Indian Journal of Agricultural Sciences **84** (6): 765–9, June 2014/Short communication
https://doi.org/10.56093/ijas.v84i6.41485

Characterization of Cucumber (*cucumis sativus*) genotypes through principle component and regression analyses

RAMESH KUMAR[1], SANDEEP KUMAR[2], DHARMINDER KUMAR[3] and R K GUPTA[4]

Dr Y S Parmar University of Horticulture and Forestry, Nauni, Solan, Himachal Pradesh 173 230

Received: 20 May 2012; Revised accepted: 20 March 2014

Key words: Cucumber, *Cucumis sativus,* Genetic diversity, Principal component, Regression analysis

Cucumber (*Cucumis sativus* L.) is one of the most important cucurbitaceous vegetable crops grown extensively in tropical and sub-tropical parts of the country. It is grown for its tender fruits, which are consumed either raw as salad, cooked as vegetable or as pickled in its immature stage. The genetic improvement of any crop mainly depends upon the amount of genetic variability present in the population and India, being the primary centre of origin of cucumber, has accumulated a wide range of variability. Because of the genetic diversity present in cucumber, there is an opportunity to select superior genotypes.

Indian subcontinent has a rich and varied heritage of genetic resources but these resources have not been exploited fully due to their inherent problems of large size and lack of sufficient evaluation and classification. Knowledge of association of various characters provides the basis of selection for yield and its components for crop improvement. Since yield is a complex quantitative trait, simple correlation and regression of characters provides limited insight into the association of various traits to yield. Few reports are available on phenotypic variability, correlation and path analysis in cucumber (Afangideh and Uyoh 2007, Kumar *et al.* 2008, Hanchinamani and Patil 2009, Hossain *et al.* 2010). Principle component analysis (PCA) helps in identifying the most relevant characters that can be used as descriptors by explaining as much of total variation in the original set of variables as possible with as few components as possible and reducing the dimension of the problem. Therefore, an attempt has been made in the present investigation to access and analyze the extent of genetic diversity through principle component and regression analyses for yield improvement in cucumber.

The present investigations were carried out at Experimental Research Farm of the Department of Vegetable Science, Dr Y S Parmar University of Horticulture and Forestry, Nauni, Solan (HP) during *kharif* 2009-10. The experimental material consisted of diverse group of 30 genotypes of cucumber, including check cultivars, i e K 75 and K 90 (Table 1). The experiment was laid out in Randomized Complete Block Design (RCBD) with three replications of each genotype. Standard cultural practices recommended in the package of practices for vegetable crops, were followed to ensure a healthy crop stand (Anonymous 2009). Seeds were directly sown in the field and three to four seeds per basin were sown at a spacing of 100 cm × 75 cm in a plot having size of 3.0 m × 2.25 m. After the emergence of seedlings, only one healthy seedling per hill was retained. The observations were recorded on node number bearing first female flower (x_1), number of marketable fruits/plant (x_2), fruit length (x_3), fruit breadth (x_4), average fruit weight (x_5), days to marketable maturity (x_6), harvest duration (x_7), total soluble solids (x_8), seed germination (x_9), seed vigour index-I (x_{10}) and II (x_{11}) and

Table 1 List of cucumber genotypes studied along with their sources

Genotype	Source
LC 1, LC 2, LC 3, LC 4, LC 5, LC 6	Hamirpur, Himachal Pradesh, India
LC 7, LC 8	Bilaspur, Himachal Pradesh, India
LC 9, LC 10, LC 11, LC 12	Kangra, Himachal Pradesh, India
LC 13, LC 14, LC 15	Mandi, Himachal Pradesh, India
LC 16, LC 17	Shimla, Himachal Pradesh, India
LC 18, LC 19, LC 20	Kullu, Himachal Pradesh, India
LC 21, LC 22	Chamba, Himachal Pradesh, India
LC 23, LC 24, LC 25	Solan, Himachal Pradesh, India
LC 26, LC 27	Una, Himachal Pradesh, India
LC 28	Jammu, Jammu & Kashmir, India
K 75*, K 90*	UHF, Nauni, Solan, Himachal Pradesh, India

