

Quantitative Evaluation of Soybean (*Glycine max* L. Merr.) Leaflet Shape by Principal Component Scores Based on Elliptic Fourier Descriptor

Naoya Furuta¹⁾⁴⁾, Seishi Ninomiya²⁾, Nobuo Takahashi³⁾, Hiroshi Ohmori¹⁾ and Yasuo Ukai¹⁾

¹⁾ Faculty of Agriculture, University of Tokyo, Yayoi, Bunkyo-ku, Tokyo 113, Japan

²⁾ National Institute of Agro-Environmental Sciences, Tsukuba, Ibaraki 305, Japan

³⁾ Nagano Prefecture Chushin Agricultural Experimental Station, Shiojiri, Nagano 391-64, Japan

⁴⁾ Present address ; Mitsubishi Research Institute, Otemachi, Tokyo 100, Japan

Summary

Leaflet shape of thirty-nine soybean cultivars/strains selected to cover the possible diversity of leaf shape, was quantitatively evaluated by principal components scores based on the elliptic Fourier descriptor of contours. After central leaflets of fully expanded compound-leaves of the cultivars/strains were videotaped, binary images of the leaflets were obtained from those video images by image processing. Then, the closed contour of each leaflet was extracted from the binary images and chain-coded by image processing. Because the first twenty harmonics could sufficiently represent soybean leaf contours, 77 elliptic Fourier coefficients were calculated for each chain-coded contour. Then, the Fourier coefficients were standardized so that the coefficients were invariant of the size, rotation, shift and chain-code starting-point of any contour. The principal component analysis about the standardized Fourier coefficients, showed that the cumulative contribution at the fifth principal component was about 96 %. Moreover, the effect of each principal component on the leaf shape was clarified by drawing the contours of leaflets using the Fourier coefficients inversely estimated under some typical values of the principal component scores. Consequently, it was indicated that the principal components scores about the standardized elliptic Fourier coefficients gave us powerful quantitative measures to evaluate soybean leaf shape. The analysis of variance and multiple comparison indicated that the genotypic differences on the first, the second and the fifth principal components were significantly large. Because the variations of those principal components were continuous, the effects of the polygenes on the (size-invariant) shape were also suggested.

Key Words : *Glycine max* L. Merr., leaflet shape, elliptic Fourier descriptor, image analysis, principal component.

Introduction

Quantitative evaluation of contour shape is often important in morphological analyses. There have been several methods suggested to evaluate contour shape. Length of a contour, moments of a contour region and Fourier de-

scriptors are the examples. Among those methods, a series of Fourier descriptors has taken important roles (Bierbaum and Ferson 1986, Bookstein *et al.* 1982, Diaz *et al.* 1989, Ehrlich and Weinberg 1970, Ehrlich *et al.* 1983, Ferson *et al.* 1985, Rohlf and Archie 1984, White *et al.* 1988), because those methods are good at directly representing contour shape itself. The common features of the Fourier methods are to express a contour by some periodical function(s), to expand the function(s) by Fourier series expansion and to represent the contour shape by Fourier coefficients. The elliptic Fourier descriptor is one of those Fourier descriptors (Giardina and Kuhl 1977, Granlund 1972, Kuhl and Giardina 1982). As will be explained precisely later, elliptic Fourier method uses the periodical variation of x and y coordinates of a mass point that moves on a contour at the same speed. The elliptic Fourier method has been used in morphological analyses of some species (Rohlf and Archie 1984, Ferson *et al.* 1985, Bierbaum and Ferson 1986, White *et al.* 1988, Diaz *et al.* 1989).

In this study, we applied the elliptic Fourier method to develop quantitative measures to evaluate soybean leaf shape (*Glycine max* L. Merr.), because soybean leaf shape is one of the important factors that compose soybean plant shape and its inheritance needs to be clarified. When the inheritance of soybean leaf shape was ever investigated (Takahashi 1934, Domingo 1945), the shape was evaluated quite qualitatively (e.g., ovate, normal and narrow) and a locus with incomplete dominance was suggested to explain the leaflet shape. We, however, empirically know that the leaf shape of soybean shows much more complicated inheritance. One may imagine that aspect ratio of leaf (often referred as a leaf shape index) can be a good quantitative measure. It, however, does not represent the contour shape that subtly varies depending on genotypes. Therefore, the development of quantitative measures to evaluate the contour shape, itself, is inevitable.

