

Genetic Diversity of Soybean Pod Shape Based on Elliptic Fourier Descriptors

Ngon T. Truong***, Jae-Gyun Gwag***, Yong-Jin Park***, and Suk-Ha Lee*†

*Dep. of Plant Science, Seoul National University, Seoul 151-742, Korea

**Dep. of Crop Sciences, Can Tho University, Can Tho, Viet Nam

***Genetic Resources Div., National Institute of Agricultural Biotechnology, Suwon 441-707, Korea

ABSTRACT: Pod shape of twenty soybean (*Glycine max* L. Merrill) genotypes was evaluated quantitatively by image analysis using elliptic Fourier descriptors and their principal components. The closed contour of each pod projection was extracted, and 80 elliptic Fourier coefficients were calculated for each contour. The Fourier coefficients were standardized so that they were invariant of size, rotation, shift, and chain code starting point. Then, the principal components on the standardized Fourier coefficients were evaluated. The cumulative contribution at the fifth principal component was higher than 95%, indicating that the first, second, third, fourth, and fifth principal components represented the aspect ratio of the pod, the location of the pod centroid, the sharpness of the two pod tips and the roundness of the base in the pod contour, respectively. Analysis of variance revealed significant genotypic differences in these principal components and seed number per pod. As the principal components for pod shape varied continuously, pod shape might be controlled by polygenes. It was concluded that principal component scores based on elliptic Fourier descriptors yield seemed to be useful in quantitative parameters not only for evaluating soybean pod shape in a soybean breeding program but also for describing pod shape for evaluating soybean germplasm.

Keywords: *Glycine max*, soybean, pod shape, elliptic Fourier descriptors, principal component analysis (PCA).

Variation in shape has always been an important means of distinguishing individuals. As quantity and quality of crop products are always affected directly or indirectly by crop organs, these phyto-organ shapes are important target for crop improvement. Pod shape was suggested to be one of the important descriptors for evaluating soybean genetic resources (IPGRI, 1998; USDA, 2001).

The agronomic traits are always considered the main criteria for selecting parent materials to make crosses with high yield or resistance to insect or disease. However, pod shape

has less attention in soybean breeding program. Therefore, soybean pod shape and systematic relation between shape and agronomic traits have not been clear yet.

Quantitative evaluation of contour shape is often important in morphological analysis. Contour shape has been evaluated in several different ways such as length of a contour, moments of a contour region, and Fourier descriptors. Among those methods, a series of Fourier descriptors was thought to be efficient (Bierbaum & Ferson, 1986; Bookstein *et al.*, 1982; Diaz *et al.*, 1989; Ehrlich *et al.*, 1983; Ferson *et al.*, 1985; Rohlf & Archie, 1984; White *et al.*, 1988), as those methods were 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 & Kuhl 1977; Granlund, 1972; Kuhl & Giardina, 1982). As 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 (Rohlf & Archie, 1984; Ferson *et al.*, 1985; Bierbaum & Ferson, 1986; White *et al.*, 1988; Diaz *et al.*, 1989).

The digital image analysis has been applied for seed discrimination (Keefe & Draper, 1986; Neuman *et al.*, 1987; Myers & Edsall, 1989). Recently, Furuta *et al.* (1995) successfully evaluated soybean leaflet shape quantitatively by principal component scores based on the elliptic Fourier descriptor. Following this study, preliminary applications for buckwheat kernel shape analysis were also carried out for both common buckwheat and tartary buckwheat (Ninomiya *et al.*, 1995; Yoshida *et al.*, 1995). It was revealed that principal component analysis based on the elliptic Fourier descriptor was able to classify buckwheat cultivars/strains, and was also practically useful for analysis of buckwheat kernel shape.

In this study, the elliptic Fourier method for developing quantitative measures was used to evaluate soybean pod shape, because soybean pod shape is one of important factors

†Corresponding author: (Phone) +82-2-880-4545 (E-mail) sukhalee@snu.ac.kr

<Received March 11, 2005>

determining seed shape and seed size. The objective of this study is to test the applicability of elliptic Fourier method for evaluating genetic diversity of pod shape in soybean.