*Check cultivar

[1] Professor (e mail: rameshkbhardwaj@rediffmail.com), [2] Ph D Vegetable Science (e mail: sandeepkdhatwalia@gmail.com), [3] Assistant Professor (e mail: dharmruder@gmail.com), Department of Vegetable Science; [4] Professor (e mail: guptark@gmail.com), Department of Basic Science

severity of powdery mildew (x_{12}), anthracnose (x_{13}) and angular leaf spot (x_{14}) and yield/plot (x_{15}) from five randomly selected plants in each replication for all characters except for fruit characters for which observations were recorded on ten randomly selected fruits per replication. Seed germination of each genotype was tested in accordance with ISTA (Anonymous 1985) and seed vigour index-I and II were calculated as per the formula given by Abdul-Baki and Anderson (1973). The disease severity of anthracnose and angular leaf spot was recorded on 0-5 scale as suggested by Bhat (2007) and disease severity for powdery mildew was recorded by adopting the scale given by Ransom *et al.* (1991).

The data were subjected to analysis of variance as per the procedure described by Gomez and Gomez (1983). The treatment means were tested at 5 per cent or 1 per cent level of significance. The calculated F-value was compared with tabulated F-value. When F-test was found significant, standard error and critical differences were calculated to find out the superiority of one entry over the others. If the number of the variables (p) are measured for each observation, then 'p' separate univariate statistical analyses are required. These analyses apply only to the individual components of the factor, not to the factor itself. Furthermore, the variables are usually highly inter-correlated since biological systems, being complex and highly integrated; contain a great number of interacting components which are interrelated. Consequently these variables should not be treated as independent components of the factor in question in statistical analyses. Principal component analysis (Hotelling 1933) restructures the data so that a general factor can be measured by 'p' correlated variables and could be expressed in terms of n<p uncorrelated variables would be highly desirable. The first few components usually account for most of the variation of the original variables. Contribution of different characters towards the divergence was estimated with the help of principle component analysis in accordance with Lawley and Maxwell (1963). Multiple linear regression equation was used to predict average fruit weight and yield per plot.

Significant differences were observed among all the genotypes for all the characters under study (Table 2). Among the horticultural traits, comparatively wide range was observed for node number bearing first female flower (3.53-13.53) and days to marketable maturity (55.67-78.33), which determine the earliness of a variety. Fruit length,

Table 2 Estimates of range, mean, standard error of difference, phenotypic and genotypic coefficient of variation and loadings of different characters on the first four principal components in cucumber

Characters	Range		Mean	± SE(d)	Coefficients of variability (%)		Critical difference	Principle component*			
	Maximum	Minimum			Phenotypic	Genotypic		PC_1#	PC_2	PC_3	PC_4
Node number bearing first female flower	3.53	13.53	8.63	1.88	39.12	28.63	3.76	–0.664	0.300	0.486**	0.339
Number of marketable fruits per plant	5.01	8.57	6.64	0.44	18.00	16.06	0.88	0.939	–0.129	–0.132	-0.007
Fruit length (cm)	8.11	22.76	14.13	0.81	21.63	20.47	1.61	0.526	0.507	–0.276	0.406
Fruit breadth (cm)	3.08	7.18	5.01	0.24	15.71	14.57	0.48	0.362	0.741	0.327	-0.195
Average fruit weight (g)	95.00	430.00	249.63	20.80	27.19	25.20	41.65	0.664	0.712	0.027	0.020
Days to marketable maturity	55.67	78.33	71.20	3.20	9.93	8.26	6.40	–0.542	0.303	0.527	0.485
Harvest duration (days)	14.00	28.67	19.36	2.68	27.44	21.59	5.36	0.901	0.023	–0.030	0.039
TSS (^{0}B)	2.03	4.07	2.75	0.28	22.85	19.21	0.55	0.794	-0.233	–0.129	0.204
Seed germination (%)	61.00	87.67	70.89	2.88	12.30	11.24	3.83	0.651	-0.448	0.460	0.098
Seed vigor index-I	438.33	1930.00	1002.12	253.15	53.34	43.45	506.80	0.829	-0.246	0.351	0.026
Seed vigor index-II	1642.28	3167.28	2250.03	253.84	24.04	19.67	508.18	0.778	-0.263	0.336	-0.023
Severity of powdery mildew (%)	8.50	29.40	19.62	3.57	34.00	25.66	0.88	–0.648	0.339	0.016	-0.456
Severity of anthracnose (%)	7.70	26.20	15.20	3.52	38.02	25.29	1.04	–0.482	–0.161	–0.643	0.438
Severity of angular leaf spot (%)	6.50	18.30	11.62	2.60	36.44	23.96	0.83	–0.377	–0.284	0.656	0.038
Yield per plot (kg)	4.37	27.31	15.14	1.00	36.12	35.20	2.01	0.903	0.416	–0.037	0.005
Eigen value								7.231	2.292	2.026	1.009
Percentage of variance								48.204	15.279	13.507	6.729
Cumulative % of variance								48.204	63.483	76.990	83.719

