

MARCIN STUDNICKI <sup>1</sup>  
WIESŁAW MĄDRY <sup>1</sup>  
TADEUSZ ŚMIAŁOWSKI <sup>2</sup>

<sup>1</sup> Katedra Doświadczalnictwa i Bioinformatyki, SGGW Warszawa

<sup>2</sup> Instytut Hodowli i Aklimatyzacji Roślin, Zakład Roślin Zbożowych w Krakowie

## Porównanie efektywności metod statystycznych tworzenia kolekcji podstawowej na przykładzie pszenicy jarej

### Comparison of statistical methods to development of a core collection for a spring wheat collection

Kolekcje podstawowe powstają z ograniczonej liczby obiektów zgromadzonych dotychczas w kolekcji zasobów genowych, wybranych tak aby reprezentowały zmienność całej kolekcji. Celem tworzenia kolekcji podstawowych jest efektywniejsze wykorzystanie aktualnie zgromadzonych zasobów genowych w programach hodowlanych roślin. Celem pracy było zaprezentowanie, a także ocena efektywności różnych metod wyboru obiektów, metod grupowania i wielkości kolekcji podstawowych. W pracy, kolekcje podstawowe, powstały w oparciu o kolekcję roboczą pszenicy jarej, zgromadzoną w Zakładzie Roślin Zbożowych Instytutu Hodowli i Aklimatyzacji Roślin w Krakowie. 196 obiektów pszenicy jarej było ocenionych pod względem 10 cech fenotypowych w latach 1996–2001. Kolekcja podstawowa pszenicy jarej powstała w wyniku trzyetapowego podejścia. W pierwszej kolejności przeprowadzono grupowanie 196 obiektów. Oceniono dwie metody analizy skupień — metodę Warda i metodę UPMGA (średniej odległości między obiektami w skupieniach). W drugim kroku wyznaczono liczbę obiektów wchodzących do kolekcji podstawowej z wcześniej wyznaczonych grup. W pracy zastosowano dwa podejścia metodę proporcjonalną i metodę logarymiczną. W ostatnim etapie wybrano obiekty z kolekcji wyjściowej do tworzonej kolekcji podstawowej. Wielkość utworzonych kolekcji podstawowych wynosiła 10%, 15%, 20%, 25% i 30% kolekcji wyjściowej. Procent istotnie zróżnicowanych średnich oraz procent istotnie zróżnicowanych wariancji, zostały wykorzystane do wskazania najbardziej efektywnej metody wyboru obiektów, techniki grupowania i wielkości utworzonych kolekcji podstawowych.

**Słowa kluczowe:** analiza skupień, kolekcje podstawowe, pszenica jara, zasoby genowe roślin uprawnych

A core collection consists of a limited number of the accessions gathered in an existing collection, chosen to represent the genetic variability within the whole collection of plant genetic resources. The purpose of creating core collections is to improve the utilization of genetic resources in crop improvement programs. The aim of this work was both to present and evaluate different procedures of sampling and clustering and to assess the size of core collections. This paper describes the development of a core subset of the spring wheat germplasm collection maintained at the Department of Cereal

Breeding and Quality Evaluation of the Plant Breeding and Acclimatization Institute in Kraków. 196 accessions of spring wheat were evaluated for 10 phenotypic traits in the seasons 1996–2001. The spring wheat core collection was established in three steps. In the first step a total of 196 accessions were stratified into homogenous groups. Two methods of cluster analysis: Ward's method and UPGMA (Unweighted Pair Group Method with Arithmetic mean) were comparatively used. In the second step, a number of entries per group was determined. Two strategies: proportional and logarithmic were used. In the final step of establishing core collections the actual entries were chosen. Five core collections contained 10%, 15%, 20%, 25% and 30% of objects from the whole collection. The mean percentage difference (MD%) and the variance difference percentage (VD%) were applied for screening optimal sampling strategy, cluster methods and size of core collections.

**Key words:** cluster analysis, core collections, plant genetic resources, spring wheat

## WSTĘP

Kolekcje roślinnych zasobów genowych, zwane też bankami genów, zakładane są na całym świecie, zabezpieczają istniejącą zmienność biologiczną roślin uprawnych i roślin spokrewnionych z nimi oraz innych ważnych roślin, w celu jej długoterminowego zabezpieczenia, oceny oraz przekazania dla potrzeb hodowców i innych badaczy. Od ponad 30 lat obserwujemy rozwój działalności na rzecz zasobów genowych, która polega przede wszystkim na systematycznym włączaniu nowych obiektów do kolekcji. W wyniku tych działań wiele kolekcji roślinnych zasobów genowych staje przed problemami związanymi z dużą liczbą zgromadzonych obiektów, polegającymi na trudnościach w realizacji podstawowych celów, dla jakich zostały one stworzone (Brown 1989; Frankel, 1984; Frankel i Brown, 1984; van Hintum i in., 2000). Utrzymanie dużych kolekcji przyczynia się do powstawania trudności w zachowaniu na odpowiednim poziomie żywotności zgromadzonych obiektów, utrudnione jest także pełne i właściwe wykorzystanie zgromadzonej tam zmienności przez hodowców i badaczy. Prowadzenie bardzo licznych kolekcji jest często nieuzasadnione ekonomicznie i trudne organizacyjnie (Frankel, 1984; Frankel i Brown, 1984).