MATERIALS AND METHODS

Plant materials

Nineteen soybean local collections from different regions, which were kindly provided by Gene Bank, Rural Development Administration in Korea and one recommended soybean variety, Sinpaldalkong 2, were used for this study (Table 1). Soybean seeds were sown on May 31, 2003 and harvested in October, 2003. Each variety was planted in a single two-meter row with 60 cm between rows and 15 cm between plants. Two seeds were sown per hole.

Agronomic traits, yield and yield components were determined. Measurements were recorded as the average from five representative plants of each variety selected randomly. In addition, morphological traits such as seed coat, hilum color, were recorded, and seed and pod shapes were photographed by digital camera Olympus D550.

Image processing and contour recording

In order to evaluate pod shape, twelve mature pods for each variety were examined by image analysis. These typical pods were selected at random from each variety. After harvesting all plants, images of pod and seed were directly taken by digital camera (Camedia Olympus D550, Olympus America Incorporation) with the 45-centimeter fixed distance from objects to camera. Each digital image has 640 x 480 spatial resolution and 8-bit quantization (256 gray levels), and then images were transferred into the computer.

To process images, two different types of software were used. The PhotoImpact Bundle 5.0 (Ulead Systems Incorporation) is used to adjust contrast and categorize types of pod such as one-seeded, two-seeded, or three-seeded pods. SHAPE analysis package (Iwata, 2002) was used to analyze pod shape.

Elliptic Fourier descriptors

Pod shape of twenty local cultivars was analyzed by elliptic Fourier descriptors method. The coefficients of elliptic Fourier descriptors are calculated by the discrete Fourier transformation of chain-coded contours according to the method by Kuhl & Giardina (1982) with the following main steps: A contour of the digitized shapes can be represented as a sequence of the x- and y- coordinates of ordered points measured contour-clockwise from an arbitrary starting

point. Assuming that the contour between two adjacent points is linearly interpolated, and the length of the linear segment between the $(i - 1)$ th and the i th points is Δt_i , then the length of the contour from starting point to the p th point is $\Delta t_p = \sum_{i=1}^p \Delta t_i$, and the perimeter of the contour is $T = t_k$, where K is the total number of the points on the contour. The K th point is equivalent to the starting point. The x- coordinate of the p th point is $x_p = \sum_{i=1}^p \Delta x_i$, where Δx_i is the displacement along the x- axis of the contour between the $(i - 1)$ th and i th points. Then, elliptic Fourier expansion of the sequences of the x - coordinates gives

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

where

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

and

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

In the above equation, x_{cen} is the coordinate of the center point, and n is the harmonic number of the coefficients (a_n and b_n). The coefficient values for the y- coordinates, c_n and d_n , are found in the same way.

The coefficients of elliptic Fourier descriptors can be mathematically normalized to be invariant to size, rotation and starting point of the contour trace. In SHAPE, the coefficients can be normalized using two types of procedures; one based on the ellipse of the first harmonic and the other based on the longest radius. For more detailed information about the normalization, see Kuhl & Giardina (1982).

Data analysis

Agronomic traits, yield, and yield components were determined as average values of five plant samples for each genotype. Principal component analysis of pod shape and redrawing pod shape were conducted by SHAPE analysis package (Iwata, 2002). Dendrogram of 20 genotypes was constructed by the unweighted pair-group method using arithmetic averages as described by Sneath & Sokal (1973) and drawing using STATISTICA 6.0 software (Jaccard, 1908).

RESULTS

Agronomic traits

Agronomic characters of twenty soybean genotypes were presented (Table 1). Days to flowering ranged from 49 days

Table 1. Agronomic traits of twenty soybean genotypes.