#PC: Principal component, *Extracted through principle component analysis, **Bold value indicates the highest Eigen Vector for the corresponding trait amongst the four principal components

breadth and weight are the major yield contributing traits, wide variations were observed with respect to these traits (8.11-22.76 cm, 3.08-7.18 cm, 95.00-430.00 g, respectively). Tremendous variations with respect to number of marketable fruits per plant (5.01-8.57), harvest duration (14.00-28.67 days), total soluble solids (2.03-4.07 ^{0}B) and yield per plot (4.37-27.31 kg) were obtained. Wide variations with respect to various horticultural characters were also reported earlier by Kumar *et al.* (2008), Hanchinamani *et al.* (2008), Yogesh *et al.* (2009) and Hossain *et al.* (2010) in cucumber. For seed characters, *viz.* seed germination (61.00–87.67%), seed vigor index-I (438.33–1930.00) and seed vigor index-II (1642.28–3167.28), a wide variation was observed. All the genotypes studied, responded differently to the attack of different diseases, *viz.* powdery mildew (8.50-29.40%), anthracnose (7.70-26.20%) and angular leaf spot (6.50-18.30%). These findings are in agreement with Wehner and Shetty (2000). These wide variations in the genotypes for different characters would help in selecting the best genotypes from existing collections.

The estimates of phenotypic and genotypic coefficients of variability gave a clear picture of amount of variations present in the available germplasm (Table 2). For all the characters studied, phenotypic coefficients of variability were higher in magnitude than genotypic coefficients of variability, though difference was not much in all the cases. Thus, showing that these traits are not much influenced by environmental factors. Genotypic coefficient of variability (GCV) was ranged from 8.26 to 43.45% with maximum value for seed vigor index-I followed by yield per plot. This reflects greater genetic variability among the genotypes for these characters for making further improvement by selection. Whereas, moderate GCV were recorded for node number bearing first female flower, severity of powdery mildew, anthracnose and angular leaf spot, average fruit weight, harvest duration, fruit length, seed vigor index-II, total soluble solids, number of marketable fruits per plant and for fruit breadth. Similar results had also been reported by Yogesh *et al.* (2009) for different estimates of variability in cucumber.

The characters contributing more to the divergence gave greater emphasis for deciding on the cluster for the purpose of further selection and the choice of parents for hybridization (Jagadev *et al.* 1991). The results of principal component analysis indicated that the first four components account for the maximum explained variations (83.72%). Factor analysis was applied to extract the basic factors underlying the observed traits of cucumber. The factors were extracted individually on the basis of eigen values (Table 2) and revealed the pattern and principal component analysis of the data. The first four components having eigen values greater than 1 were retained in the analysis because of the substantial amount of the variations. The factors corresponding to eigen values less than 1 were not considered. These factors were ignored due to "Guttmans lower bound principle" according to which eigen values less than unity (λ<1) should be ignored (Kaiser 1958). The orthogonal factors were also extracted. The centroid method of analysis (Lawley and Maxwell 1963) was used. The four factors were obtained for the estimation of components as detailed follows:

Factor 1: $0.93X_2 + 0.52X_3 + 0.90X_7 + 0.79\ X_8 + 0.65\ X_9 + 0.82\ X_{10} + 0.77\ X_{11} + 0.90X_{15}$