Materials and Methods

Thirty-nine soybean cultivars were selected considering the diversity of leaflet shape (Table 1). They were grown at the Nagano Prefecture Chushin Agricultural Experimental Station in 1990 under the ordinal cultivation of the station. After the uppermost compound

leaves were fully expanded, the central leaflets of the second uppermost leaves were removed at pulvinus. Each sampled leaflet was placed on a white board with a ruler and video-taped one by one.

The original analogue video image of NTSC standard composite video was digitized into R (red), G (green) and B (blue) digital images by a real time A/D converter (Nexus 68320, Kashiwagi Res., Corp., Tokyo) after being decoded into analogue RGB (red, green and blue) video signals by a composite video decoder (ED-1000 N, Photoron Co., Ltd. Tokyo). Each digital image is composed of 512 X 480 pixels and each pixel is 8-bits deep (256 gray levels). The digitized images were processed in the color image processor (Nexus 6410, Kashiwagi Res., Corp., Tokyo) which is controlled by a workstation

Table 1. List of the cultivars and strains examined in this study. The code number indicates the cultivar/strain code of the Chushin Agricultural Experimental Station

Code No. Name	Number of samples
19 Yamabedaizu	10
31 Karikei 408	10
42 Karikei 429	10
48 Touhoku 95	10
53 Touhoku 103	10
57 Chuiku 200	10
62 Juiku 10	10
63 Keburi	10
67 Tachikogane	11
75 Katsumokugokushouzairai	10
79 Tamahikari	10
201 Bay	10
202 N83-640	10
203 D74-9820	11
204 Young	11
207 NCBPMR-3	10
214 Tracy	11
215 Essex	11
216 Monroe	10
217 Aksarben	10
219 Suwon 83	10
227 Chulhu 19	13
230 Bosserer	8
303 Tanbaguro	10
401 Ao 1038 Tsurunoko	9
417 cf 1010 Keihandaizu	10
420 Cg 2002 Kamiwa	8
528 Iwate 489	10
530 Iwate 577	7
548 Iwate 668	11
602 Sato daizu	12
606 Hitashimame	10
607 Aomame	6
615 Kuromame	9
620 Batouzairai	10
625 Kasugaodaizu	10
631 Tenryuzairai 5	6
653 Nakaowase 2	5
654 Kuromame	6
Total	375

(NEWS 830, SONY Corp., Tokyo) through a VMS-bus interface. The binary image of each leaflet was obtained by thresholding the original red digital image of leaflet. The level of the thresholding was adjusted visually. The adjustment was easy because the contrast of the leaflets on the white board was high. After noise reduction and filling holes on the binary image, only the closed contour of each leaflet was extracted by edging the binary image. Then, the extracted contour was chain-coded (Rosenfeld and Kak 1982). The contour of each sampled leaflet was composed of about 300 to 400 chain-coded points.

Each chain-coded contour was projected onto XY plane so that the starting point of the chain-code was located at the origin of the coordinate (Fig. 1). Assumed that a mass point moves on the contour with the same speed, the x and y coordinates of the mass point periodically oscillate and they can be expanded by Fourier series expansion. This expression for a closed contour is called the elliptic Fourier descriptor. Assumed that the contour between the $(i-1)$ -th and the i -th chain coded points is linearly interpolated and that the length of the contour from the starting point to the p -th point and the perimeter of the contour are denoted to be t_p and T , respectively, then, $t_p = \sum_{i=1}^p \Delta t_i$ and $T = t_K$, where Δt_i and K are the distance between the $(i-1)$ -th and the i -th points and the total number of the chain-coded points on the contour respectively. Notice that the K -th point is equivalent to the starting point. Let x and y coordinates of the p -th points denoted as x_p and y_p . Then, $x_p = \sum_{i=1}^p \Delta x_i$ and $y_p = \sum_{i=1}^p \Delta y_i$, where Δx_i and Δy_i are the distance along the x and y axes between the $(i-1)$ -th and the i -th point (Fig. 1). The elliptic Fourier expansions of the coordinates on the contour are

$$x_p = A_o + \sum_{n=1}^{\infty} \left(a_n \cos \frac{2n\pi t_p}{T} + b_n \sin \frac{2n\pi t_p}{T} \right)$$

