

JOANNA UKALSKA ¹
KRZYSZTOF UKALSKI ¹
TADEUSZ ŚMIAŁOWSKI ³
WIESŁAW MĄDRY ²

¹ Katedra Biometrii, Szkoła Główna Gospodarstwa Wiejskiego, Warszawa

² Katedra Doświadczalnictwa i Bioinformatyki, Szkoła Główna Gospodarstwa Wiejskiego, Warszawa

³ Instytut Hodowli i Aklimatyzacji Roślin, Zakład Roślin Zbożowych w Krakowie

Badanie zmienności i współzależności cech użytkowych w kolekcji roboczej pszenicy ozimej (*Triticum aestivum* L.) za pomocą metod wielowymiarowych

Część II. Analiza składowych głównych na podstawie macierzy korelacji fenotypowych i genotypowych

An examination of diversity and interrelationships among traits in a winter wheat (*Triticum aestivum* L.) germplasm collection by multivariate methods
Part II. Principal component analysis using the phenotypic and genotypic correlation matrix

W pracy przedstawiono dwa podejścia do badania zależności między cechami w kolekcji roboczej pszenicy ozimej wykorzystujące zastosowania metody głównych składowych (PCA). Pierwsze, klasyczne podejście, polegało na przeprowadzeniu PCA na podstawie macierzy współczynników korelacji fenotypowych tj. macierzy uzyskanej w oparciu o estymatory średnich fenotypowych z rozważanych lat, a następnie wyznaczeniu współczynników korelacji między składowymi, a wartościami estymatorów średnich dla rozważanych cech. Drugie podejście, przedstawione w pracy, polegało na wykonaniu analizy składowych głównych na macierzy współczynników korelacji genotypowych analizowanych cech, czyli macierzy korelacji pomiędzy nieobserwowalnymi efektami genotypowymi badanych cech. Podejście takie umożliwia zbadanie genetycznych uwarunkowań, niepodatnych na wpływy środowiska, pomiędzy badanymi cechami. Uzyskano dużą zbieżność wyników obu metod. Jednak zastosowanie PCA na podstawie macierzy korelacji genotypowych pozwoliło na wyjaśnienie ok. 15% zmienności wielocechowej, więcej niż przy zastosowaniu klasycznej metody PCA. Ponadto stwierdzono większe bezwzględne wartości współczynników korelacji badanych cech ze składowymi głównymi.

Słowa kluczowe: analiza składowych głównych, korelacje fenotypowe, korelacje genotypowe, pszenica ozima, zasoby genowe

In the paper, the diversity among quantitative traits in the winter wheat germplasm working collection using Principal Component Analysis (PCA) was studied. Fifty-one genotypes (cultivars and clones) from the Plant Breeding and Acclimatization Institute, Department of Cereal Crops in Cracow, were evaluated in the years 1999–2002. Two approaches of the PCA were applied. The first, classical approach, involved the phenotypic correlation matrix, i.e. correlation coefficients between across year phenotypic means. In the second, new approach, application of genotypic correlation matrix, i.e. correlation coefficients matrix between the unobservable genotypic effects for considered traits in PCA, has been proposed. Coincident results were obtained using the above methods. However, the first three principal components obtained using the genotypic correlation matrix compared to the classical PCA accounted for 15% more of the overall variation among genotypes. Moreover, higher absolute values of correlation coefficients between principal components and the evaluated traits were recorded.

Key words: genotypic correlation matrix, germplasm collection, phenotypic correlation matrix, principal component analysis, winter wheat

WSTĘP

Charakterystyka różnorodności oraz analiza podobieństwa obiektów w kolekcjach roboczych ma na celu ocenę ich przydatności do potrzeb hodowlanych. Niezbędna jest również przy prowadzeniu i organizacji kolekcji, a zwłaszcza jej redukcji do kolekcji podstawowej (Skinner i in., 1999; Mohammadi i Prasanna, 2003; Mądry, 2007). Coroczna ocena odmian i rodów hodowlanych w kolekcji najczęściej wykonywana jest na podstawie obserwacji fenotypowych cech jakościowych i ilościowych i nazywana jest zróżnicowaniem fenotypowym (Skinner i in., 1999; Mohammadi i Prasanna, 2003; Upadhyaya i in., 2005; Mądry, 2007).

