DE GRUYTER OPEN
VOL. 26 NO 4(66): 30-34
DOI 10.1515 /OSZN-2015-0026

Agnieszka Pyza *, Helena Kubicka-Matusiewicz *, Dariusz Gozdowski **

Assessment of similarity of inbred lines and F_1 generations in collection of winter rye (*Secale cereale* L.)

Ocena podobieństwa linii wsobnych i pokolenia F_1 w kolekcji żyta ozimego (*Secale cereale* L.)

*Mgr inż. Agnieszka Pyza, prof. dr hab. Helena Kubicka-Matusiewicz, Polish Academy of Sciences Botanical Garden – Center for Biological Diversity Conservation in Powsin, Prawdziwka 2 St., 02-973 Warsaw, e-mail: pyza-aga@wp.pl

**Dr Dariusz Gozdowski, Warsaw University of Life Sciences, Faculty of Agriculture and Biology, Department of Experimental Design and Bioinformatics, Nowoursynowska 159 St., 02-776 Warsaw.

Keywords: cluster analysis, principal component analysis (PCA), variability, inbred line, rye
Słowa kluczowe: analiza skupień, analiza składowych głównych, zmienność, linie wsobne, żyto

Abstract
Effective tools for evaluation of diversity in the collected gene resources of a given species are multivariate statistical methods. They provide information on phenotypic and genetic variability of collected material. The subjects of this study were nine inbred lines and three F_1 generations of winter rye (*Secale cereale* L.), growing in experimental plots of the Polish Academy of Sciences Botanical Garden, Centre for Biological Diversity Conservation in Powsin. The evaluation was performed over the period of 3 years. Observations were made of the following traits: length of stem, length of spike, number of nodes in spike rachis, number of kernels per spike, weight of kernels per spike, weight of 1000 kernels, length of flag leaf, length of subflag leaf, length of third leaf, width of flag leaf, width of subflag leaf, width of third leaf, area of leaves per stem, number of stems per plant and area of leaves per plant.
On the basis of cluster analysis and principal components analysis, two genetically homogeneous groups were identified. Mean values and standard deviations were calculated for each trait in each group and for all genotypes together. Multivariate distance matrix permitted identification of the most genetically similar and most distant forms.

Streszczenie
Ważnym elementem oceny różnorodności w obrębie gromadzonych kolekcji zasobów genowych danego gatunku jest zastosowanie wielocechowych metod statystycznych. Umożliwiają one poszerzenie wiedzy o zmienności fenotypowej oraz genetycznej posiadanych materiałów kolekcyjnych. Przedmiotem badań było dziewięć linii wsobnych i trzy pokolenia F_1 żyta ozimego (*Secale cereale* L.), rosnące na polu doświadczalnych Polskiej Akademii Nauk Ogrodzie Botanicznym – Centrum Zachowania Różnorodności Biologicznej w Powsinie, które oceniano w cyklu trzyletnim. Obserwacji poddano następujące cechy: długość źdźbła (LSM), długość kłosa (LSK), liczba węzłów na osadce kłosowej (NNSR), liczba ziarniaków w kłosie (NKS), masa ziarniaków w kłosie (WKS), masa 1000 ziarniaków (WTK), długość liścia flagowego (LFL), długość liścia podflagowego (LSL), długość trzeciego liścia (LTL), szerokość liścia flagowego (WFL), szerokość liścia podflagowego (WSL), szerokość trzeciego liścia (WTL), powierzchnia liści na źdźble (ALS), krzewistość (NSP) i powierzchnia liści na roślinie (ALP).
Na podstawie analizy skupień i składowych głównych PCA wydzielono dwie jednorodne genetycznie grupy. Obliczono wartości średnie i odchylenia standardowe dla poszczególnych cech z podziałem na grupy i dla wszystkich genotypów razem. Wielocechowa macierz odległości między poszczególnymi genotypami pozwoliła na określenie form najbardziej podobnych do siebie i najbardziej różniących się od siebie.