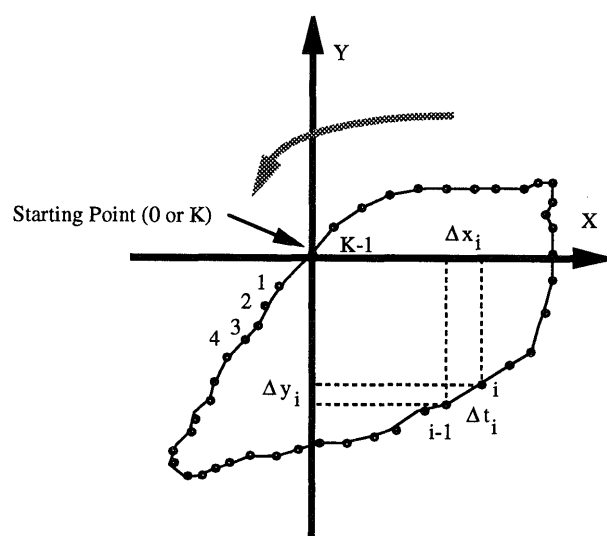


Fig. 1. Schematic diagram of a leaflet contour. Circles show the chain-coded points on the contour of a digital image. See the text for details.

$$\text{and } y_p = C_0 + \sum_{n=1}^{\infty} \left(c_n \cos \frac{2n\pi t_p}{T} + d_n \sin \frac{2n\pi t_p}{T} \right).$$

Then, assuming the linear interpolation between the neighboring points, the elliptic Fourier coefficients of the n -th harmonic (a_n , b_n , c_n , d_n) are given as

$$a_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta x_p}{\Delta t_p} \left(\cos \frac{2n\pi t_p}{T} - \cos \frac{2n\pi t_{p-1}}{T} \right)$$

$$b_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta x_p}{\Delta t_p} \left(\sin \frac{2n\pi t_p}{T} - \sin \frac{2n\pi t_{p-1}}{T} \right)$$

$$c_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta y_p}{\Delta t_p} \left(\cos \frac{2n\pi t_p}{T} - \cos \frac{2n\pi t_{p-1}}{T} \right)$$

$$\text{and } d_n = \frac{T}{2n^2\pi^2} \sum_{p=1}^K \frac{\Delta y_p}{\Delta t_p} \left(\sin \frac{2n\pi t_p}{T} - \sin \frac{2n\pi t_{p-1}}{T} \right).$$

In this study, we used the first twenty harmonics to express the contour shape of soybean leaflets, because the approximation of the soybean leaf shape had been sufficiently good with the number in our preliminary examination. Because those coefficients are not invariant in the size, rotation, shift and starting point of chain coding about a contour, they cannot be indicators to compare the shape of objects. For example, the Fourier coefficients differ even on completely the same shape when the starting point of the chain-coding differs. In this study, the standardization of the Fourier coefficients were made according to Kuhl and Giardina (1982). Let the standardized coefficients of the n -th harmonic be a_n^{**} , b_n^{**} , c_n^{**} and d_n^{**} . Then,

$$\begin{bmatrix} a_n^{**} & b_n^{**} \\ c_n^{**} & d_n^{**} \end{bmatrix} = \frac{1}{E^*} \begin{bmatrix} \cos \Psi & \sin \Psi \\ -\sin \Psi & \cos \Psi \end{bmatrix} \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix}$$

where $E^* = (a_1^{*2} + c_1^{*2})^{1/2}$,

$$\begin{bmatrix} a_1^* \\ c_1^* \end{bmatrix} = \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}$$

$$\Psi = \arctan \left[\frac{c_1^*}{a_1^*} \right] \quad (0 \leq \Psi < 2\pi)$$

$$\text{and } \theta = \frac{1}{2} \arctan \left[\frac{2(a_1 b_1 + c_1 d_1)}{(a_1^2 + c_1^2 - b_1^2 - d_1^2)} \right] \quad (0 \leq \theta < \pi).$$

