Euphytica **43**: 97–103, 1989.

Geographic distribution of variation in quantitative traits in a world lentil collection

W. Erskine, Y. Adham and L. Holly
International Center for Agricultural Research in the Dry Areas (ICARDA), P.O. Box 5466, Aleppo, Syria

Received 4 January 1988; accepted in revised form 27 June 1988

Key words: Lens culinaris, lentil, variation, germplasm, geographical distribution

Summary

In a world lentil collection the distribution of variation amongst accessions from 13 major lentil-producing countries was examined on the basis of nine quantitative morphological characters by discriminant analysis and canonical analysis. Stepwise discriminant analysis revealed major differences between accessions from different countries. Three major regional groups were apparent: 1) a levantine group (Egypt, Jordan, Lebanon and Syria, 2) a more northern group composed of Greece, Iran, Turkey, and USSR, and 3) accessions from India and Ethiopia with strikingly similar quantitative morphological characters. Misclassifications of individuals within groups were frequent. Characters useful in discriminating between accessions from different countries were in descending order of importance: time to maturity, lowest pod height and 100-seed weight. The regional grouping indicates the importance of local adaptation through clusters of associated characters with phenological adaptation to the ecological environment as the major evolutionary force in the species.

Introduction

A comprehensive study of variation in the lentil (*Lens culinaris* Med.) was undertaken by Barulina (1930), who examined the large collections made by many extensive Russian expeditions under the auspices of N.I. Vavilov. She examined the pattern of geographic variation in detail and classified cultivated forms into the subspecies *macrosperma* and *microsperma* on the basis of a suite of related qualitative and quantitative characters, which were relatively insensitive to environmental conditions. Barulina also described the geographic distribution of associated characters into regional groups or *grex* on the basis of several qualitative traits. This classic work has no antecedents.

The International Center for Agricultural Research in the Dry Areas (ICARDA) has assembled a large lentil germplasm collection with a view firstly to its conservation and secondly to its exploitation by breeding. As a pre-requisite to breeding, a total of 4036 accessions were evaluated for a range of morphological characters from 1978 to 1982. A listing of the characteristics of the accessions and univariate analysis of the traits were published in a germplasm catalogue (Erskine & Witcombe, 1984). Analyses of the variation in the collection have been made for an aspartate amino-transferase locus (Skibinski et al., 1984) and between and within accessions collected within one area North Yemen (Erskine & Choudhary, 1986). In this study we examine the differentiation between populations from major growing countries for a range of quantitative morphological characters using discriminant analysis and canonical variate analysis.

Materials and methods

A total of 1370 accessions of lentil were evaluated at ICARDA Tel Hadya (35° 55′ N, 36° 55′ E) during the 1978-79 season. They comprised a random selection of the ICARDA world collection arranged in an augmented randomized complete block design. Sowing was in November 1978 into dry soil, and germination started after the first rain on 30 November. Seed was drilled in 6 row plots, 5 m-long at a rate of 200 seeds/m^2 with rows 25 cm apart. Fertilizer was applied at the rate of 50 kg P_2O_5/ha. Inoculation was not undertaken but nodulation was adequate throughout. During the growing season only 247 mm of rain were received but its distribution was even.

The following characters were measured during growth, as described in Erskine & Witcombe (1984): time to 50% flowering (days), time to 90% pod maturity (days), plant height (cm), and lowest pod height (cm). After harvesting the central 4 cm^2 of plots, measurements were made on biological yield (kg/ha), seed yield (kg/ha), 100 seed weight (g), and seed number per pod. Harvest index was calculated as seed yield divided by biological yield.

A subset of the 1370 accessions was used for statistical analysis comprising those accessions from countries represented by 10 or more accessions with an upper limit of 100 accessions per country. India and Iran, originally over the limit, were represented by 100 randomly chosen accessions. The subset for analysis was composed of 615 accessions from a total of 13 major lentil producing countries (Table 1). The unadjusted plot data were used in the analysis of the nine quantitative characters.