Przy analizie różnorodności fenotypowej obiektów kolekcyjnych wykorzystuje się informacje dotyczące wielu cech użytkowych. Niezbędne jest zastosowanie wielowymiarowych metod statystycznych. Są one podstawowymi metodami dla dużych układów danych. Pozwalają na efektywne zbadanie związków pomiędzy badanymi zmiennymi, jak również na lepsze zrozumienie struktury obserwowanej kolekcji. Najczęściej stosowaną metodą do analizy różnorodności obiektów pod względem cech ilościowych jest analiza składowych głównych (PCA). Jest stosowana do badania kolekcji zasobów genowych dla wielu gatunków roślin (Iezzoni i Pritts, 1991; Perez-Gonzalez, 1992; Flores i in., 1997; Ayana i Bekele, 1999; Mars i Marrakchi, 1999; Rojas i in., 2000; Abdi i in., 2002; Leguizmon i Badenes, 2003; Rotondi i in., 2003; Martinez-Calvo, 2007; Ukalska i in., 2007). PCA to jedna z najstarszych metod wielowymiarowych. Zapoczątkowana została przez Karla Pearsona (1901), później rozwijana przez Hottelinga (1933; 1936), Rao (1964), Erikssona i in. (1999). Metoda ta pozwala na wyznaczenie udziału rozważanych cech w wielo cechowej zmienności obiektów kolekcyjnych. Cechy, które tworzą pierwszą, drugą i kolejne składowe główne wykazują najsilniejszą moc dyskryminacyjną, tj. różnicującą badane obiekty. Jeżeli pomiary różnych zmiennych są podobnego rzędu wielkości a ich odchylenia standardowe nie różnią się znacząco, proponuje się wykonanie analizy PCA na podstawie macierzy kowariancji. W praktyce jednak obserwowane cechy różnią się znacznie rzędem wielkości oraz wartościami

odchyłeń standardowych. Zatem przy wielowymiarowej analizie kolekcji zasobów genowych PCA najczęściej wykonywana jest na podstawie macierzy współczynników korelacji (Jolliffe, 1986; Krzanowski, 1988; Everitt i Donn, 1992). Macierz ta odpowiada macierzy współczynników korelacji fenotypowych.

Niniejsza praca jest kontynuacją opracowania zawartego w części I (Ukalski i in., 2008). Jej celem jest zastosowanie analizy składowych głównych na podstawie macierzy współczynników korelacji fenotypowych oraz na podstawie macierzy współczynników korelacji genotypowych 14 cech użytkowych pszenicy ozimej *Triticum aestivum* L., a także porównanie obu zastosowanych podejść.

MATERIAŁ I METODY

Material

Materiał badawczy stanowiła kolekcja robocza pszenicy ozimej, zgromadzona przez Zakład Roślin Zbożowych Instytutu Hodowli i Aklimatyzacji Roślin w Krakowie, opisana w I części pracy dotyczącej korelacji fenotypowych i genotypowych (Ukalski i in., 2008). Analizy zastosowane w niniejszej pracy wykonano w oparciu o te same obiekty kolekcyjne (51 odmian i rodów hodowlanych odrzuconych na etapie doświadczeń przedrejestracyjnych), a także wykorzystano 14 tych samych cech, które analizowano w części I pracy. Obserwacje stanowiły dwukierunkową klasyfikację niekompletną obiekty \times lata.

Metody statystyczne

Do statystycznej analizy danych, odrębnie dla każdej z analizowanych cech, przyjęto losowy model postaci $y_{ij} = m + g_i + r_j + e_{ij}$, którego opis zawarto w części I (Ukalski i in., 2008).

Analizę składowych głównych (PCA) wykonano na dwa sposoby. Pierwsze, klasyczne podejście, polegało na przeprowadzeniu PCA na podstawie macierzy korelacji fenotypowych tj. macierzy uzyskanej w oparciu o estymatory średnich fenotypowych z rozważanych lat. Ze względu na niekompletny układ doświadczalny losowe efekty główne obiektów oszacowano za pomocą najlepszego nieobciążonego predyktora (BLUP). Następnie wyznaczono współczynniki korelacji między składowymi głównymi, a wartościami BLUP dla poszczególnych cech. Składowa główna jest kombinacją liniową wszystkich analizowanych zmiennych (cech). Po wyznaczeniu współczynników korelacji pomiędzy cechami a wybraną składową, można stwierdzić, na podstawie ich bezwzględnych wartości, które cechy charakteryzują się największą mocą dyskryminacyjną, czyli różnicującą badane obiekty kolekcyjne.

Drugie podejście przedstawione w pracy polega na wykonaniu analizy składowych głównych na macierzy korelacji genotypowych analizowanych cech, czyli macierzy korelacji pomiędzy nieobserwowalnymi efektami genotypowymi. Macierz korelacji genotypowych wyznaczono w sposób opisany w I części pracy (Ukalski i in., 2008). Proponowane podejście umożliwia zbadanie genetycznych uwarunkowań, niepodatnych na wpływy środowiska, pomiędzy badanymi cechami. Podobnie jak w podejściu klasycznym, wyznaczono współczynniki korelacji między składowymi głównymi a wartościami efektów genotypowych dla poszczególnych cech.

Jako uzupełnienie opisanych analiz przedstawiono wykresy badanych cech w układzie trzech pierwszych składowych głównych, a także rozmieszczenie obiektów kolekcyjnych w układzie dwóch pierwszych składowych głównych.

Obliczenia potrzebne do wykonania przedstawionych analiz statystycznych oraz prezentowane w pracy wykresy zostały wykonane przy użyciu pakietu SAS, za pomocą procedur: MIXED i PRINCOMP (Littell i in., 1996; Khattree i Naik, 2000; SAS/STAT User's Guide, 2002).

