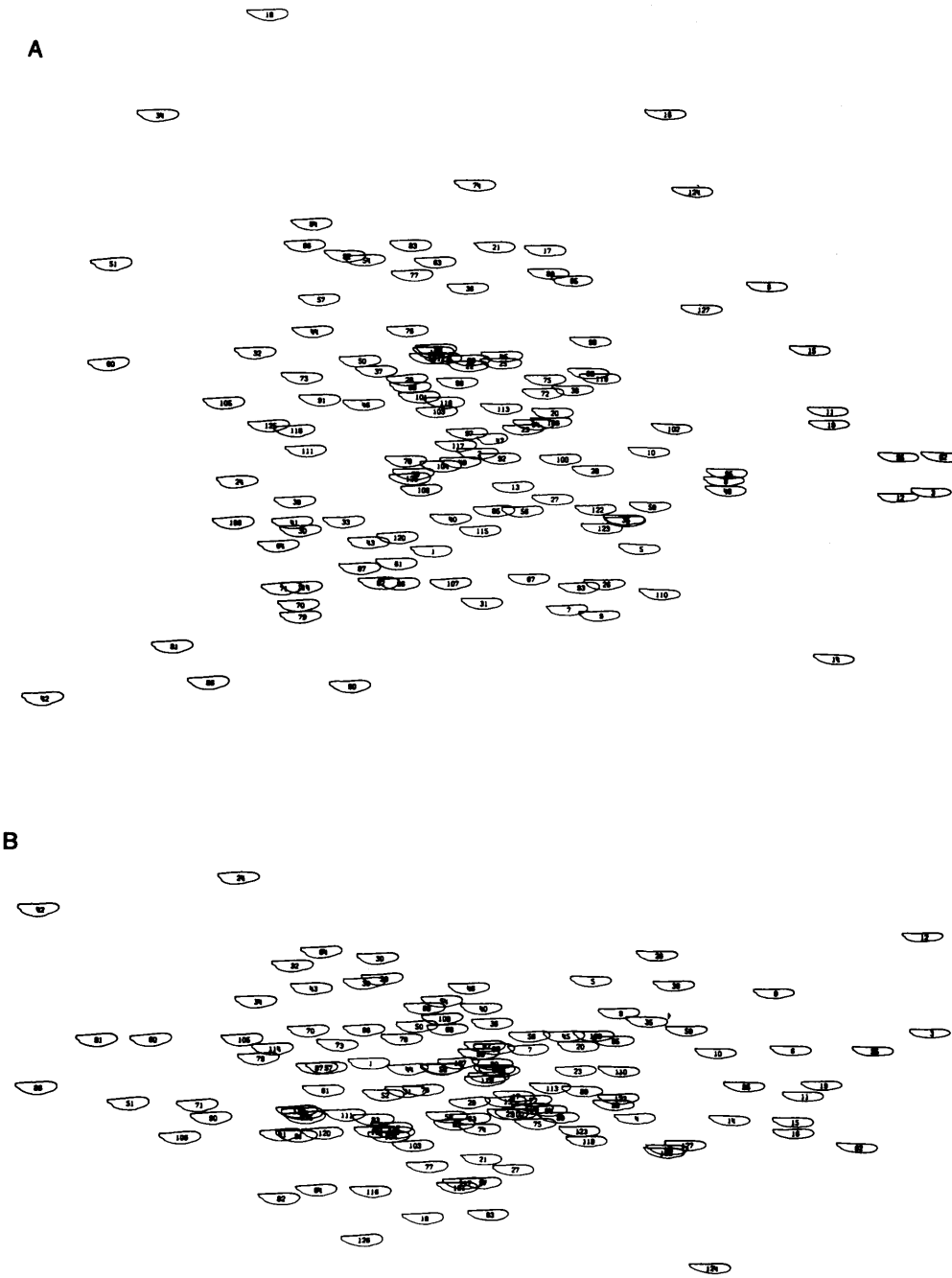

FIG. 8. Plots of the first two dimensions from principal components analyses of unstandardized Fourier coefficients. (A) *CP* data set. (B) *NEF* data set.