Stepwise discriminant analysis on the basis of country of origin was undertaken to test if origin affects the assortment of lentil variability in a multivariate sense. The stepwise discriminant analysis was performed using the program BMDP7M (Jennich & Sampson, 1985). Characters for inclusion were selected stepwise to minimize Wilk's lambda between groups. An approximate F ratio provided

Table 1. The mean (X) and standard deviation (SD) of accessions from different countries for each character, together with the number of accessions from each country of origin

Country	No. acc.	Time to flower days		Time to maturity days		Plant height cm		Lowest pod height cm		Biological yield kg/ha		Seed yield kg/ha		Harvest index		100 seed weight g		Seed number per pod	
		X	SD	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD	X	SD
Afghan.	32	142	6.6	172	4.9	27.3	3.7	14.0	2.6	2432	560	666	310	0.27	0.10	2.23	0.40	1.59	0.19
Chile	49	133	5.4	174	4.6	29.5	3.4	18.3	2.9	3754	1076	709	338	0.19	0.07	4.65	1.37	1.26	0.25
Egypt	41	121	9.7	165	5.5	32.8	4.3	15.5	2.6	4133	1660	1230	434	0.32	0.11	3.57	0.45	1.62	0.20
Ethiopia	98	117	4.8	155	3.0	24.8	2.5	11.3	2.1	2506	821	1126	388	0.45	0.06	3.00	0.28	1.52	0.15
Greece	17	136	4.6	172	3.5	30.9	2.6	18.8	2.8	4147	929	817	383	0.19	0.06	3.02	1.15	1.41	0.24
India	100	119	3.4	155	2.9	22.9	2.1	9.8	1.7	3236	1056	1339	447	0.42	0.04	2.60	0.56	1.59	0.20
Iran	100	136	7.2	174	4.8	27.6	3.5	15.0	2.6	2837	896	767	290	0.27	0.06	3.30	0.94	1.55	0.20
Jordan	19	121	1.9	164	5.7	29.5	2.0	15.0	2.5	4912	1388	1803	607	0.36	0.05	4.63	0.53	1.22	0.16
Lebanon	25	128	6.0	167	4.7	30.0	2.6	16.0	2.9	4291	1396	1231	600	0.28	0.08	4.41	0.99	1.33	0.20
Pakistan	14	129	5.7	167	4.6	27.4	4.6	14.1	2.7	3269	939	1027	304	0.32	0.05	2.90	0.55	1.57	0.21
Syria	31	125	4.1	164	3.9	30.1	3.0	17.4	2.7	4901	1481	1558	526	0.32	0.06	4.72	1.75	1.34	0.27
Turkey	77	135	5.3	172	4.8	30.1	3.6	18.1	3.2	4019	1090	978	383	0.24	0.06	3.49	1.05	1.48	0.21
USSR	12	137	4.0	172	4.4	31.3	4.3	18.7	4.2	3112	1146	726	443	0.22	0.09	3.03	1.05	1.52	0.21
Overall mean		128	5.7	166	4.3	27.6	3.2	14.5	2.6	3414	1085	1062	406	0.32	0.07	3.37	0.88	1.49	0.20
F value[a]		113.8		173.6		45.6		73.2		24.4		25.7		84.5		34.0		16.3	

[a] F value of one-way analysis of variance by country of origin (d.f. 12,602)

a test of significance of the Mahalanobis distance between group centroids. A measure of goodness of the classification procedure consists of the probabilities of misclassification. This was done using the jacknife procedure (Afifi & Clark, 1984) which omits one observation from the first group and computes the discriminant function on the basis of the remaining observations. The excluded observation is then classified. The procedure is repeated for each observation. The proportion of misclassified individuals is the jacknife estimate of the probability of misclassification.

Canonical analysis (Seal, 1964), which has been used by plant breeders (Riggs, 1973; Phillips & Powell, 1984) and evaluators of germplasm (Damania & Porceddu, 1983), was also applied to the data. This is a technique by which multivariate data are represented on orthogonal axes such that maximum discrimination is obtained between groups when tested against the variation within groups.