WYNIKI I DYSKUSJA

Wyniki analizy składowych głównych przedstawiono tabelarycznie oraz w postaci wykresów. W tabeli 1 podano wartości własne odpowiadające pierwszym trzem składowym głównym oraz procent wyjaśnianej przez nie całkowitej wariancji.

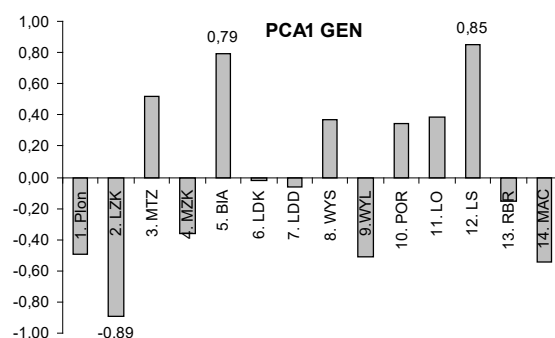
Tabela 1

Wartości własne oraz procent zmienności wyjaśniony przez 3 pierwsze składowe główne uzyskane na podstawie macierzy współczynników korelacji genotypowych (GEN) oraz macierzy współczynników korelacji fenotypowych (FEN) w kolekcji roboczej pszenicy ozimej
Eigen values and proportion of overall variation among genotypes explained by the first 3 principal components, based on genotypic correlation matrix (GEN) and phenotypic correlation matrix (FEN) in the winter wheat germplasm working collection

Składowa główna Principal component	Wartości własne Eigen values		Wyjaśniona wariancja (%) Explained variance	
	GEN	FEN	GEN	FEN
I	3,79	2,75	27,08	19,61
II	3,06	2,28	21,86	16,32
III	2,22	1,93	15,87	13,81
Skumulowana wariancja (%) Cumulative variance			64,81	49,74

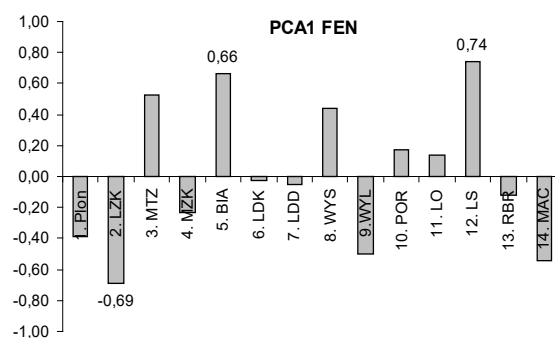
Natomiast współczynniki korelacji pomiędzy danymi składowymi głównymi, a badanymi cechami podano za pomocą wykresów (rysunki 1, 2, 3). Rysunki 1a, 2a, 3a dotyczą wyników PCA w oparciu o macierz współczynników korelacji genotypowych, natomiast rysunki 1b, 2b, 3b są wynikiem PCA na podstawie macierzy korelacji fenotypowych. W tabeli i na wykresach, przedstawionych w dalszej części pracy, PCA wykonana na macierzy współczynników korelacji genotypowych będzie oznaczana skrótem GEN, natomiast wykonana na macierzy współczynników korelacji fenotypowych skrótem FEN.

Trzy pierwsze składowe główne uzyskane na podstawie macierzy korelacji genotypowych wyjaśniały 64,81% wariancji całkowitej (tab. 1), czyli łącznej, wielowymiarowej zmienności rozważanych cech u obiektów w badanej kolekcji roboczej. Natomiast analogiczne składowe uzyskane w klasycznej PCA (na podstawie macierzy współczynników korelacji fenotypowych) wyjaśniały 49,74% wariancji całkowitej, czyli o około 15% wielocехowej zmienności mniej.



Rys. 1 a. Współczynniki korelacji pomiędzy zmiennymi a pierwszą składową główną. PCA wykonana na macierzy współczynników korelacji genotypowych

Fig. 1 a. Correlation coefficients between the variables and the first principal component. The PCA results are based on genotypic correlation matrix

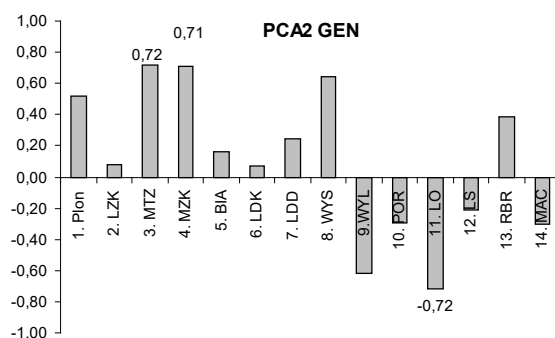


Rys. 1 b. Współczynniki korelacji pomiędzy zmiennymi a pierwszą składową główną. PCA wykonana na macierzy współczynników korelacji fenotypowych

Fig. 1 b. Correlation coefficients between the variables and the first principal component. The PCA results are based on phenotypic correlation matrix