1. INTRODUCTION

High degree inbred lines are used as components for obtaining hybrids. Thus, the agricultural worth of hybrids is determined to a high degree by the traits of the lines [Grudkowska *et al*. 2001]. Inbred lines should be characterised by good equalisation high vigour and high degree of homozygosity, abundant production of pollen and high combination ability [Kadłubiec *et al*. 2000], to be suitable for breeding programme that is time-consuming and long-lasting [Lewandowska *et al*. 2012]. Multivariate statistical methods provide information on phenotypic and genetic variability of the collection of cultivated plants [Ukalski *et al*. 2009]. Such methods permit classification of the objects studied into groups having one or many traits the same and allows characterisation of the groups [Kubicka *et al*. 2006, 2012, 2013]. The method most often used for investigation of multivariate variability of cultivated plants is the principal component analysis (PCA) [Rao 1964, Dudnikov 2000, Studnicki *et al*. 2010]. So far, the multivariate statistical methods have been successfully used for determination of phenotypic variability and classification of collections of many species of cultivated plants, including rye, oat, wheat and maize [Bujak *et al*. 2006, Studnicki *et al*. 2009, Ukalski *et al*. 2009, Ukalski, Śmiałowski 2011, Lewandowska *et al*. 2012].
The aim of this study is evaluation of morphological diversity of inbred lines and hybrid forms of winter rye (*Secale cereale* L.) from our collection with the help of multidimensional statistical methods.

2. MATERIALS AND METHODS

The study was performed on nine inbred lines and three F_1 generations of winter rye (*Secale cereale* L.). The inbred lines differed in the origin and degree of inbred (Table 1). The subjects of observation were the following traits: length of stem (LSM), length of spike, number of nodes in spike rachis, number of kernels per spike, weight of kernels per spike, weight of 1000 kernels, length of flag leaf (LFL), length of subflag leaf, length of third leaf, width of flag leaf (WFL), width of subflag leaf (WSL), width of third leaf (WTL), area of leaves per stem, number of stems per plant (NSP) and area of leaves per plant. The measurements were performed on 3 stems from 10 plants representing each analysed forms over the period of 3 years of studies at the Polish Academy of Sciences Botanical Garden, Centre for Biological Diversity Conservation in Powsin.

The experimental data were analysed using univariate and multivariate statistical methods. Cluster analysis was used for multivariate classification of the examined genotypes. The variables were standardised because of various units that were measured. Squared Euclidean distance was used as a measure of distance between genotypes. Ward's method was used for agglomeration of the clusters.

For evaluation of differences between the distinguishes groups means (centroids) for the clusters were calculated and analysis of variance was performed to estimate significance of the differences. Pearson's linear correlations were calculated for pairs of all examined traits. PCA was performed to evaluate multivariate relationships between variables and for evaluation of multivariate differences between genotypes.

The analyses were conducted using Statistica, version 10 (StatSoft 2010). Significance level was set for all analyses at $p = 0.05$.

Table 1. Origin of genotypes and their generation.

Number and name of genotype	Origin and generation
1-mk, 2-MK	Shortness form from Jeleniec (J) - S_{24}
3-kn, 4-KN	Segregant from crossing inbred lines: L145 x L31 (L145 - Smolickie variety (S), L31 - Dankowskie Selekcyjne variety (DS) - S_{24}
5-L79	Smolickie variety (S) - S_{26}
6-jeż	Dankowskie Selekcyjne variety (DS) - S_{26}
7-L299, 8-J74	Shortness form from Jeleniec (J) - S_{26}
9-M147	Breeding materials - S_{26}
10-M147xJ74, 11-jeżxL299, 12-jeżxL79	F_1 generation

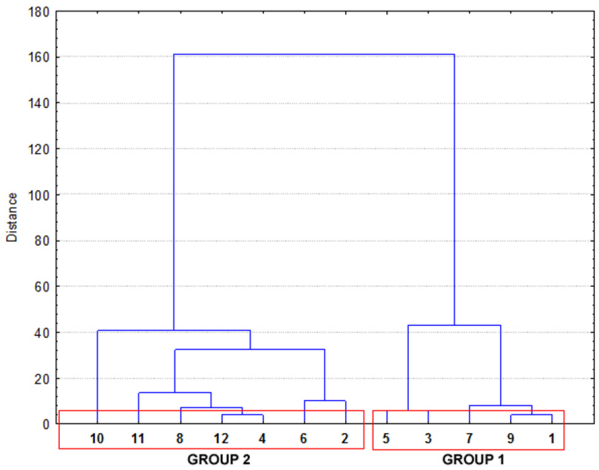

Fig. 1. Dendrogram presenting similarity of 12 genotypes on the basis of cluster analysis.