Results

The means and standard deviations for the accessions from each country are shown in Table 1. Analysis of variance by country of origin gave highly significant F-ratios for each character varying from a low of 16.3 for seed number per pod to a high of 173.6 for time to maturity with 12 and 602 degrees of freedom. Clearly there were major differences among the country means for each trait.

Stepwise discriminant analysis on the basis of country of origin resulted in an approximate F-ratio of 24.0 with 108 and 4336 degrees of freedom (Table 2). This shows that on average populations from some countries were strikingly different in a multivariate sense. The stepwise method ranks the characters in order of their inclusion within the discriminant function; this is the order of their usefulness in the classification by country of origin. The most important trait was time to maturity, and in descending order of importance the other characters were lowest pod height, 100-seed weight, time to flower, biological yield, plant height, seed yield, harvest index, and seed number per pod. The discriminant or classification functions for each country are given in Table 3.

The jacknife classification tests the robustness of the classification by the percentage of misclassifications. The percentage of correct classifications varied from only 8% for Lebanese accessions to 86% for accessions from Ethiopia with an overall mean of 55% correct classifications (Table 4). These values reflect the distinctness of accessions from a particular country from all other. For example, the Ethiopian accessions were rarely misclassified; whereas there were many misclassified Lebanese accessions which were wrongly classified as Syrian and vice versa. Another group of countries which were difficult to classify were Iran, Turkey and USSR; there were many misclassifications between these three countries reflecting the similarity of the genetic material.

Table 2. Stepwise order of inclusion of variables within the discriminant analysis together with the respective approximate F-statistics and Wilk's Lambda

Step no.	Variable	Wilk's Lambda	Approximate F-statistic	Degrees of freedom for F	
1	Time to maturity	0.224	173.6	12	602
2	Lowest pod height	0.140	83.6	24	1202
3	100 seed weight	0.098	58.9	36	1773
4	Time to flower	0.075	46.1	48	2309
5	Biological yield	0.060	38.6	60	2803
6	Plant height	0.048	33.7	72	3253
7	Seed yield	0.041	29.7	84	3658
8	Harvest index	0.037	26.5	96	4018
9	Seed number per pod	0.033	24.0	108	4336

Canonical analysis reduces the multivariate (9 variable) data into fewer orthogonal axes. In this case the first three significant canonical variates accounted for 90.0% of the total variation. The first two variates accounted for 84.6% of the variability and are shown as a scatter diagram (Fig. 1). The first variate represents a separation based mainly on size (biomass) and both flowering and maturity which were all closely correlated. In Fig. 1 accessions from Ethiopia and India have low values for axis 1 as they are early to flowering and maturity and have low biological yields; whereas the population means for Afghanistan, Chile, Greece, Iran, Turkey and USSR are at the other extreme with high values of variate 1 representing their high biomass production and late flowering and maturi-

Table 3. Classification functions for each country for nine characters used in discriminant analysis

Country	Character									Constant
	Time to flower	Time to maturity	Plant height	Lowest pod height	Biological yield	Seed yield	Harvest index	100 Seed weight	Seed no per pod	
Afghanistan	1.89	10.10	3.58	− 1.47	0.01	− 0.02	381.08	9.83	73.47	− 1171.6
Chile	1.62	10.25	3.19	− 0.93	0.02	− 0.03	381.05	11.86	71.76	− 1167.3
Egypt	1.25	10.19	3.81	− 1.44	0.01	− 0.03	390.95	10.15	78.01	− 1124.9
Ethiopia	1.33	9.60	3.08	− 1.42	0.02	− 0.03	427.36	9.90	71.84	− 1019.1
Greece	1.66	10.11	3.39	− 0.81	0.02	− 0.03	380.64	9.51	70.66	− 1148.3
India	1.30	9.69	2.87	− 1.65	0.01	− 0.03	402.79	9.38	72.10	− 1016.7
Iran	1.65	10.36	3.37	− 1.30	0.01	− 0.02	380.47	10.81	75.93	− 1189.9
Jordan	1.31	10.19	3.24	− 1.42	0.01	− 0.02	374.57	10.72	68.06	− 1104.9
Lebanon	1.60	10.04	3.38	− 1.36	0.01	− 0.03	384.17	11.63	72.03	− 1127.2
Pakistan	1.42	10.21	3.28	− 1.30	0.01	− 0.03	389.00	9.61	74.14	− 1126.4
Syria	1.62	9.87	3.21	− 1.06	− 0.01	− 0.03	389.21	12.16	73.37	− 1109.2
Turkey	1.67	10.20	3.34	− 0.91	0.01	− 0.03	385.52	10.64	74.47	− 1171.3
USSR	1.67	10.19	3.64	− 0.80	0.01	− 0.02	369.63	9.91	74.34	− 1169.2