Factor 2: $0.74X_4 + 0.71X_5 + 0.33X_{12}$

Factor 3: $0.48X_1 + 0.52X_6 + 0.65X_{14}$

Factor 4: $0.43X_{13}$

The first four factors had the variances of 7.231, 2.292, 2.026 and 1.009 with 48.204, 15.279, 13.507 and 6.729% of total variation, respectively and aggregating to 83.72% of total variation. The first factor extracted had the combination of number of marketable fruits per plant (x_2), fruit length (x_3), harvest duration (x_7), total soluble solids (x_8), seed germination (x_9), seed vigour index-I (x_{10}) and II (x_{11}) and yield per plot (x_{15}). The first factor had high positive loadings for all variables except fruit length. The second factor was a combination of fruit breadth (x_4), average fruit weight (x_5) and severity of powdery mildew (x_{12}). The third factor accounted for the combination of node number bearing first female flower (x_1), days to marketable maturity (x_6) and angular leaf spot (x_{14}) and the fourth factor accounted for the only one variable, i e, anthracnose (x_{13}). The positive value of different characters under study in different components indicated its importance in divergence among 30 genotypes of cucumber, whereas negative values showed the lowest contribution to the divergence (Table 2). Loading of different variables based on first two principle components indicates that fruit length (x_3), fruit breadth (x_4), average fruit weight (x_5), harvest duration (x_7) and yield per plot (x_{15}) are the main components of divergence between 30 genotypes of cucumber, whereas contribution of anthracnose (x_{13}) and angular leaf spot (x_{14}) was found least in divergence (Fig 1). Hence, main emphasis should be given on the fruit length and breadth, average fruit weight, number of marketable fruits per plant and harvest duration to increase the fruit yield of cucumber

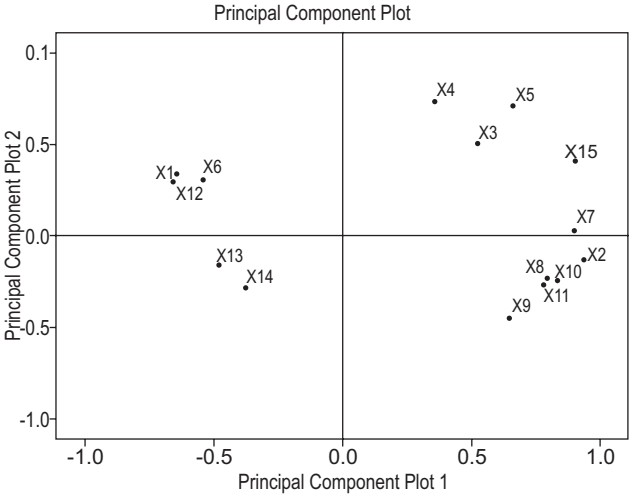

Fig 1 Loading of different characters based on first two principle components

crop. Similar findings have also been reported for cucumber, peppers and okra by Zhang and Cui (1993), Portis *et al.* (2006) and Koutsos *et al.* (2000), respectively. The factors could be used for further breeding programs for exploiting the hybrid vigor for higher fruit yield.

To estimate the average fruit weight and yield per plot, step wise multiple linear regression models were extracted with average fruit weight (x_5) and yield per plot (x_{15}) as dependent variables and remaining characters as independent variables. The prediction of average fruit weight (Model I) and yield per plot (Model II) with values in parentheses indicating standard errors of the regression coefficients have been given in the Table 3. Average fruit weight (Model I) can be best predicted by number of marketable fruits per plant (x_2), fruit length (x_3) and fruit breadth (x_4). These characters had positive effects on estimation of average fruit weight. The coefficient of determination (R^2) for both the models was high, *viz.* 0.93 (Model I) and 0.99 (Model II). It means that ninety nine per cent of total variation in average fruit weight and ninety three per cent of yield per plot was influenced by these characters. Therefore, use of number of marketable fruits per plant, fruit length, average fruit weight and harvest duration was the best model for predicting yield per plot. Relative importance of regression model for predicting the total yields have also been reported earlier by Kumar *et al.* (2011) in carrot. Moreover, there is an opportunity for improvement through hybridization and selection due to genetic diversity in the available germplasm. Therefore, these models can be used best for genetic improvement of cucumber by the development of new varieties/hybrids.

SUMMARY

Thirty diverse genotypes of cucumber (*Cucumis sativus* L.) collected from different indigenous sources were characterized with respect to economically important traits by using principal component and regression analyses during *kharif* 2009-10. The effect and contribution of each character on fruit yield per plot was measured. Principal component analysis characterized the genotypes into four principal components based on their total variation (83.72%). The first principal component accounted for more than 48% of the total variation and was the combination of number of marketable fruits per plant, fruit length, harvest duration, total soluble solids, seed germination, seed vigour index-I and II and yield per plot. The second, third and fourth principle components contributed only 15.27%, 13.50% and 6.72% of total variations, respectively. To quantify the importance of each variable in predicting average fruit weight and yield per plot, multiple linear regression models were developed. Model-I indicated that average fruit weight can be predicted satisfactorily on the basis of number of marketable fruits per plant, fruit length and breadth, while, Model-II indicated that yield per plot can be best predicted by with the help of number of marketable fruits per plant, fruit length, average fruit weight, harvest duration, seed germination, seed vigour index-II and severity of powdery mildew and anthracnose. Therefore, on the basis of information on genetic diversity through principal component and regression analyses, suitable selection strategy can be formulated for getting higher yield in cucumber.