Genotype	Days to flowering (days)	Maturity (days)	Branch number	One+two seeded pod (A)	Three seeded pod (B)	Ratio (A/B)
IT102668	60	140	7 ± 1.3	138	8	17:1
IT108886	53	133	5 ± 0.5	57	2	29:1
IT101280	55	136	8 ± 1.7	221	6	37:1
IT104411	70	150	5 ± 1.1	100	4	25:1
IT104535	56	135	6 ± 0.8	87	1	87:1
IT104554	56	134	6 ± 0.8	40	3	13:1
IT141548	53	131	6 ± 0.8	38	0	38:1
IT141594	69	144	4 ± 1.0	75	4	19:1
IT100826	70	147	7 ± 0.8	123	4	31:1
IT113446	78	156	8 ± 1.1	127	7	18:1
IT113141	69	143	7 ± 1.7	127	9	14:1
IT103906	70	157	4 ± 1.1	44	10	4:1
IT101075	71	159	4 ± 1.7	51	12	4:1
IT100914	51	148	7 ± 0.8	67	3	22:1
IT104704	59	149	8 ± 0.8	86	23	4:1
IT103788	68	146	7 ± 1.1	58	12	5:1
IT105686	67	147	8 ± 1.9	73	9	8:1
IT100804	53	131	8 ± 1.1	78	5	16:1
IT108810	58	151	5 ± 0.7	30	1	30:1
Sinpaldalkong 2	49	140	5 ± 0.7	50	22	2:1
Mean	62	144	6	84	7	2:1 – 87:1
CV (%)	13.75	6.02	23.14	54.89	87.11	

to 78 days, and four soybean genotypes (Sinpaldalkong 2, IT100914, IT141548, and IT108886) were characterized with early flowering. There was also significant difference among soybean genotypes in other characteristics such as maturity, one plus two-seeded pod, and three-seeded pod number. There were six soybean genotypes with earlier maturity than Sinpaldalkong 2 that were IT108886, IT101280, IT104535, IT104554, IT141548, and IT100804.

The number of one-seeded plus two-seeded pod and three-seeded pod is quite different among genotypes. These traits showed wide variation (Table 1). Significant correlation was observed between number of branches and pod number ($r = 0.61^*$). Pod number and the ratio between the one-seeded plus two-seeded pod and three-seeded pod seem to be controlled by genetic traits but were also influenced by cultural practices such as plant spacing.

Table 2 indicated genotypic differences in yield and yield components. IT104704, IT113141, and IT100914 showed high plant yield, as compared with Sinpaldalkong 2. Generally, large seed size and great number of seeds per unit area resulted in the high plant yield.

In soybean, a pod may contain from one to five seeds, but in most common cultivars they usually have two or three seeds per pod. However, a few of them have four or five-seeded pod. There was genotypic variation in the ratio between one and two-seeded pods and three-seeded pod (Table 1). The more pod number, the higher ratio especially in late maturing genotype.

Principal Component Analysis

Soybean pod shape can be evaluated mainly by human visual judgment and categorical classifications based on rough estimate. Fig. 1 indicated the genetic diversity in typical pod shape collected from different regions. Of specific interest was that Sinpaldalkong 2, one of the recommended soybean varieties in Korea rather than soybean collections, has four-seeded pod. This indicated that soybean breeder selected higher number of seeds per pod in formal breeding programs for improved soybean cultivars.

There was wide variation in soybean pod shape (Fig. 1). The mean pod shape of each variety was drawn using the

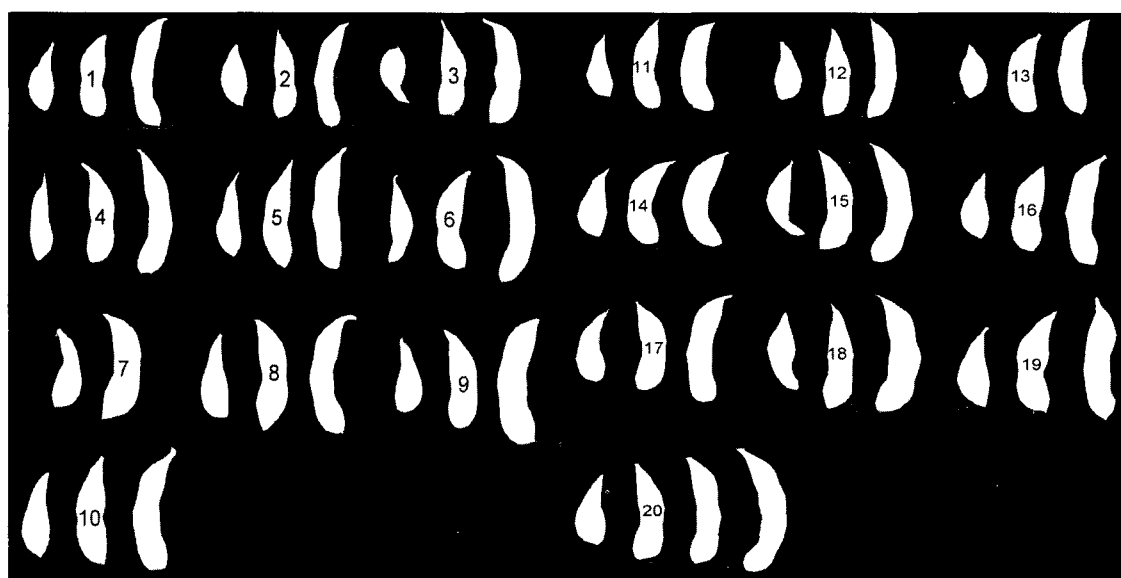
Table 2. Yield and yield components of twenty soybean genotypes.