3. RESULTS AND DISCUSSION

Cluster analysis revealed statistically significant differences in the traits studies and according to its results, two homogeneous groups were identified, (Fig. 1). The plants from group I were characterised by a shorter stem and a shorter spike, smaller number of kernels in the spike, lower mass of kernels in a spike and lower mass of one thousand kernels (Table 2). Moreover, the plants from group I have leaf area by over 5 times smaller and number of stems per plant by almost twice smaller than those from group II. In general, statistically significant differences were found between the groups in the majority of traits considered, at $p < 0.05$.

Results of correlation analysis (Table 3) and PCA (Fig. 2a) show that all the traits were positively correlated to a higher or lesser degree. It means that usually, a higher value of one trait was correlated with the higher values of the other traits. Slightly poorer was the correlation of the width of flag leaf, width of subflag leaf and width of the third leaf with the other traits, although these widths were strongly correlated with each other. Slightly poorer were also the correlations of the length of stem and number stems per plant with the other traits.

The first principal component (PC1) explains over 2/3 of the total variation of the data, which means that the diversity of genotypes in the PCA plot (Fig. 2b) is the strongest along the horizontal axis. The group I genotypes, including 1 - mk; 9 - M147; 7 - L299; 3 - kn and 5 - L79, show rather low values of almost all traits. Group II genotypes (2 - MK; 6 - jeż; 4 - KN; 12 - F_1 (jeżxL79); 8 - J74; 11 - F_1 (jeżxL299) and 10 - F_1 (M147xJ74) show the majority of traits of higher values than those of group I. No statistically significant differences between the groups were found only for four traits length of flag leaf, width flag leaf, width subflag leaf and width third leaf. Within the groups, some of the genotypes were considerably different from the other ones; in group I, the most distant was genotype 3 (kn), while in group II, genotype 10 (F_1 (M147xJ74). These observations were confirmed by the multivariate Euclidean distances between genotypes presented in Table 4.

Table 2. Means, standard deviations (SD) and results of analysis of variance (F and P-value) for comparison of clusters distinguished in cluster analysis (significant correlations are in bold, $p < 0.05$).

Traits	Clusters				F	P-value	Total	
	1		2					
	mean	SD	mean	SD			mean	SD
Length of stem [cm]	69.65	11.59	115.10	32.21	8.909	**0.014**	96.16	34.10
Length of spike [mm]	81.98	11.25	114.19	12.58	20.783	**0.001**	100.77	20.18
Number of nodes in spike rachis	14.16	1.35	19.37	1.30	45.335	**0.000**	17.20	2.97
Number of kernels per spike	12.73	9.36	48.51	21.16	12.292	**0.006**	33.60	24.81
Weight of kernels per spike [g]	0.20	0.16	1.22	0.76	8.500	**0.015**	0.79	0.77
Weight of 1000 kernels [g]	13.71	4.55	24.38	7.92	7.210	**0.023**	19.93	8.48
Length of flag leaf [mm]	104.78	39.88	143.39	25.11	4.284	0.065	127.30	36.30
Length of subflag leaf [mm]	163.66	42.45	210.97	20.83	6.654	**0.027**	191.26	38.54
Length of third leaf [mm]	159.24	28.86	203.74	17.81	11.035	**0.008**	185.20	31.64
Width of flag leaf [mm]	10.43	2.83	12.48	2.39	1.852	0.203	11.63	2.68
Width of subflag leaf [mm]	13.15	3.12	15.79	2.83	2.346	0.157	14.69	3.12
Width of third leaf [mm]	12.57	2.21	15.28	2.22	4.365	0.063	14.15	2.53
Area of leaves per stem [dm^2]	0.38	0.09	0.87	0.15	39.000	**0.000**	0.66	0.28
Number stems per plant	5.03	1.54	9.71	2.24	16.085	**0.002**	7.76	3.07
Area of leaves per plant [dm^2]	1.87	0.53	9.50	3.96	17.820	**0.002**	6.32	4.91

Table 3. Correlations coefficients between examined traits (significant correlations are in bold, $p < 0.05$).