REFERENCES

Abdul-Baki A A and Anderson J D. 1973. Vigour germinated in soybean seed by multiple criteria. *Crop Science* **13**: 630–3.

Afangideh U and Uyoh E A. 2007. Genetic variability and correlation studies in some varieties of cucumber (*Cucumis sativus* L.). *Jordan Journal of Agricultural Sciences* **3** (4): 376–84.

Anonymous. 1985. International rules for seed testing. *Seed Science and Technology* **13**: 293–353.

Anonymous. 2009. *Package of Practices for Vegetable Crops*. Directorate of Extension Education, Dr Y S Parmar University of Horticulture and Forestry, Nauni, Solan, p 202.

Bhat N A, Masoodi S D, Ahmad M and Zargar M Y. 2007. Occurrence and severity of angular leaf spot of cucumber in Kashmir. *Annals of Plant Protection Sciences* **15** (2): 410–3.

Gomez K A and Gomez A A. 1983. *Statistical Procedures for Agricultural Research,* pp 357–427. John Wiley and Sons Inc., New York.

Hanchinamani C N and Patil M G. 2009. Correlation studies in cucumber (*Cucumis sativus* L.). *Asian Journal of Horticulture* **4**(1): 121–5.

Hanchinamani C N, Patil M G, Dharmatti P R and Mokashi A N. 2008. Studies on variability in cucumber (*Cucumis sativus* L.). *Crop Research* **36** (1-3): 273–6.

Hossain M F, Rabbani M G, Hakim M A, Amanullah A S M and Ahsanullah A S M. 2010. Study on variability character association and yield performance of cucumber (*Cucumis sativus* L.). *Bangladesh Research Publications Journal* **4** (3): 297–311.

Hotelling H. 1933. Analysis of complex statistical variables into principal components. *Journal of Education Psychology* **24**: 417–41.

Jagdev P N, Samal K M and Lenka L. 1991. Genetic divergence in rape mustard. *Indian Journal of Genetics* **51**: 465–6.

Kaiser H F. 1958. The varimax criteria for analytic rotation in factor analysis. *Psychomethka* **23**: 187–200.

Koutsos T V, Koutsika S M, Gouli V E, Tertivanidis K. 2000. Study of genetic relationship of Greek okra cultivars (*Abelmoschus esculentus* (L.) Moench.) by using agronomic traits, heterosis and combining ability. *Journal of Vegetable Crop Production* **6** (1): 25–35.

Kumar A, Kumar S and Pal A K. 2008. Genetic variability and characters association for fruit yield and yield traits in cucumber. *Indian Journal of Horticulture* **65** (4): 423–8.

Kumar R, Vashisht P and Gupta R K. 2011. Characterization of European carrot genotypes through principal components and regression analyses. *International Journal of Vegetable Science* **17** (1): 3–12.

Lawley D N and Maxwell A E. 1963. *Factor Analysis as a Statistical Method,* p 117. Butterworth, London.

Portis E, Nervo G, Cavallanti F, Barchi L and Lanteri S. 2006. Multivariate analysis of genetic relationships between Italian pepper landraces. *Crop Science* **46**: 2 517–25.

Ransom L M, Briens R G O and Glass R J. 1991. Chemical control of powdery mildew in green peas. *Australian Plant Pathology* **20** (1): 16–20.

Wehner T C and Nischit V S. 2000. Screening the cucumber germplasm collection for resistant to gummy stem blight in North Carolina field tests. *Hort Science* **35**(6): 1 132–40.

Yogesh C, Yadav S K, Brijpal B and Dixit S K. 2009. Genetic variability, heritability and genetic advance for some traits in cucumber. *Indian Journal of Horticulture* **66** (4): 488–91.

Zhang M and Cui H W. 1993. Application of factor analysis to cucumber breeding. *Cucurbit Genetics Cooperative* **16**: 27–9.