Genotype	Seed number per m ²	100-seed weight (g)	Plant yield (g/ five plants)
IT102668	4951 ± 24	15.5 ± 0.5	119
IT108886	2109 ± 80	55.9 ± 1.1	192
IT101280	7770 ± 105	14.5 ± 1.3	172
IT104411	3641 ± 27	16.8 ± 1.1	105
IT104535	2797 ± 73	38.3 ± 1.4	231
IT104554	1465 ± 169	40.5 ± 1.5	241
IT141548	1110 ± 37	34.5 ± 1.5	73
IT141594	2797 ± 72	17.3 ± 1.2	91
IT100826	4154 ± 67	16.7 ± 1.4	119
IT113446	4640 ± 94	15.1 ± 0.8	191
IT113141	4862 ± 52	27.6 ± 1.5	257
IT103906	2197 ± 51	14.6 ± 1.4	128
IT101075	2531 ± 99	23.5 ± 1.9	182
IT100914	2620 ± 141	31.6 ± 2.1	257
IT104704	4462 ± 168	36.3 ± 1.7	273
IT103788	2731 ± 143	31.5 ± 2.2	162
IT105686	3086 ± 118	27.8 ± 1.5	183
IT100804	2864 ± 165	22.2 ± 2.0	109
IT108810	999 ± 199	26.7 ± 1.5	52
Sinpaldalkong 2	3241 ± 127	25.1 ± 1.7	133
Mean	3251	26.6	164
CV (%)	48.5	40.95	39.9

mean values of the standardized Fourier coefficients (Fig. 2). Then, the principal component analysis was made with the covariance matrix of the standardized Fourier coefficients by SHAPE analysis package (Iwata, 2002).

Elliptic Fourier coefficients were inversely estimated for the case that the score at a principal component took value of $\pm 2s$ (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 are given.

Then, we could visually recognize the relationship between each principal component and the pod shape by redrawing the contour with the estimated coefficients for the corresponding principal component (Fig. 2). The first principal component represents the aspect ratio of the pod 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 outstanding 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 scores is, somehow, equivalent to the pod shape index (aspect ratio of pods). The second principal component score represent the location of the centroid of pods along the midrib. The third and fourth principal component scores represent the straightness or distortion of pod shape. A principal component was thought to be independent of any other principal

**Fig. 1.** Typical pod shape of twenty soybean genotypes.

(1, IT102668, 2, IT108886, 3, IT101280, 4, IT104411, 5, IT104535, 6, IT104554, 7, IT141548, 8, IT141594, 9, IT100826, 10, IT113446, 11, IT113141, 12, IT103906, 13, IT101075, 14, IT100914, 15, IT104704, 16, IT103788, 17, IT105686, 18, IT100804, 19, IT108810, 20; Sinpaldalkong 2)