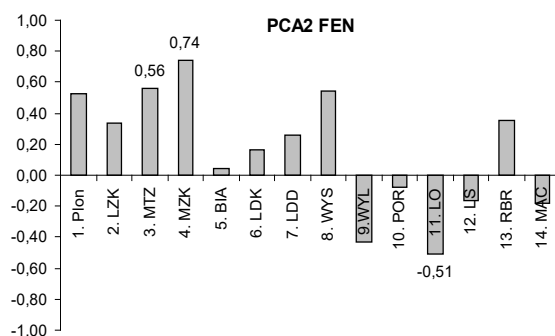
Dla obu proponowanych podejść (GEN i FEN) uzyskano zbieżne wyniki, z tym, że wartości bezwzględne współczynników korelacji dla klasycznej PCA (FEN) są znacznie mniejsze dla wszystkich rozpatrywanych cech (rys. 1a i 1b). Najsilniejszą korelację ujemną pierwszej składowej głównej stwierdzono z liczbą ziaren z kłosa ($r_{GEN} = -0,89$, $r_{FEN} = -0,69$), natomiast dodatnią z zawartością białka ($r_{GEN} = 0,79$, $r_{FEN} = 0,66$) i liczbą sedymentacji ($r_{GEN} = 0,85$, $r_{FEN} = 0,74$). Przeciwne znaki współczynników korelacji dla tych cech wskazują, że obiekty w kolekcji odznaczające się dużą liczbą ziaren z kłosa charakteryzowały się jednocześnie niską zawartością białka i LS. Dla cech skorelowanych z pierwszą składową główną stwierdzono duże wartości współczynników powtarzalności oraz bardzo silne korelacje genotypowe (Ukalski i in., 2008). Zatem, pierwszą składową główną można traktować jako przejaw czynnika genetycznego (poligenów) warunkującego te cechy.

Selekcja obiektów hodowlanych na podstawie wartości fenotypowych (tzn. średnich z lat) każdej z tych cech współlistnieje z równoczesną selekcją na wszystkie cechy skorelowane z pierwszą składową główną. Wobec tego postępowi hodowlanemu dla jednej z tych cech towarzyszą równocześnie niekonicznie korzystne zmiany pozostałych cech uwarunkowanych genetycznie przez ten czynnik.



Rys. 2 a. Współczynniki korelacji pomiędzy zmiennymi a drugą składową główną. PCA wykonana na macierzy współczynników korelacji genotypowych

Fig. 2 a. Correlation coefficients between the variables and the second principal component. The PCA results are based on genotypic correlation matrix



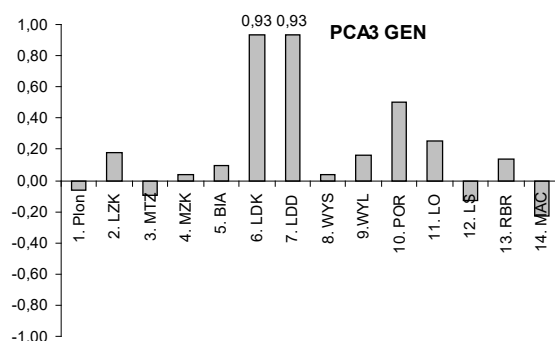
Rys. 2 b. Współczynniki korelacji pomiędzy zmiennymi a drugą składową główną. PCA wykonana na macierzy współczynników korelacji fenotypowych

Fig. 2 b. Correlation coefficients between the variables and the second principal component. The PCA results are based on phenotypic correlation matrix

Druga składowa główna wyjaśniła 21,86% zmienności całkowitej badanych odmian i rodów w kolekcji w podejściu GEN, a w podejściu FEN 16,32% (tab. 1). Najwyższe bezwzględne wartości współczynników korelacji w przypadku podejścia GEN (rys. 2a) uzyskano pomiędzy omawianą składową a cechami: MTZ ($r_{GEN}=0,72$), LO ($r_{GEN}=-0,72$) i MZK ($r_{GEN}=0,71$). Zatem można stwierdzić, że występujące w badanej kolekcji roboczej formy plenne charakteryzowały się jednocześnie niskimi właściwościami

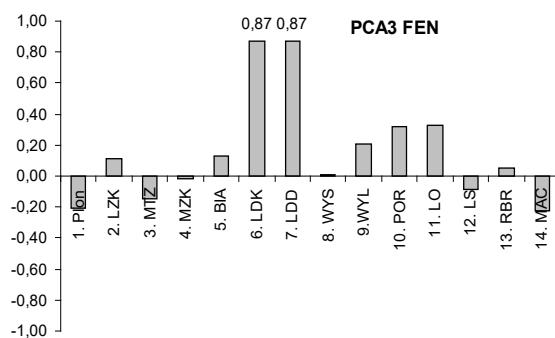
technologicznymi. Natomiast w podejściu FEN podobną wartość współczynnika korelacji stwierdzono tylko dla MZK ($r_{FEN} = 0,74$; rys. 2b). MTZ ($r_{FEN} = 0,56$) i LO ($r_{FEN} = -0,51$) to cechy, dla których, w podejściu FEN (rys. 2b), uzyskano niższe bezwzględne wartości współczynników korelacji z drugą składową główną. Liczba opadania jest cechą, dla której stwierdzono (Ukalski i in., 2008) najmniejszą, wśród badanych cech, wartość współczynnika powtarzalności. Zatem PCA na podstawie macierzy współczynników korelacji fenotypowych nie wychwyciła znaczącego wpływu dyskryminacyjnego cechy podatnej na warunki środowiskowe.