Table 4. Percentage of correctly classified (underlined) and mis-classified accessions from each country of origin from the jackknife classification

Origin	Classified as												
	Af	Ch	Eg	Et	Gr	In	Ir	Jo	Le	Pa	Sy	Tu	US
Afghanistan	66	0	0	0	3	0	3	0	0	13	0	0	16
Chile	4	57	0	0	6	0	8	2	4	2	6	4	6
Egypt	2	0	68	5	0	0	5	2	0	10	0	2	6
Ethiopia	1	0	0	86	0	10	1	1	0	0	1	0	0
Greece	0	12	0	0	53	0	6	0	6	0	0	12	12
India	0	0	1	29	0	68	0	1	1	0	0	0	0
Iran	25	6	4	0	3	0	43	1	2	8	1	1	6
Jordan	0	0	5	0	0	0	5	63	5	5	16	0	0
Lebanon	4	8	4	0	8	0	4	20	8	4	20	8	12
Pakistan	0	0	14	0	7	14	7	7	0	36	0	7	7
Syria	0	3	3	3	3	3	0	16	6	13	45	3	0
Turkey	5	13	0	0	17	0	12	0	5	3	10	26	9
USSR	8	17	0	0	8	0	17	0	0	0	0	25	25

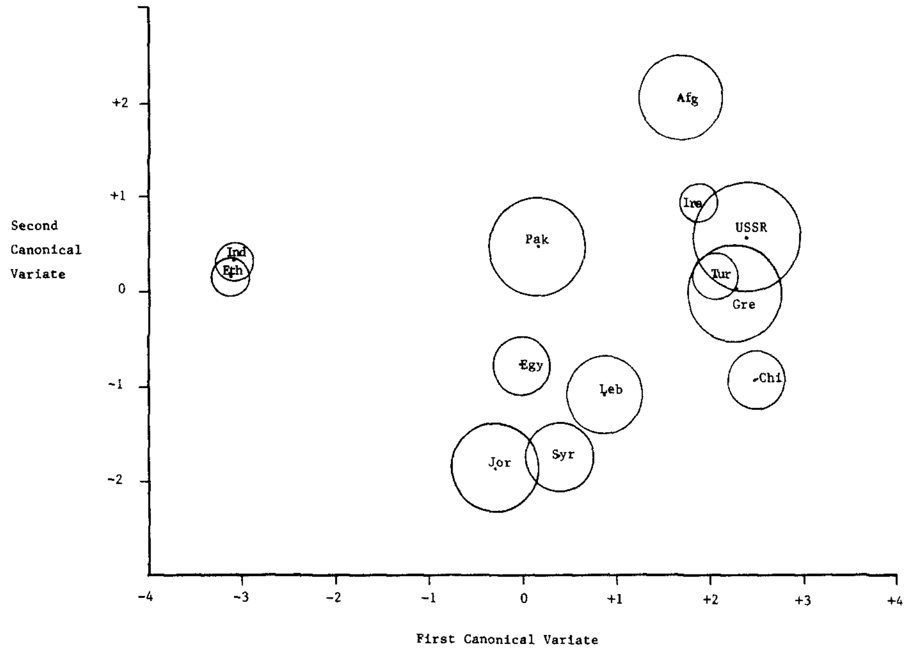

Fig. 1. Scatter diagram of the first two canonical variable mean values for the 13 countries of origin. Country names are abbreviated to their first three letters (except USSR). Circles represent confidence limits at P = 0.95.