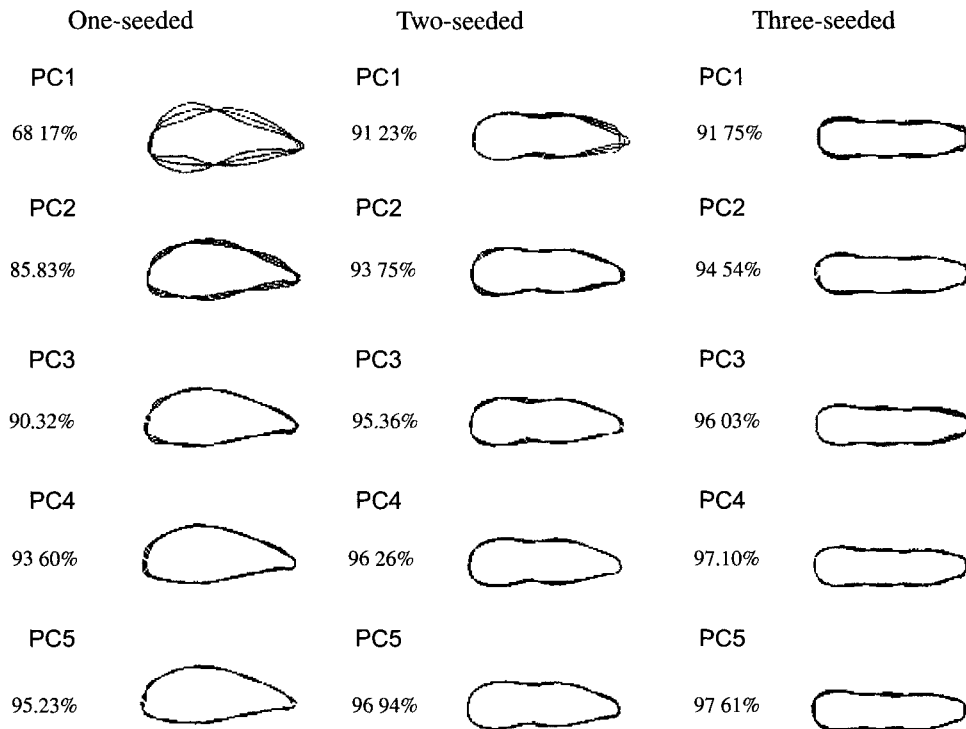


Fig. 2. Each column shows pod contour shape drawn using the Fourier coefficients estimated under principal component score (+ 2 s, - 2 s, and zero); and the redrawings of pod shape. 1st column for one-seeded, 2nd for two-seeded, and 3rd for three-seeded pod

Table 3. Principal component analysis of one-seeded, two-seeded, and three-seeded pod using standardized Fourier coefficients.

Principal Component	Cumulative of pod types (%)		
	One-seeded	Two-seeded	Three-seeded
First	68.2	91.2	91.7
Second	85.8	93.8	94.5
Thurd	90.3	95.4	96.0
Fourth	93.6	96.3	97.1
Fifth	95.2	96.9	97.6

component. As the cumulative contribution at the fifth principal component was over 95%, the first five principal components were considered as significant for this study.

Table 3 shows the result of principal component analysis. The contributions of first, second, third, fourth and fifth principal components to the whole variation of pod shape were 0.9523 in one-seeded pod, 0.9693 in two-seeded pod, and 0.9761 in three-seeded pod. That is, the first five principal components can together explain for more than 95% of total shape variation of pod shape.

We were able to visually recognize the relationship between each principal component and pod shape by redrawing the contour with estimated coefficients for the corresponding principal component (Fig. 2). From these

results, we are able to discriminate soybean genotypes based on the pod shape.

Cluster analysis

Cluster analysis was performed using UPGMA method based on the principal component scores (Fig. 3). Apparently, soybean genotypes are classified into four clusters. Cluster I involved three soybean genotypes with long two-seeded pod shape, with maturity over three months, with high branching, with high number of seeds and large seed size (26–27 g per 100 seed). Cluster II consisted of six soybean genotypes having same shape in one and three-seeded pod, plenty of branch number, late flowering and maturity. Cluster III was characterized with large size of three-seeded pod shape, late maturity (over three months), relative high plant seed yield. Cluster IV included six soybean genotypes that showed the same size of two and three seeded pod shape, early flowering time (before two months after sowing), high number of branches and seeds. It is suggested that pod shape can be used as a qualitative marker to identify and classify varieties.