Z trzecią składową główną (rys. 3 a i 3 b) najsilniej skorelowane były LDK ($r_{GEN} = 0,93$, $r_{FEN} = 0,87$) i LDD ($r_{GEN} = 0,93$, $r_{FEN} = 0,87$). Wyniki dla obu proponowanych podejść są całkowicie zbieżne.



Rys. 3 a. Współczynniki korelacji pomiędzy zmiennymi a trzecią składową główną. PCA wykonana na macierzy współczynników korelacji genotypowych

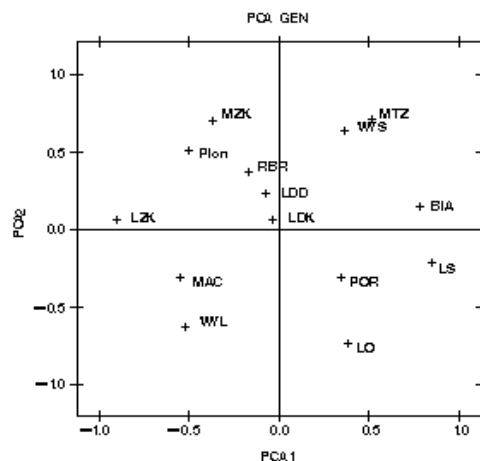
Fig. 3 a. Correlation coefficients between the variables and the third principal component. The PCA results are based on genotypic correlation matrix



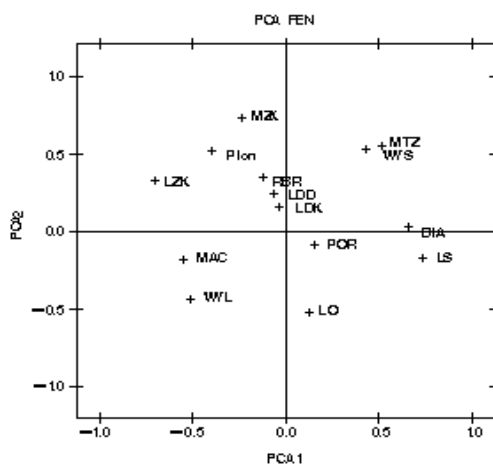
Rys. 3 b. Współczynniki korelacji pomiędzy zmiennymi a trzecią składową główną. PCA wykonana na macierzy współczynników korelacji fenotypowych

Fig. 3 b. Correlation coefficients between the variables and the third principal component. The PCA results are based on phenotypic correlation matrix

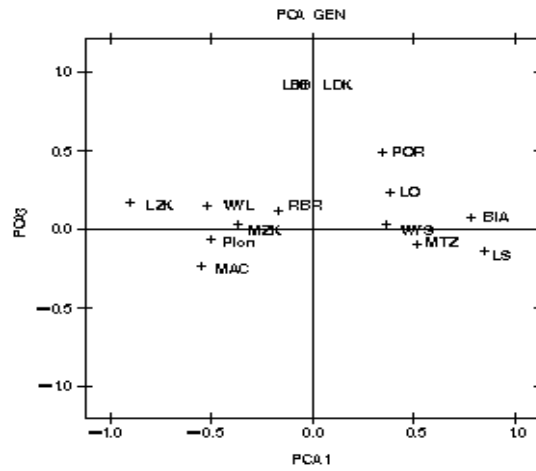
Rozpatrywane cechy można przedstawić graficznie w przestrzeni trzech składowych głównych. Wykres trójwymiarowy może być nieczytelny, dlatego w pracy przedstawiono dwa wykresy dwuwymiarowe obrazujące rozmieszczenie cech w układzie wybranych dwóch składowych głównych (rysunki 4 i 5).