This standardizing procedure is based on the ellipse of the first harmonic. E^* and Ψ are the length and the direction of the major axis of the first harmonic ellipse, respectively. The former is the parameter for the size-invariance and the latter is that for the rotation invariance. Because there are two solutions of Ψ , we choose one of them considering the comparative position of the pulvinus of each sample leaflet to the first harmonic ellipse. θ is the phase shift of the radius vector of the first harmonic ellipse at $p=0$ and the parameter for the chain-code starting point invariance. The shift-invariance is obtained simply by setting the location pa-

rameters of the Fourier expansions (A_0 and C_0) to be zero. This standardization makes a_n^{**} , b_n^{**} , c_n^{**} and d_n^{**} independent of the size, rotation, shift and chain-code starting point of a contour. Because a_1^{**} , b_1^{**} and c_1^{**} become constant by the standardization, we, consequently, have 77 standardized coefficients excluding the constant coefficients ($20 \times 4 - 3$).

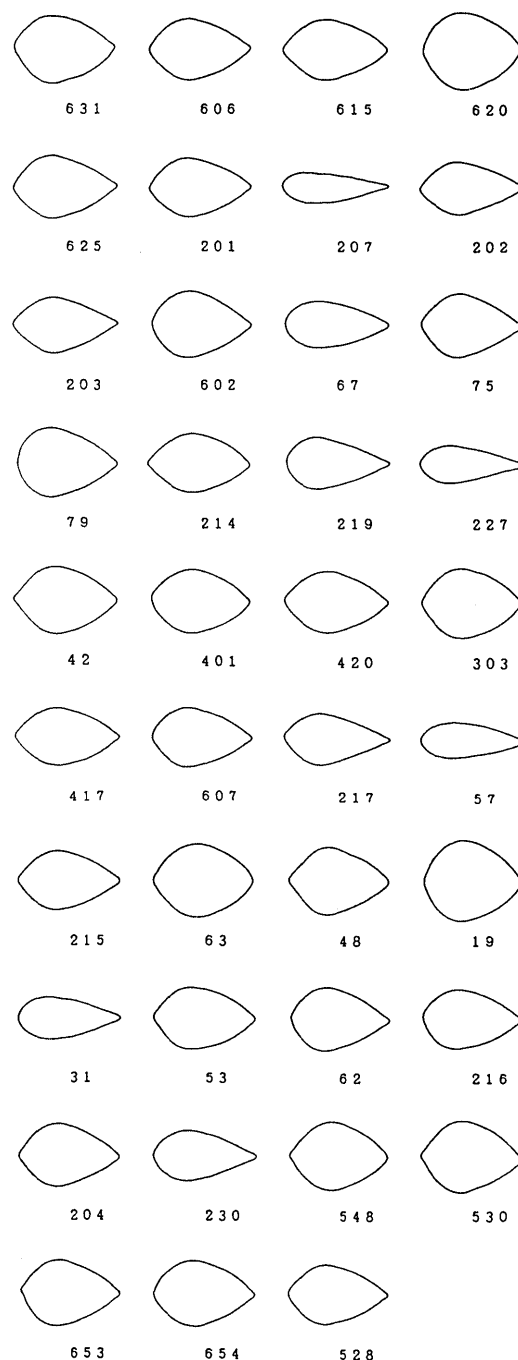


Fig. 2. Mean leaflet shape of the thirty-nine soybean cultivars/strains examined in this study. The numbers show the cultivars/strains codes in Table 1.