Traits	LSM	LSK	NNSR	NKS	WKS	WTK	LFL	LSL	LTL	WFL	WSL	WTL	ALS	NSP	ALP
LSM		**0.59**	**0.66**	**0.82**	**0.84**	**0.85**	0.52	0.55	**0.72**	0.06	0.06	0.13	**0.65**	**0.90**	**0.75**
LSK	**0.59**		**0.96**	**0.72**	**0.69**	**0.64**	**0.80**	**0.82**	**0.77**	**0.61**	**0.66**	**0.70**	**0.93**	**0.65**	**0.64**
NNSR	**0.66**	**0.96**		**0.80**	**0.74**	**0.69**	**0.75**	**0.80**	**0.81**	0.57	**0.62**	**0.70**	**0.93**	**0.77**	**0.76**
NKS	**0.82**	**0.72**	**0.80**		**0.96**	**0.90**	**0.60**	**0.66**	**0.69**	0.41	0.46	0.51	**0.72**	**0.83**	**0.69**
WKS	**0.84**	**0.69**	**0.74**	**0.96**		**0.94**	**0.58**	**0.60**	**0.63**	0.43	0.46	0.52	**0.73**	**0.79**	**0.68**
WTK	**0.85**	**0.64**	**0.69**	**0.90**	**0.94**		**0.67**	**0.69**	**0.69**	0.50	0.49	0.52	**0.68**	**0.83**	**0.73**
LFL	0.52	**0.80**	**0.75**	**0.60**	**0.58**	**0.67**		**0.98**	**0.85**	**0.68**	**0.69**	**0.62**	**0.73**	**0.58**	0.56
LSL	0.55	**0.82**	**0.80**	**0.66**	**0.60**	**0.69**	**0.98**		**0.91**	**0.66**	**0.68**	**0.64**	**0.75**	**0.64**	**0.58**
LTL	**0.72**	**0.77**	**0.81**	**0.69**	**0.63**	**0.69**	**0.85**	**0.91**		0.40	0.42	0.43	**0.75**	**0.76**	**0.66**
WFL	0.06	**0.61**	0.57	0.41	0.43	0.50	**0.68**	**0.66**	0.40		**0.98**	**0.96**	0.57	0.27	0.39
WSL	0.06	**0.66**	**0.62**	0.46	0.46	0.49	**0.69**	**0.68**	0.42	**0.98**		**0.98**	**0.63**	0.25	0.37
WTL	0.13	**0.70**	**0.70**	0.51	0.52	0.52	**0.62**	**0.64**	0.43	**0.96**	**0.98**		**0.69**	0.33	0.42
ALS	**0.65**	**0.93**	**0.93**	**0.72**	**0.73**	**0.68**	**0.73**	**0.75**	**0.75**	0.57	**0.63**	**0.69**		**0.67**	**0.76**
NSP	**0.90**	**0.65**	**0.77**	**0.83**	**0.79**	**0.83**	**0.58**	**0.64**	**0.76**	0.27	0.25	0.33	**0.67**		**0.88**
ALP	**0.75**	**0.64**	**0.76**	**0.69**	**0.68**	**0.73**	0.56	**0.58**	**0.66**	0.39	0.37	0.42	**0.76**	**0.88**	

ALP: area of leaves per plant ; ALS: area of leaves per stem; LFL: length of flag leaf; LSK: length of spike; LSL: length of subflag leaf; LSM: lengh of stem; LTL: length of third leaf; NKS: number of kernels per spike; NNSR: number of nodes in spike rachis; NSP: number stems per plant; WFL: width of flag leaf; WKS: weight of kernels per spike; WSL: width of subflag leaf; WTK: weight of 1000 kernels; WTL: width of third leaf.

Table 4. Multivariate standardised squared Euclidean distances between genotypes (significant correlations are in bold, $p < 0.05$).