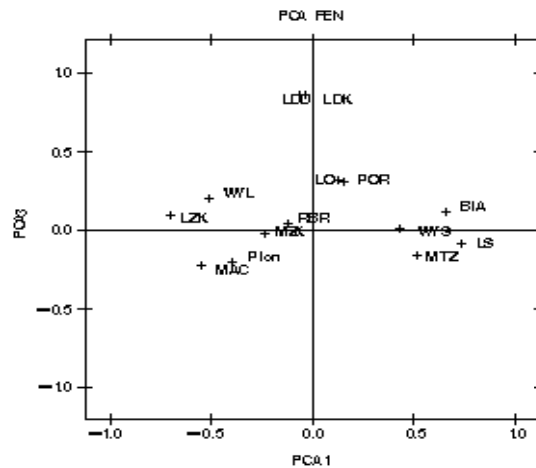
Rys. 4 a. Rozmieszczenie 14 zmiennych w przestrzeni pierwszej i drugiej składowej głównej. PCA wykonana na macierzy współczynników korelacji genotypowych
 Fig. 4 a. Scatter plot of 14 variables for the first and second principal components. The PCA results are based on genotypic correlation matrix



Rys. 4 b. Rozmieszczenie 14 zmiennych w przestrzeni pierwszej i drugiej składowej głównej. PCA wykonana na macierzy współczynników korelacji fenotypowych
 Fig. 4 b. Scatter plot of 14 variables for the first and second principal components. The PCA results are based on phenotypic correlation matrix

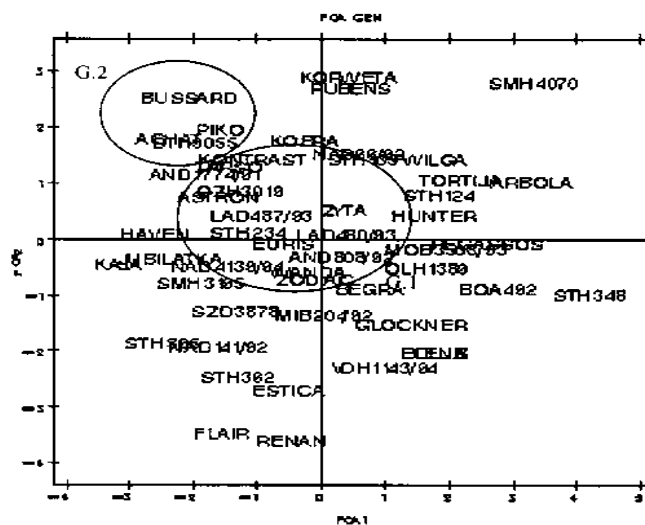


Rys. 5 a. Rozmieszczenie 14 zmiennych w przestrzeni pierwszej i trzeciej składowej głównej. PCA wykonana na macierzy współczynników korelacji genotypowych
Fig. 5 a. Scatter plot of 14 variables for the first and third principal components. The PCA results are based on genotypic correlation matrix

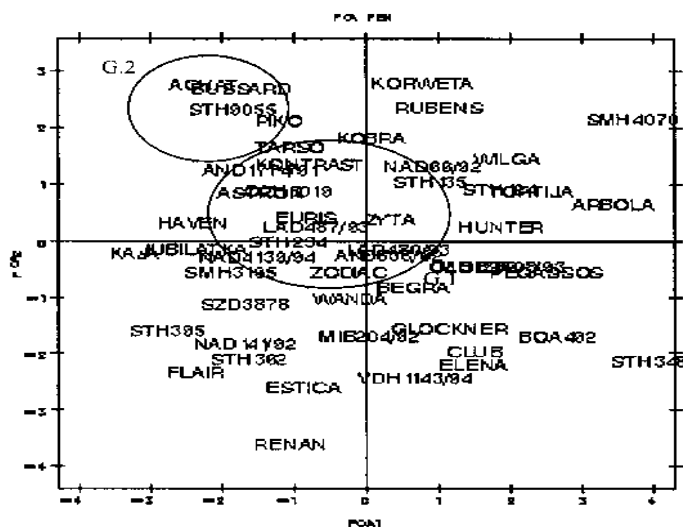


Rys. 5 b. Rozmieszczenie 14 zmiennych w przestrzeni pierwszej i trzeciej składowej głównej. PCA wykonana na macierzy współczynników korelacji fenotypowych
Fig. 5 b. Scatter plot of 14 variables for the first and third principal components. The PCA results are based on phenotypic correlation matrix

Wykresy te, podobnie jak omówione wcześniej (rysunki 1–3), mogą być pomocne w wykrywaniu cech najsilniej skorelowanych ze składowymi głównymi, a zatem są alternatywnym sposobem przedstawienia omówionych wyników.



Rys. 6 a. Rozmieszczenie 51 obiektów w przestrzeni pierwszej i drugiej składowej głównej. PCA wykonana na macierzy współczynników korelacji genotypowych
 Fig. 6 a. The distribution of 51 genotypes for the first and second principal components. The PCA results are based on genotypic correlation matrix



Rys. 6 b. Rozmieszczenie 51 obiektów w przestrzeni pierwszej i drugiej składowej głównej. PCA wykonana na macierzy współczynników korelacji fenotypowych
 Fig. 6 b. The distribution of 51 genotypes for the first and second principal components. The PCA results are based on phenotypic correlation matrix

Cechy charakteryzujące się największymi bezwzględnymi wartościami pierwszej (LZK, LS, BIA, rys. 4 a i b), drugiej (LO, MZK, MTZ) i trzeciej składowej głównej (LDD, LDK, rys. 5 a i b) najsilniej różnicują badane obiekty kolekcyjne. Natomiast cechy, których współrzędne, w układzie omawianych składowych głównych, są bliskie wartościom zerowym charakteryzują się małym zróżnicowaniem w obrębie rozważanej kolekcji, czyli są to cechy o słabej sile dyskryminacyjnej.