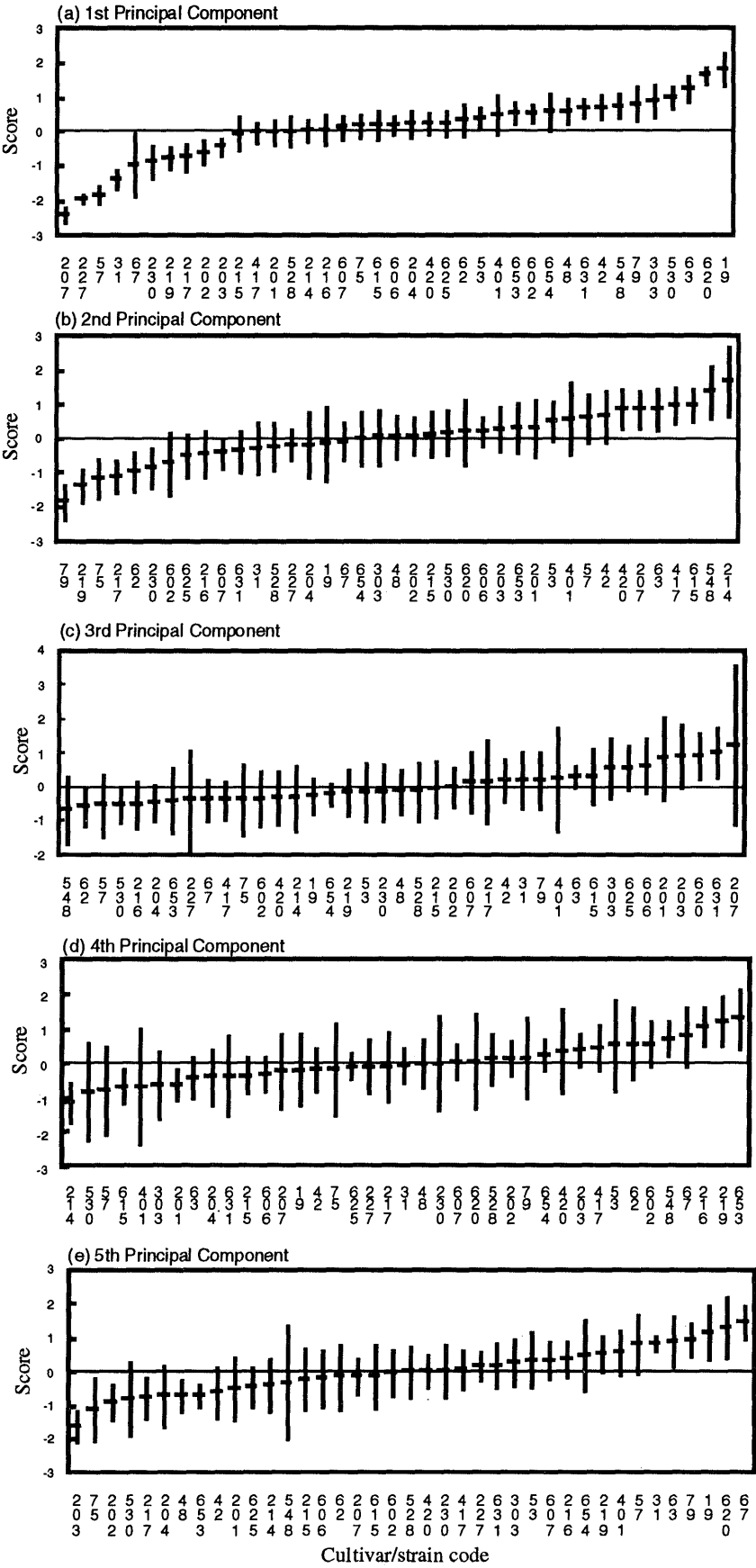


Fig. 3. Variation of the first to fifth principal component scores based on 77 standardized elliptic Fourier coefficients among the thirty-nine soybean cultivars/strains. Vertical bars indicate the standard errors. For each principal component, the cultivars/strains are arranged in the order of the scores.

Results

The standardized elliptic Fourier coefficients of 375 soybean leaflets in thirty-nine cultivars/strains were calculated. Then, the mean leaflet shape of each cultivar/strain was drawn using the mean values of the standardized Fourier coefficients within each cultivar/strain (Fig. 2). The wide variation of soybean leaflet shape is clear in Fig. 2.