ty. The second canonical variate (axis 2) represents the variation in mainly seed size and the number of seeds per pod. Thus on axis 2 accessions from Jordan, Syria and Chile have larger seed size and lower number of seeds per pod than those from Afghanistan at the other extreme, which have small seeds with a high number of seeds per pod.

Discussion

The evolution within a crop of ecotypes adapted to specific environments is a function of populations rather than individuals. Populations subjected to particular environmental conditions evolve adaptive gene complexes which are conserved by genetic linkage or natural or human selection. This study has focussed on these adaptive gene complexes. For their analysis a multivariate approach was required and univariate analysis of the variation can only give an incomplete picture.

An observer in a field of lentil germplasm can recognize accessions from different geographic areas by the presence of clusters of characters, both quantitative and qualitative. Barulina (1930) classified lentil germplasm into regional forms or *grex* on the basis of the qualitative characters, pubescence, degree of pod dehiscence, relative length of the calyx teeth, number of flowers per peduncle, branching pattern and flower colour. The grouping in the present study embraced quantitative characters which can be compared with the earlier work.

The grouping by country of origin is summarized visually in the canonical scatter diagram, in which the two axes account for 84.6% of the variability, (Fig. 1) with differences between countries reflecting genetic differentiation or similarity. There is a Levantine group comprising the accessions from Egypt, Jordan, Lebanon and Syria. Remembering that canonical variate 1 represents mainly biomass and phenology, within the levantine group accessions from Jordan and Egypt are on average earlier to flowering and maturity than those from Syria and Lebanon. In the second canonical axis representing predominantly seed size and seed number per pod, the group means from Syria and Jordan reflect a larger average seed size than amongst Egyptian and Lebanese accessions. Despite these small local differences there is a close similarity in quantitative traits between accessions from Egypt, Jordan, Lebanon and Syria. There were many misclassifications between accessions from these four countries by the Jack-knife technique. These countries, being neighbours, have relatively similar agro-ecological conditions. In Barulina's classification the accessions from these countries are mostly within Grex *asiaticae* with some accessions from G. *intermediae* in Jordan and Syria and a few *macrosperma* lines from Syria.

A second clear group is that comprising accessions from Greece, Iran, Turkey and USSR or the geographic region of South-East Europe and North-West Asia. This comprises the latest accessions to flower and maturity, which develop a very large biomass, but a low harvest index.

There are two outliers in Fig. 1 from the South-East Europe and North-West Asia group on the second canonical axis, namely Afghanistan and Chile. Afghan accessions are characterized by very small seeds and a high number of seeds per pod. In contrast, Chilean material is predominantly *macrosperma* with few seeds per pod. The nearest country in the canonical scatter diagram (Fig. 1) to the Chilean accessions is Greece, the sole representative of Europe (Turkish accessions were from Asian Turkey). This is commensurate with the European introduction of lentil into Chile in the *post* Colombine period.

The Pakistan group is distinctive but intermediate in time to flower and maturing between the highland, late maturity Afghan germplasm and the very early maturing Indian material.

The Indian and Ethiopian germplasm is all early to flower and mature, has a low biological yield, plant height and lowest pod height, and is consequently distinct from other material. Barulina classified the Indian germplasm as *pilosae* on the basis of its strong pubescence; but Ethiopian material was classified into G. *aethiopicae* because of the endemic qualitative character of an elongated pod apex. She did, however, note the similarity between these two groups in other characters.

The discussion has emphasized the regional differentiation of lentils through clusters of associated quantitative characters revealed by the multivariate analysis of quantitative characters. It confirms to a large extent the regional differentiation made by Barulina on the basis of groups of qualitative characters.