W przestrzeni składowych głównych można przedstawić również rozmieszczenie wszystkich badanych 51 odmian i rodów hodowlanych (rys. 6). Wykres taki pozwala na scharakteryzowanie kolekcji jednocześnie pod względem tych cech, które wykazywały silne skorelowanie ze składowymi głównymi. Na wykresach 6a i 6b zaznaczono dwie przykładowe grupy obiektów oznaczone jako G.1 i G.2. Grupy te na wykresach 6a i 6b nie zawierają tych samych obiektów ze względu na wykorzystanie w analizie PCA różnych macierzy współczynników korelacji. Grupa G.1 (rys. 6 a) charakteryzuje się, na poziomie genotypowym, wartościami cech LZK, LS i BIA (skorelowanych z pierwszą składową główną) zbliżonymi do ich średnich wartości w całej kolekcji i podobnie średnimi wartościami cech MTZ, LO i MZK (skorelowanych z drugą składową główną), natomiast grupa G.1 na rys. 6 b obejmuje obiekty o przeciętnych wartościach cech LZK, LS i BIA oraz MZK. Obiekty z grupy G.2 (rys. 6a) charakteryzują się, w stosunku do średnich wartości cech, podwyższoną wartością LZK i obniżonymi wartościami cech LS i BIA, oraz obniżoną wartością LO i podwyższonymi wartościami MTZ i MZK. Natomiast na wykresie 6b obiekty w obrębie grupy G.2 charakteryzują się na poziomie fenotypowym, w stosunku do średnich wartości cech, podwyższoną wartością cechy LZK i obniżonymi wartościami cech LS i BIA oraz podwyższonymi wartościami MZK. W podobny sposób można scharakteryzować wszystkie obiekty kolekcyjne.

Warto podkreślić, że przedstawione wykresy (rysunki 4–6) są tym lepszym obrazem relacji rozważanych cech (rys. 4 i 5), a także zróżnicowania wielocechowego obiektów kolekcyjnych (rys. 6), im większa część ogólnej zmienności (ogólnej wariancji) między badanymi obiektami została wyjaśniona przez składowe główne. Zatem wykresy uzyskane na podstawie macierzy współczynników korelacji genotypowych są dokładniejszym odzwierciedleniem relacji między obiektowych i między-cechowych w badanej kolekcji roboczej pszenicy ozimej.

WNIOSKI

Analizę składowych głównych dla rozważanych cech i obiektów w kolekcji roboczej pszenicy ozimej przeprowadzono dwójako: klasycznie, tj. na podstawie macierzy korelacji fenotypowych, oraz na podstawie macierzy korelacji genotypowych, czyli nieobserwowalnych efektów genotypowych badanych cech. Porównanie obu podejść pozwala na wyciągnięcie następujących wniosków:

1. Pierwsze trzy składowe główne uzyskane na podstawie macierzy współczynników korelacji genotypowych wyjaśniły o 15% więcej ogólnej zmienności między obiektowej niż przy zastosowaniu macierzy współczynników korelacji fenotypowych. Dla większości badanych cech uzyskano także bezwzględnie większe wartości

- współczynników korelacji pomiędzy cechami a trzema pierwszymi składowymi głównymi wyprowadzonymi z macierzy korelacji genotypowych niż pochodzących z macierzy korelacji fenotypowych. Spowodowane jest to tym, że współczynniki korelacji fenotypowych pomiędzy cechami są wypadkową współzależności genotypowych i środowiskowych.
2. Oba zastosowane podejścia w PCA pozwoliły na dokładne, wielocechowe scharakteryzowanie badanej kolekcji i znalezienie cech najsilniej różnicujących badane obiekty kolekcyjne. Analiza składowych głównych wykazała istnienie trzech najważniejszych czynników genetycznych, którymi są zespoły poligeniczne odpowiedzialne kolejno za: kształtowanie się liczby ziaren z kłosa, zawartości białka i liczby sedymentacji; masy ziaren z kłosa, masy tysiąca nasion oraz liczby opadania; cechy fenologiczne, tj. liczbę dni do kłoszenia i liczbę dni do dojrzałości.
 3. Wykorzystanie w PCA macierzy współczynników korelacji genotypowych pozwoliło dodatkowo na stwierdzenie znaczącego zróżnicowania obiektów w kolekcji roboczej pod względem liczby opadania. Oznacza to, że dyskryminacyjna siła cech podatnych na wpływy środowiska może nie zostać stwierdzona na podstawie analizy opartej na macierzy współczynników korelacji fenotypowych.