Then, the principal components analysis was made with the covariance matrix of the standardized Fourier coefficients by a SAS procedure, PRINCOMP (SAS Institute Inc. 1985). Because the cumulative contribution at the fifth principal component was over 96 %, we considered only the first five principal components in this study. The contribution of the first, second, third, fourth and fifth principal components were 88.5 %, 3.0 %, 1.9 %, 1.5 % and 1.1 %, respectively.

Fig. 3 shows the variation of the first, second, third, fourth and fifth principal component scores of thirty-nine soybean cultivars/strains. One-way analyses of variance on the first five principal components showed that the differences between cultivars/strains were significant at 0.1 % level for all of these five components, while Scheffe's multiple comparisons showed no significant combination at the third and fourth principal components.

Finally, elliptic Fourier coefficients were inversely estimated for the case that the score at a principal component took the value of $\pm 2\sigma$ (standard deviation of the scores at the principal component) while the scores for the remaining principal components were kept zero (Notice that the mean value of the scores at each principal component equals zero). This is a problem of solving a simultaneous equation if the eigen vectors obtained in the above principal component analysis, is given. Then, we could visually recognize the relationship between each principal component and the leaf shape by redrawing the contour with the estimated coefficients for the corresponding principal component. Fig. 4 shows the estimated leaf shape in this way. The first principal component represents the aspect ratio of the leaflet shape. This can be simply explained by the fact that the element of the eigen vector for d_1^{**} at the first principal component was outstandingly large, because d_1^{**} represents the relative length of the minor axis to the major axis of the first harmonic ellipse. That is, the first principal component score is, somehow, equivalent to the leaf shape index (aspect ratio of leaves). The second principal component score represents the location of the centroid of leaflets along the midrib. The third and fourth principal component scores represent the straightness or distortion of leaflet shape. The fifth principal component score represents the roundness of leaflet shape. We should notice that a principal component is independent of any other principal component.

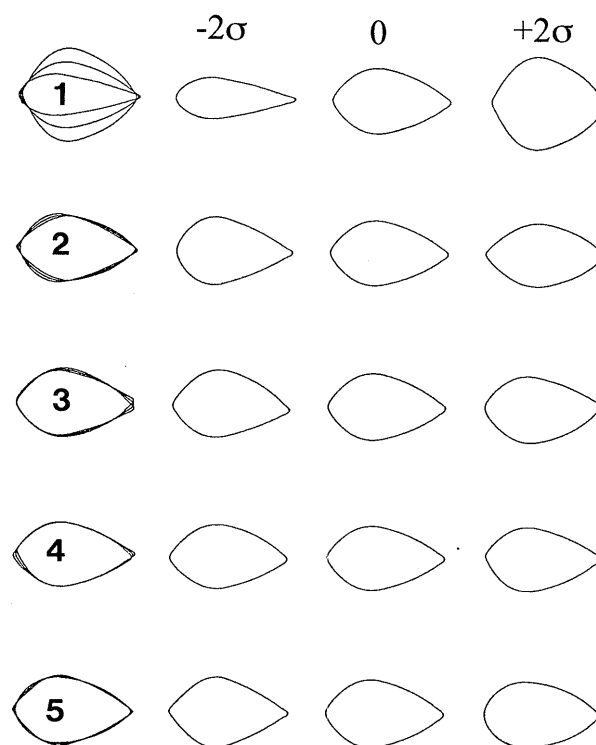


Fig. 4. Leaflet contour shape drawn using the Fourier coefficients estimated under three typical values of each principal components score ($+2\sigma$, -2σ and zero). The numbers, 1 to 5 correspond to the first to fifth principal components, respectively. Each column shows the case that the score takes either $+2\sigma$, zero or -2σ from the right and the leftmost column shows the overlaid drawings of the three cases. See the text for details.