DISCUSSION

Fourier descriptor was applied extensively for morpholog-

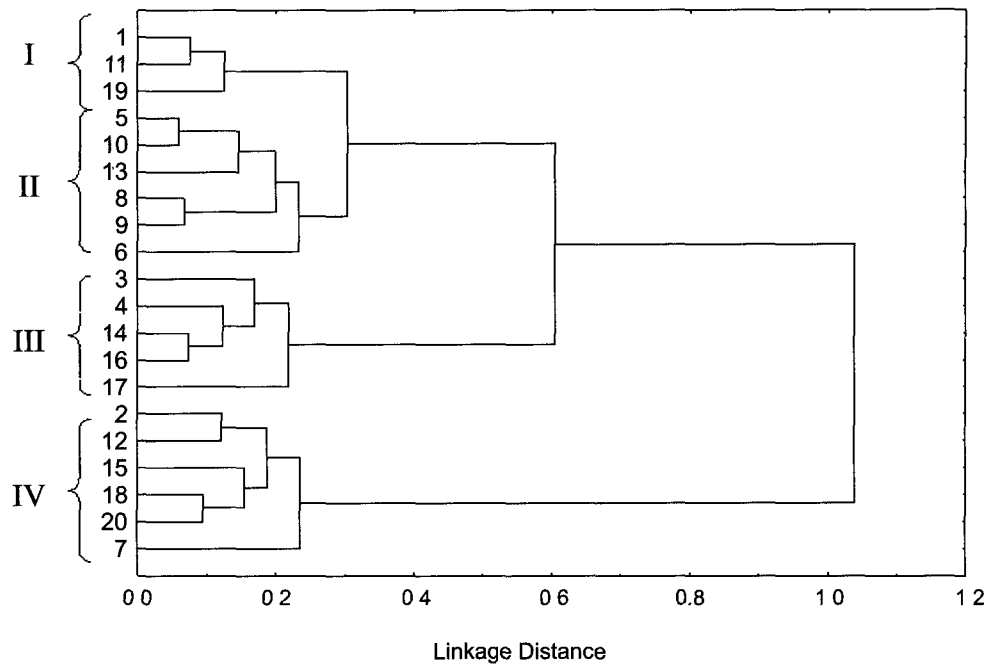


Fig. 3. The dendrogram of twenty soybean genotypes differentiated by using principal component scores of three pod types. (1; IT102668, 2; IT108886, 3; IT101280, 4; IT104411, 5; IT104535, 6; IT104554, 7, IT141548, 8, IT141594, 9; IT100826, 10; IT113446, 11, IT113141, 12; IT103906, 13; IT101075, 14; IT100914, 15; IT104704, 16; IT103788, 17, IT105686, 18; IT100804, 19; IT1088109, 20, Sinpaldalkong 2)

ical analysis. In fact, 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 coefficient, because there are usually too many coefficients to be considered at once and it is almost impossible to understand the effect of each coefficient on shape. In this study, it was found that the first five principal component scores represented more than 95% of the soybean pod shape expressed by 77 Fourier coefficients. Moreover, the effect of each principal component on the shape by redrawing the contours with the Fourier coefficients can be estimated for the cases that the principal component score took some typical values. That is, we could successfully partition the shape information of soybean pod 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 pod shape.

The analysis of variance and multiple comparisons indicated the large genotypic differences in the first, the second, and the fifth principal components. Because the variations of those principal components were continuous, the effects of the polygenes for the shape were also suggested.

Finally, the features of principal components are fully data-dependent. That is, the shape features represented by each principal component 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 pod shape as possible. For example, the development of a database for the standardized Fourier coefficients of soybean pods can be strongly suggested. Whenever a new data set is available, the data set should be added to the database and should be subject to a principal component analysis using the whole data set in the database repeatedly, so that the above problem gets solved gradually as database grows.

ACKNOWLEDGMENTS

This work was supported by a grant from BioGreen 21 Program, Rural Development Administration, Republic of Korea. Ngon T. Truong thanks his colleagues in Can Tho University for their covering research and teaching works during his PhD study in Korea.