LITERATURA

- Abdi A., Bejkele E., Asfaw Z., Teshome A. 2002. Patterns of morphological variation of sorghum (*Sorghum bicolor* (L.) Moench) landraces in qualitative characters in North Shewa and South Welo, Ethiopia. *Hereditas* 137: 161 — 172.
- Ayana A., Bekele E. 1999. Multivariate analysis of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from Ethiopia and Eritrea. *Gen. Res. Crop Evol.* 46: 273 — 284.
- Erikson L., Johansson E., Kettaneh-Wold N., Wold S. 1999. Introduction to multi- and megavariate data analysis using projection methods (PCA & PLS). Umetrics AB, Umea, Sweden.
- Everitt B. S., Dunn G. 1992. *Applied Multivariate Data Analysis*. Oxford University Press, New York.
- Flores F., Gutierrez J. C., Lopez J., Moreno M. T., Cuberto J. I. 1997. Multivariate analysis approach to evaluate a germplasm collection of *Hedysarum coronarium* L. *Gen. Res. Crop Evol.* 44: 545 — 555.
- Hotteling H. 1933. Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology* 24: 417 — 441, 498 — 520.
- Hotteling H. 1936. Simplified Computation of Principal Components. *Psychometrika* 1:27 — 35.
- Iezzoni A. F., Pritts M. P. 1991. Application of Principal Components Analysis to Horticultural Research. *Hort. Sci.* 26(4):334 — 338.
- Jolliffe I. T. 1986. *Principal components analysis*. Springer-Verlag, New York.
- Khattree R., Naik D. N. 2000. *Multivariate data reduction and discrimination with SAS software*. SAS Institute Inc., Cary, NC.
- Krzyszowski W. J. 1988. *Principles of multivariate analysis: a users's perspective*. Oxford University Press, Oxford. 563 pp.
- Leguizamón J., Badenes M. L. 2003. Multivariate analysis as a tool for germplasm studies, Example of Analysis of Germplasm Loquat Data. *Acta Hort.* 606: 29 — 35.
- Littell R. C., Milliken G. A., Stroup W.W., Wolfinger R. D. 1996. *SAS system for mixed models*. SAS Institute Inc., Cary, NC.
- Mars M., Marrakchi M. 1999. Diversity of pomegranate (*Punica granatum* L.) germplasm in Tunisia. *Genet. Resour. Crop. Evol.* 46: 461 — 467.
- Martinez-Calvo J. 2007. Study of a germplasm collection of loquat (*Eriobotrya japonica* Lindl.) by multivariate analysis. *Genet. Resour. Crop Evol.* 55(5):695 — 703.

- Mądry W. 2007. Metody statystyczne do oceny różnorodności fenotypowej dla cech ilościowych w kolekcjach roślinnych zasobów genowych. Zesz. Probl. Post. Nauk Rol. 517: 21 — 41.
- Mohammadi S.A., Prasanna B.M. 2003. Analysis of genetic diversity in crop plants-Salient statistical tools and considerations. Crop Sci. 43:1235 — 1248.
- Pearson K. 1901. On lines planes of closest fit to a system of points in space. Philosophical Magazine 2:557 — 572.
- Perez-Gonzalez S. 1992. Associations among morphological and phenological characters representing apricot germplasm in Central Mexico. J. Am. Soc. Hort. Sci. 117:486 — 490.
- Rao C.R. 1964. The Use and Interpretation of Principal Components in Applied Research. Sankhya A26:329 — 358.
- Rojas W., Barriga P., Figueroa H. 2000. Multivariate analysis of the genetic diversity of Bolivian quinoa germplasm. Plant Genetic Resources Newsletter 122: 16 — 23.
- Rotondi A., Magli M., Ricciolini C., Baldoni L. 2003. Morphological and molecular analyses for the characterization of a group of Italian olive cultivars. Euphytica 132: 129 — 137.
- SAS/STAT User's Guide, Version 8.2. 2002. SAS Institute, Cary NC.
- Skinner D.Z., Barchan G.R., Auricht G., Hughes S. 1999. A method for the efficient management and utilization of large germplasm collections. Crop Sci. 39:1237 — 1242.
- Ukalska J., Mądry W., Ukalski K., Masny A., 2007. Wielowymiarowa ocena różnorodności fenotypowej w kolekcji zasobów genowych truskawki (*Fragaria* × *ananassa* Duch.) Część I: analiza zmienności. Zesz. Probl. Post. Nauk Rol. 517: 749 — 758.
- Ukalski K., Ukalska J., Śmiałowski T., Mądry W. 2008. Badanie zmienności i współzależności cech użytkowych w kolekcji roboczej pszenicy ozimej (*Triticum aestivum* L.) za pomocą metod wielowymiarowych. Część I. Korelacje fenotypowe i genotypowe. Biul. IHAR 249: 35 — 43.
- Upadhyaya H. D., Mallikarjuna Swamy B. P., Kenchana Goudar P. V., Kullaiswamy B.Y., Singh S. 2005. Identification of diverse groundnut germplasm through multienvironment evaluation of a core collection for Asia. Field Crops Research 93: 293 — 299.