Discussion

There have been many discussions about applying the Fourier descriptor to morphological analysis. In fact, the Fourier descriptor has been a powerful method to describe the biological contour shape. It has been, however, very difficult to understand the morphological meanings of the Fourier coefficients, because there are usually too many coefficients to be considered at once and because it is almost impossible to understand the effect of each coefficient on shape. In this study, we found that the first five principal component scores represented more than 96 % of the soybean leaf shape expressed by 77 Fourier coefficients. Moreover, we also clarified the effect of each principal component on the shape by redrawing the contours with the Fourier coefficients estimated for the cases that the principal component score took some typical values. That is, we could successfully partition the shape information of soybean leaflets into only a few quantitative variables (principal components) that were independent of each other and each of which represented the visually recognizable feature of the soybean leaflet shape. It is also important that, with our standardized Fourier method, we can evaluate only the shape of the objects independently of the size factor.

The analysis of variance and multiple comparison indicated that the genotypic differences on the first, the second and the fifth principal components were large. Because the variations of those principal components were continuous, the effects of the polygenes for the shape were also suggested. The discussion on the inheritance of the shape features represented by those principal components will be made in the future.

Finally, we should discuss the features of principal components that are fully data-dependent. That is, the shape features represented by each principal components in this study may vary in different data sets. This is theoretically unavoidable. To overcome this problem, we have to select a data set that contains as large variation of leaflet shape as possible. For example, the development of a database for the standardized Fourier coefficients of soybean leaflets can be strongly suggested. Whenever we have a new data set, we add the data set to the database and make a principal component analysis on the whole data set in the database repeatedly, so that the above problem gets solved gradually as the database grows.

Acknowledgments

We would like to express our special thanks to all the members of the soybean breeding section of Nagano Prefecture Chushin Agricultural Experimental Station for their kind help, especially for sampling data and cultivating soybean for this study.

Literature Cited

- Bierbaum, R. M. and S. Ferson (1986) Do symbiotic pea crabs decrease growth rate in mussels? *Biol. Bull.* 170 : 51-61.
- Bookstein, F. L., R. E. Strauss, J. M. Humphries, B. Cheronoff, R. L. Elder and G. R. Smith (1982) A Comment upon the uses of Fourier methods in systematics. *Syst. Zool.* 31 : 85-92.
- Diaz, G., A. Zuccarelli, I. Pelligra and A. Ghiani (1989) Elliptic Fourier analysis of cell and nuclear shapes. *Comp. Biomed. Res.* 22 : 405-414.
- Domingo, W. E. (1945) Inheritance of number of seeds per pod and leaflet shape in the soybean. *J. Agric. Res.* 70 : 251-268.
- Ehrlich, R. and B. Weinberg (1970) An exact method for characterization of grain shape. *J. Sed. Pet.* 40 : 205-212.
- , 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 : 202-206.
- Ferson, S., F. J. Rohlf and R. K. Koehn (1985) Measuring Shape Variation of Two-dimensional Outlines. *Syst. Zool.* 34 : 59-68.
- Giardina, C. R. and F. P. Kuhl (1977) Accuracy of curve approximation by harmonically related vectors with elliptical loci. *Comp. Graph. Ima. Proc.* 6 : 277-285.
- Granlund, G. H. (1972) Fourier preprocessing for hand print character recognition. *IEEE Trans. Comput.* C-21 : 269-281.
- Rosenfeld, A. and A. C. Kak (1982) "Digital Picture Processing Vol. 2", 2nd ed., Academic Press, INC. San Diego, pp 349.
- Kuhl, F. P. and C. R. Giardina (1982) Elliptic Fourier features of a closed contour. *Comp. Graph. Ima. Proc.* 18 : 236-258.
- Rohlf, F. J. and J. W. Archie (1984) A comparison of Fourier methods for the description of wing shape in mosquitoes (*Diptera culicidae*). *Syst. Zool.* 33 : 302-317.
- SAS Institute Inc. (1985) "SAS User's Guide : Statistics Version 5 Edition", Cary, SAS Institute Inc.
- Takahashi, N. (1934) Linkage relation between the genes for the forms of leaves and the number of seeds per pod of soybeans. (In Japanese) *Jpn. J. Genet.* 9 : 208-225.
- White, R., H. C. Rentice and T. Verwist (1988) Automated image acquisition and morphometric description. *Can. J. Bot.* 66 : 450-459.