Genotype	1	2	3	4	5	6	7	8	9	10	11	12
1	0	23.2	29	30.0	13	17.0	**9.6**	26.0	**4.0**	83	38.3	27.3
2	23.2	0	84	17.7	55	10.1	22.3	17.4	22.5	28	30.2	15.9
3	28.7	83.7	0	58.1	**6**	63.5	37.0	58.7	24.4	145	60.8	51.9
4	30.0	17.7	58	0	36	17.2	14.5	**6.2**	16.6	21	**8.6**	**3.9**
5	13.1	54.6	**6**	36.2	0	40.2	16.6	36.9	**8.5**	108	39.7	32.7
6	17.0	10.1	64	17.2	40	0	15.2	12.4	15.3	40	16.3	15.2
7	**9.6**	22.3	37	14.5	17	15.2	0	15.5	**4.4**	58	17.4	16.4
8	26.0	17.4	59	**6.2**	37	12.4	15.5	0	17.1	28	13.2	**6.5**
9	**4.0**	22.5	24	16.6	**9**	15.3	**4.4**	17.1	0	66	22.9	15.0
10	82.9	28.4	145	21.3	108	39.9	58.0	28.3	65.9	0	31.0	27.6
11	38.3	30.2	61	**8.6**	40	16.3	17.4	13.2	22.9	31	0	10.9
12	27.3	15.9	52	**3.9**	33	15.2	16.4	**6.5**	15.0	28	10.9	0

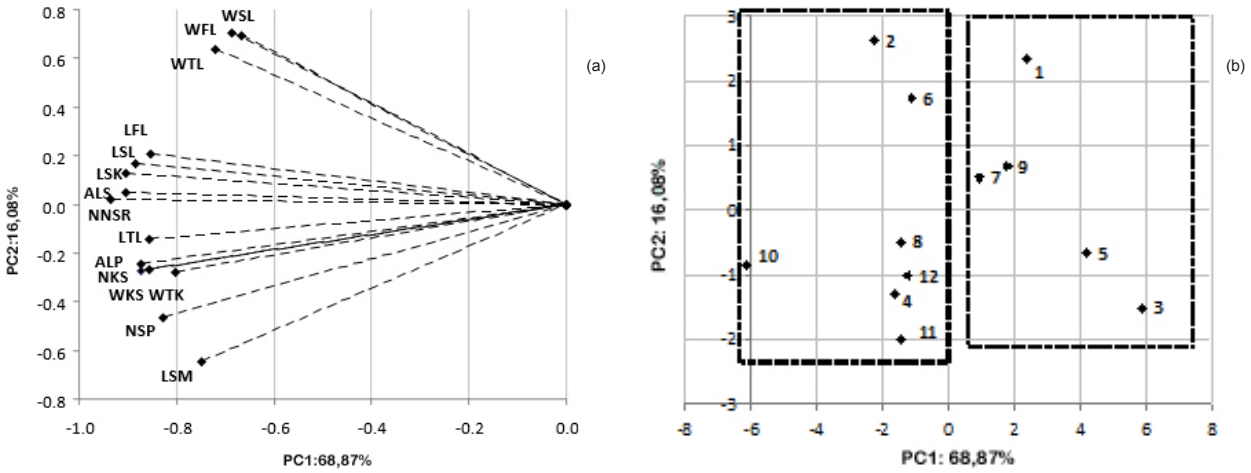

Fig. 2. Results of PCA presenting relationships between traits (a) and multivariate differences between genotypes (b).
ALP: area of leaves per plant ; ALS: area of leaves per stem; LFL: length of flag leaf; LSK: length of spike; LSL: length of subflag leaf; LTL: length of third leaf; NKS: number of kernels per spike; NNSR: number of nodes in spike rachis; WFL: width of flag leaf; WKS: weight of kernels per spike; WSL: width of subflag leaf; WTK: weight of 1000 kernels; WTL: width of third leaf; PCA: principal component analysis.

All traits studied were mutually positively correlated, some of them very strongly and some others poorly. The two groups of uniform genotypes showed statistically significant differences in the majority of traits studied, except for the width of leaf and length of flag leaf.

Statistical methods applied permitted separation of two groups of genotypes and their multivariate characterisation. Similar statistical methods have been applied by Ukalski *et al*. [2011] for evaluation of rye genotypes diversity and by, for example, Kuleung *et al*. [2006], Shang *et al*. [2006] and Bolibok-Brągoszewska *et al*. [2009] for determination of molecular markers.

Analysis of similarity based on a large number of traits permits separation of genotypes similar in many traits and those much genetically different, which is important for the choice of genotypes for gene banks [Studnicki *et al*. 2010]. The choice of genotypes showing significant differences in many traits for core collections ensures high genotypic and phenotypic diversity, which makes the collection a valuable material for breeding new hybrid varieties.