The close resemblance between germplasm from adjacent countries, climatically similar such as within the levantine group, indicates that the ecological environment has been the major evolutionary force in the cultivated lentil. Furthermore, with the character time to maturity as the most important trait in the classification, it appears that phenological adaptation to the ecological environment has been central to the evolution of the crop.

In chickpea a similar association between geographic origin and phenotype was found by cluster analysis of many primitive varieties (Naryan & Macefield, 1976). However in durum wheat, a more heavily exploited crop in which material has been widely exchanged between breeders, the differences between germplasm from countries with widely divergent agro-ecological conditions have been reduced (Spagnoletti Zeuli & Qualset, 1987).

The regional grouping in lentil has a practical use in the utilization of germplasm by breeders. Collections often contain so many entries and such a multiplicity of potentially interesting characters that they are formidable to potential users. The grouping found in this study could be used in the selection of parental stocks from the germplasm collection and in establishing a core collection (Frankel & Brown, 1984). The aim of a core collection is to include, with minimum redundancy, the genetic diversity of a species in a condensed (and hence manageable) yet representative assembly of accessions.

Acknowledgements

The authors wish to gratefully acknowledge the help of Dr. M. Solh, the technical assistance of Miss. L. Khoury, Mr. O. Obaji and Mr. A. Ismail and the comments on the manuscript of Drs. A.B. Damania and R.S. Malhotra.

References

Afifi, A.A. & V. Clark, 1984. Computer-aided multivariate analysis. Lifetime Learning Publications, Belmont, California, U.S.A.

Barulina, H. 1930. Lentils of the USSR and other countries. Supplement 40 to the Bulletin of Applied Botany, Genetics and Plant Breeding, Leningrad, 265–304. (English Summary).

Damania, A.B. & E. Porceddu, 1983. Variation in landraces of *Turgidum* and bread wheats and sampling strategies for collecting wheat genetic resources. In: Proceedings of 6th International Wheat Genetics Symposium, Kyoto, Japan. Pages 123–136.

Erskine, W. & M.A. Choudhary, 1986. Variation between and within lentil landraces from Yemen Arab Republic. Euphytica 35: 695–700.

Erskine, W. & J.R. Witcombe, 1984. Lentil Germplasm Catalog. International Center for Agricultural Research in the Dry Areas (ICARDA), P.O. Box 5466, Aleppo, Syria.

Frankel, O.H. & A.H.D. Brown, 1984. Plant genetic resources today: a critical appraisal. In: J.H.W. Holden & J.T. Williams (Eds). Crop Genetic Resources: conservation and evaluation. George Allen and Unwin. 249–257.

Jennich, R. & P. Sampson, 1985. Stepwise Discriminant Analysis. In: W.J. Dixon (Ed.). BMDP Statistical Software Manual. University of California Press, London, U.K. 519–537.

Narayan, R.K.J. & A.J. Macefield, 1976. Adaptive responses and genetic divergence in a world germplasm collection of chickpea (*Cicer arietinum* L.). Theoretical and Applied Genetics 47: 179–187.

Phillips, M.S. & W. Powell, 1984. The effect of natural selection selection on composite cross populations of oats (*Avena sativa*) grown at contrasting sites in Scotland. Journal of Agricultural Science, Cambridge 102: 469–473.

Riggs, T.J., 1973. Studies in barley breeding. Ph.D. thesis, University of Edinburgh, U.K.

Seal, H.L., 1964. Multivariate Statistical Analysis for Biologists. Methuen, London, U.K.

Skibinski, D.O.F., D. Rasool & W. Erskine, 1984. Aspartate animo-transferase allozyme variation in a germplasm collection of the domesticated lentil (*Lens culinaris*). Theoretical and Applied Genetics 68: 441–448.

Spagnoletti Zeuli, P.L. & C.O. Qualset, 1987. Geographical diversity for quantitative spike characters in a world collection of durum wheat. Crop Science 27: 235–241.