based on tangent angles has two serious problems. The first is that, as pointed out by Zahn and Roskies (1972), reconstructions with relatively few harmonics do not usually result in figures with closed contours (see Fig. 4). This makes it more difficult to interpret trends in reconstructed images (such as shown in Fig. 9 for the *CP* and *NEF* methods) since they may not look realistic. The second problem is that the estimation of the coefficients seems to be more sensitive to "noise" in the image than the other methods (as mentioned above species CX116 was considered to be an outlier only by this method). This leaves the methods based on elliptic Fourier coefficients as the most generally useful. The use of the normalized coefficients, *NEF*, is a convenience since it saves the investigator from having to worry about aligning the images in a standard fashion. The only disadvantage is that a small amount of additional computation is required (which may be a problem if the analyses are being performed on a small microcomputer without floating-point hardware).

Our original (somewhat naive) hope was that an analysis of wing shape might yield classifications comparable to conventional taxonomic classification, but even in the absence of such congruence we feel that the present application is quite successful. The ordinations enabled us to summarize and display the diversity of shapes of wings found among the North American Culicidae—even though the wings are fairly similar, and some of the differences rather subtle and difficult to describe by conventional means.

The differences in wing shape must have functional and ecological significance, but we did not find any obvious associations with published accounts indicating that a given species was a "strong" or "weak" flier, was found in exposed or sheltered habitats, etc. The detection of such associations may require additional quantitative information about the behavior and physiology of the various species of mosquitoes.

These results are in contrast to the study

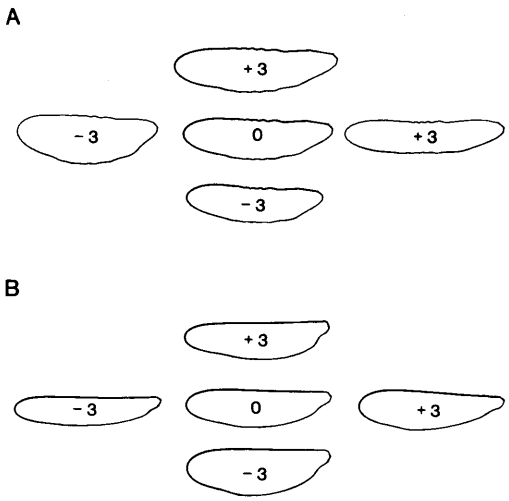

FIG. 9. Reconstructed wing outlines for points along the principal component axes shown in Figure 8. (A) *CP* data set. (B) *NEF* data set.

by Plowright and Stephen (1973) who found good agreement between their morphometric analysis of bee wings and the conventional classification. Their characters described features of the venation rather than the outline shape of the wings as in the present study. Brown (1979) and Brown and Shipp (1977, 1978) also used wing vein lengths to examine phenetic relationships within two groups of Australian Diptera and found that the results varied in their consistency with results from previous studies. The previous studies had used standard taxonomic characters including genitalic characters as well as geographic distribution for erecting species groups. Simon (1983) used a series of 49 characters involving wing vein length and was able to establish differences among year classes (broods) and life cycle forms of periodical cicadas (*Magicicada*) where no differences had previously been established. Whether differences in wing outline shape will provide useful taxonomic information when more diverse wing shapes are analyzed remains to be seen. Minor wing shape differences may also be expected between closely related

species or populations if these differences are important for ecological/functional differences in flight characteristics. One of us (JWA) is examining both wing shape and vein length differences among populations and species of Hawaiian picture-winged *Drosophila* where drastic wing shape differences between species and sexes are apparent.