4. CONCLUSIONS

Multivariate cluster analysis and PCA are the methods highly useful for evaluation of genetic diversity of collection forms. The above statistical methods permit effective management of collection and facilitate the choice of parental forms that can be used for plant breeding.

REFERENCES

BOLIBOK-BRĄGOSZEWSKA H., HELLER-USZYŃSKA K., WENZL P., USZYŃSKI G., KILIAN A., RAKOCZY-TROJANOWSKA M. 2009. DArT markers for the rye genome-genetic diversity and mapping. BMC Genomics 10(1): 578.

BUJAK H., JEDYŃSKI S., KARCZMAREK J., KARWOWSKA C., KURCZYCH Z., ADAMCZYK J. 2006. Wielocechowa analiza wartości hodowlanej linii wsobnych kukurydzy (*Zea mays* L.). Biul. IHAR 240/241: 211-216.

DUDNIKOV A. J. 2000. Multivariate analysis of genetic variation in *Aegilops tauschii* from the world germplasm collection. Gen. Res. Crop Evol. 47: 185-190.

GRUDKOWSKA M., MADEJ L. 2001. Struktura plonu wybranych linii wsobnych żyta ozimego. Biul. IHAR 218/219: 389-393.

KADŁUBIEC W., KARWOWSKA C., KURCZYCH Z., KURIATA R. 2000. Zdolność kombinacyjna linii wsobnych kukurydzy. Biul. IHAR 216: 371-378.

KUBICKA H., PUCHALSKI J., NIEDZIELSKI M., ŁUCZAK W., MARTYNISZYN A. 2006. Gromadzenie i ocena zasobów genowych żyta. Biul. IHAR 240/241: 141-150.

KUBICKA H., GOZDOWSKI D., PUCHALSKI J., ŁUCZAK W., MARTYNISZYN A. 2012. Wielocechowa ocena zróżnicowania form lokalnych żyta o różnym pochodzeniu geograficznym pod względem cech morfologicznych i użytkowych. Biul. IHAR 264: 105-115.

KUBICKA H., GOZDOWSKI D., PUCHALSKI J., WIŚNIEWSKI M. 2013. Wielocechowa analiza zmienności wybranych cech użytkowych form lokalnych żyta ozimego. J. of Agronomy 15: 54-59.

KULEUNG C., BAENZIGER P. S., KACHMAN, S. D., DWEIKAT I. 2006. Evaluating the genetic diversity of triticale with wheat and rye SSR markers. Crop Science 46(4): 1692-1700.

LEWANDOWSKA S., KURIATA R., KADŁUBIEC W. 2012. Morfologiczne i genetyczne zróżnicowanie linii wsobnych kukurydzy. Biul. IHAR 264: 177-182.

RAO C. R. 1964. The use and interpretation of principal components in applied research. Sankhya A 26: 329-358.

SHANG H. Y., WEI Y. M., WANG X. R., ZHENG Y. L. 2006. Genetic diversity and phylogenetic relationships in the rye genus *Secale* L.(rye) based on *Secale cereale* L. microsatellite markers. Genetics and Molecular Biology 29(4): 685-691.

StaSoft, INC. 2010. STATISTICA data analysis software system, version 10. www.statsoft.com.

STUDNICKI M., MĄDRY W., ŚMIAŁOWSKI T. 2009. Porównanie efektywności metod statystycznych tworzenia kolekcji podstawowej na przykładzie pszenicy jarej. Biul. IHAR 252: 105-118.

STUDNICKI M., MADRY W., KOCIUBA W. 2010. The efficiency and effectiveness of sampling strategies used to develop a core collection for the Polish spring triticale (*Triticosecale* Wittm.) germplasm resources. Communications in Biometry and Crop Science 5(2): 127-137.

UKALSKI K., ŚMIAŁOWSKI T. UKALSKA J. 2009. Analiza plonowania i stabilności genotypów owsa za pomocą metody graficznej typu GGE. Żywność. Nauka. Techn. Jakość. 3: 127-140.

UKALSKI K., ŚMIAŁOWSKI T. 2011. Wielocechowa analiza wyników doświadczeń wstępnych z żytem ozimym. Biul. IHAR 260/261: 251-262.