REFERENCES

- 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. Ghani. 1989. Elliptic Fourier analysis of cell and nuclear shape. *Comp. Biomed Res.* 22 : 405-414
- Ehrlich, R. and B. Weinberg. 1970. An exact method for characterization of grain shape. *J. Sed. Pet.* 40 : 205-212
- Ferson, S., F. J. Rohlf, and R. K. Koehn. 1985. Measuring shape variation of two dimensional outlines. *Syst Zool* 34 : 59-68.
- Furuta, N., S. Ninomiya, N. Takahashi, H. Ohmori, and Y. Ukai. 1995. Quantitative evaluation of soybean (*Glycine max* L. Merr.) leaflet shape by principal component scores based on elliptic Fourier descriptor. *Breed Sci* 45 : 315-320.
- Giardina, C. R. and F. P. Kuhl. 1977. Accuracy of curve approximation by harmonically related vectors with elliptical loci. *Comp. Grap. Ima. Proc.* 6 : 277-285.
- Granlund, G. H. 1972. Fourier preprocessing for hand print character recognition. *IEEE. Trans. Comput.* C-21 : 269-281. <http://www.ipgri.cgiar.org/system>. 1998.
- Iwata, H. and Y. Ukai. 2002. SHAPE: A computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors. *J. of Hered.* 93 : 384-385.
- Iwata, H., H. Nesumi, S. Ninomiya, Y. Takano, and Y. Ukai. 2002. The evaluation of genotype x environment interactions of citrus leaf morphology using image analysis and elliptic Fourier descriptors. *Breeding Sci.* 52 : 243-251.
- Iwata, H., H. Nesumi, S. Ninomiya, Y. Takano, and Y. Ukai. 2002. Diallel analysis of leaf shape variations of citrus varieties based on elliptic Fourier descriptors. *Breeding Sci.* 52 : 89-94.
- Iwata, H., S. Niikura, S. Matsuura, Y. Takano, and Y. Ukai. 2000. Diallel analysis of root shape of Japanese radish (*Raphanus sativus* L.) based on elliptic Fourier descriptors. *Breeding Sci* 50 : 73-80.
- Jaccard, P. 1908. Nouvelle recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.* 44 : 223-270.
- Keefe, P. D. and S. R. Draper. 1986. An automated machine vision system for the morphometry of new cultivars and plant gene bank accessions. *Plant Varieties and Seeds* 1 : 1-11
- Kuhl, F. P. and C. R. Giardina. 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* 18 : 236-258.
- Matano, T. and A. Ujihara. 1973. Agroecotypes of *Fagopyrum* II. Classification of common buckwheat (*Fagopyrum esculentum* Moench.) by kernel shape and these geographical distributions in Japan. *Jpn. J. Crop Sci.* 42 : 29-30.
- Myers, D. G. and K. J. Edsall. 1989. The application on image processing techniques to the identification of Australian Wheat Varieties. *Plant Varieties and Seeds* 2 : 109-116
- Neuman, M., H. D. Saperstein, E. Shwedy, and W. Bushuk. 1987. Discrimination of wheat class and variety by digital image analysis of whole grain samples. *J. Cereal Sci.* 6 : 125-132
- Ninomiya, S., R. Ohsawa, and M. Yoshida. 1995. Evaluation of buckwheat and tartary buckwheat kernel shape by elliptic Fourier method. In: T. Matano & Ujihara (Eds.), *Current Advances in Buckwheat Research*, pp. 389-396. Shinshu Univ. Press
- 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.
- Sneath, P. H. A. and R. R. Sokal. 1973. *Numerical Taxonomy*. WH Freeman San Francisco.
- Thseng, F. S., S. J. Tsai, J. Abe, and S. T. Wu. 1999. *Glycine formosana* Hosokawa in Taiwan : pod morphology, allozyme, and DNA polymorphism. *Bot. Bull. Acad. Sin.* 40 : 251-257.
- USDA. 2001. *Soybean Germplasm Collection Descriptors*.
- White, R. H., C. Rentice, and T. Verwist. 1988. Automated image acquisition and morphometric description. *Can. J. Bot.* 66 : 450-459.
- Yoshida, M., S. Ninomiya, M. Oide, M. Hagiwara, A. Ujihara, and T. Matano. 1995. Geographical variation of kernel shape on tartary buckwheat from the world. In: T. Matano & A. Ujihara (Eds.) *Current Advances in Buckwheat Research*, 397-404. Shinshu Univ Press