The ready availability of inexpensive coordinate digitizers attached to microcomputers (or to terminals) makes these methods practical for general use. The actual process of digitizing outlines is tedious, however. It takes several minutes to position an image, carefully trace its outline, and then store the information in a computer. This process can be greatly speeded up for conveniently-sized high-contrast images (where one does not have to adjust a microscope and the background can clearly be distinguished from the object) by the use of a television camera attached to a computer-controlled digitizer. Such a setup in our laboratory (see Ferson et al., in prep.) enables large numbers of specimens to be processed since only a few seconds are required for each image. The principal problem is the fact that the precision of the resulting measurements is somewhat limited. The scanner in our laboratory has a resolution of maximally 480 by 640 pixels compared to the 1,700 by 1,700 field on the digitizer used in the present study. Most video units presently available for microcomputers have much less resolution (undoubtedly only a temporary limitation). Charge-coupled device (CCD) cameras are also available that have a resolution of 1,700 by 2,200 (as much resolution as a digitizer), but the processing of an image with 3.74×10^6 bytes requires much more computational effort. The determination of which device should be used will depend upon the complexity of the outline of the image and how many harmonics are needed to describe it satisfactorily. This will have to be determined in each study.

ACKNOWLEDGMENTS

This paper represents contribution number 489 from Graduate Studies in Ecology and Evolution, State University of New York at Stony Brook. This work was supported in part by a grant (BSR8202269) from the National Science Foundation. Some of the preliminary computations were performed while one of us (FJR) was on leave at the IBM Thomas J. Watson Research Center. Mr. Scott Ferson assisted this project in many ways. His help was greatly appreciated. Barbara Thomson assisted in the preparation of the final figures.

REFERENCES

ANSTEY, R. L., AND D. A. DELMET. 1973. Fourier analyses of zooecial shapes in fossil tubular bryozoans. Geol. Soc. Am. Bull., 84:1753–1764.

BOOKSTEIN, F. L. 1982. Foundations of morphometrics. Annu. Rev. Ecol. Syst., 13:451–470.

BOOKSTEIN, F. L., R. E. STRAUSS, J. M. HUMPHRIES, B. CHERNOFF, R. L. ELDER, AND G. R. SMITH. 1982. A comment on the uses of Fourier methods in systematics. Syst. Zool., 31:85–92.

BROWN, K. R. 1979. Multivariate assessment of phenetic relationships within the tribe Luciliini (Diptera: Calliphoridae). Aust. J. Zool., 27:465–477.

BROWN, K. R., AND E. SHIPP. 1977. Wing morphometrics of Australian Luciliini (Diptera: Calliphoridae). Aust. J. Zool., 25:765–777.

BROWN, K. R., AND E. SHIPP. 1978. Wing morphometric analysis of Australian Sarcophaginae (Diptera: Sarcophagidae). Syst. Entomol., 3:179–188.

CARPENTER, S. J., AND W. J. LACASSE. 1955. Mosquitoes of North America (North of Mexico). Univ. California Press, Los Angeles.

CROVELLO, T. J. 1969. Numerical taxonomy: Its value to mosquito taxonomy. Mosquito Syst., 1:63–67.

EHRLICH, R., R. B. PHARR, JR., AND N. HEALY-WILLIAMS. 1983. Comments on the validity of Fourier descriptors in systematics: A reply to Bookstein et al. Syst. Zool., 32:302–306.

GOWER, J. 1971. Statistical methods of comparing different multi-variate analyses of the same data. Pages 138–149 *in* Mathematics in the archaeological and historical sciences (F. R. Hodson, D. G. Kendall, and P. Tautu, eds.). Edinburgh Univ. Press, Edinburgh.

HENDRICKSON, J. A., AND R. R. SOKAL. 1968. A numerical taxonomic study of the genus *Psorophora* (Diptera: Culicidae). Ann. Entomol. Soc. Am., 61: 385–392.

HU, M. K. 1962. Visual pattern recognition by moment invariants. IRE Trans. on Information Theory, 8:179–187.

KAESLER, R. L., AND J. A. WATERS. 1972. Fourier analysis of the ostracod margin. Bull. Geol. Soc. Am., 83:1169–1178.

KINCAID, D. T., AND R. B. SCHNEIDER. 1983. Quantification of leaf shape with a microcomputer and Fourier transform. Can. J. Bot., 61:2333–2342.

KUHL, F. P., AND C. R. GIARDINA. 1982. Elliptic Fourier features of a closed contour. Computer Graphics and Image Processing, 18:236–258.

MOELLERING, H., AND J. N. RAYNER. 1982. The dual axis Fourier shape analysis of closed cartographic forms. Cartographic J., 19:53–59.

MOELLERING, H., AND J. N. RAYNER. 1983. The har-

monic analysis of spatial shapes using dual axis Fourier shape analysis (DAFSA). Geograph. Analysis, 13:64–77.

MOSS, W. W., W. A. STEFFAN, N. L. EVENHUIS, AND D. L. MANNING. 1979. The genus *Toxorhynchites* (Diptera: Culicidae); Analysis of *T. splendens* and allies using techniques of numerical taxonomy. Mosquito Syst., 11:258–274.

NIELSEN, L. T. 1969. A critique on numerical taxonomy. Mosquito Syst., 1:23–25.

PLOWRIGHT, R. C., AND W. P. STEPHEN. 1973. A numerical taxonomic analysis of the evolutionary relationships of *Bombus* and *Psithyrus* (Apidae: Hymenoptera). Can. Entomol., 105:733–743.

RALSTON, A. 1965. A first course in numerical analysis. McGraw Hill, New York.

ROBERTS, D. MCL., A. WAREN, AND C. R. CURDS. 1983. Morphometric analysis of outline shape applied to the peritrich genus *Vorticella*. Syst. Zool., 32:377–388.

ROHLF, F. J. 1963. Classification of *Aedes* by numerical taxonomic methods (Diptera: Culicidae). Ann. Entomol. Soc. Am., 56:798–804.

ROHLF, F. J. 1967. Correlated characters in numerical taxonomy. Syst. Zool., 16:109–126.

ROHLF, F. J. 1972. Empirical comparison of three ordination techniques in numerical taxonomy. Syst. Zool., 21:271–280.

ROHLF, F. J. 1977. Classification of *Aedes* mosquitoes using statistical methods. Mosquito Syst., 9:372–388.

ROHLF, F. J. 1982. Consensus indices for comparing classifications. Math. Biosci., 59:131–144.

ROHLF, F. J., AND S. FERSON. 1983. Image analysis. Pages 583–599 *in* Numerical taxonomy (J. Felsenstein, ed.). NATO ASI Series G, Ecological Sciences No. 1. Springer-Verlag, New York.

ROHLF, F. J., AND R. R. SOKAL. 1967. Taxonomic structure from randomly and systematically scanned biological images. Syst. Zool., 16:246–260.

SIMON, C. M. 1983. Morphological differentiation in wing venation among broods of 13- and 17-year periodical cicadas. Evolution, 37:104–115.

SIMON, C. M., W. A. STEFFAN, W. W. MOSS, AND N. L. EVENHUIS. 1982. The genus *Toxorhynchites* (Diptera: Culicidae); Numerical phylogenetic analysis of *Tx. splendens* and allies with phenetic comparisons. Mosquito Syst., 14:221–261.

SNEATH, P. H. A., AND R. R. SOKAL. 1973. Numerical taxonomy. W. H. Freeman and Co., New York.

STEWARD, C. C. 1968. Numerical classification of the Canadian species of the genus *Aedes*. Syst. Zool., 17:426–437.

ZAHN, C. T., AND R. Z. ROSKIES. 1972. Fourier descriptors for plane closed curves. IEEE Trans. on Computers, 21:269–281.

Received 7 February 1984; accepted 1 March 1984.