W celu rozwiązania problemów związanych z dużą liczbą obiektów w kolekcjach zasobów genowych roślin uprawnych, proponuje się tworzenie kolekcji podstawowych (ang. core collection). Kolekcja podstawowa powstaje z określonego wyboru pewnej liczby obiektów spośród znajdujących się w kolekcji wyjściowej zasobów genowych (kolekcji aktualnie zgromadzonej), zwanej też kolekcją całkowitą (Frankel i Brown, 1984). Wybór obiektów powinien być tak przeprowadzony, aby reprezentowały one dobrze zmienność fenotypową i genotypową kolekcji wyjściowej, zawartą w stosunkowo małej liczbie obiektów. Kolekcje podstawowe tworzy się metodami próby warstwowej (Brown, 1989; Frankel i Brown, 1984; van Hintum i in., 2000).

Opracowano wiele technik służących do tworzenia kolekcji podstawowych, wykorzystują one różne kryteria oceny zmienności zgromadzonych obiektów. Stosowana jest ocena różnorodności fenotypowej, ocena zmienności genotypowej w oparciu o markery molekularne, a także ocena zmienności na podstawie danych ekologiczno–geograficznych (Balakrishnan i in., 2000; Crossa i in., 2005; Ghamkhar i in., 2008; Hu i in., 2000; Upadhyaya i in., 2007; van Hintum i in., 2000; Xu i in., 2006).

Celem pracy było badanie efektywności 8 technik statystycznych proponowanych do tworzenia kolekcji podstawowej, na podstawie obserwacji cech fenotypowych hodowlanej kolekcji roboczej pszenicy jarej. Każdą technikę stosowano dla pięciu wielkości kolekcji podstawowej, tj. 10, 15, 20, 25 i 30% kolekcji wyjściowej.

## MATERIAŁ I METODY

### Opis kolekcji wyjściowej

Materiałem źródłowym, stanowiącym kolekcję wyjściową do tworzenia kolekcji podstawowych za pomocą badanych technik statystycznych, była kolekcja robocza pszenicy jarej, zgromadzona i prowadzona dla celów hodowlanych w Zakładzie Roślin Zbożowych Instytutu Hodowli i Aklimatyzacji Roślin w Krakowie. Kolekcja wyjściowa zawierała 196 obiektów, które były zarówno liniami hodowlanymi przekazanymi przez hodowców z różnych ośrodków, odmianami zarejestrowanymi w Polsce i zagranicą, jak i obiektami uzyskanymi z różnych banków genów. Doświadczenia polowe z tymi obiektami przeprowadzono w latach 1996–2001 w miejscowości Kończewice, położonej w województwie kujawsko-pomorskim, w powiecie toruńskim w gminie Chełmża, na terenie oddziału Spółki Hodowla Roślin Strzelce Sp. z o.o. należącej do Grupy IHAR. Materiał siewny z każdego obiektu wysiano na pojedynczych (jednopowtórzeniowych) poletkach o powierzchni 10 m<sup>2</sup>. Poszczególne obiekty były badane przez różną liczbę lat, od 2 do 6. Zatem, jednostki doświadczalne (poletka) w całym doświadczeniu stanowiły niekompletną klasyfikację podwójną o postaci obiekty × lata, z co najwyżej jedną obserwacją w podklasie.

W trakcie doświadczenia obserwowano 10 ilościowych cech rolniczych. Na polu oceniano liczbę dni od 1.V. do terminu kłoszenia i dojrzałości oraz wysokość roślin. Pomiary te dokonywano na podstawie obserwacji 50 roślin losowo wybranych z poletka. W czasie wegetacji roślin określono porażenie chorobami liści i porażenie chorobami kłosa, według 9-stopniowej skali bonitacyjnej (porządkowej). W czasie zbioru określono plon ziarna z poletka i jego składowe — liczbę ziaren z kłosa, masę ziaren z kłosa i masę 1000 ziaren. Oceniano także cechę technologiczną ziarna, tj. procentową zawartość białka w ziarnie, oznaczoną metodą bliskiej podczerwieni NIR.

### Metody statystyczne do tworzenia kolekcji podstawowych

#### Ocena średnich i parametrów zmienności badanych cech w kolekcji wyjściowej

Obserwacje badanych cech 196 obiektów w niekompletnej klasyfikacji podwójnej o postaci obiekty × lata z co najwyżej jedną obserwacją w podklasie, były analizowane na podstawie następującego modelu mieszanego analizy wariancji (Hartung i Piepho, 2005; Littell i in., 2006; Piepho i Mohring, 2005):

$$y_{ij} = m + g_i + r_j + e_{ij}$$

gdzie:

$y_{ij}$  — jest obserwacją danej cechy u  $i$ -tego obiektu ( $i = 1, \dots, a$ ) w  $j$ -tym roku ( $j = 1, \dots, b$ ),

$m$  — jest średnią ogólną,

$g_i$  — jest losowym efektem głównym  $i$ -tego obiektu,

$r_j$  — jest stałym efektem głównym  $j$ -tego roku,

$e_{ij}$  — jest składnikiem losowym, obejmującym efekty interakcji obiekt  $x$  rok oraz błąd doświadczalny,

Ze względu na niekompletność sklasyfikowanych danych, losowe efekty główne dla obiektów  $g_i$ , dotyczące każdej cechy, zostały oszacowane za pomocą najlepszych nie-obciążonych liniowych predyktorów, nazywanych BLUP (ang. the best linear unbiased predictor) (Hartung i Piepho, 2005; Piepho i Mohring, 2005). W celu wyznaczenia estymatorów BLUP zastosowano metodę największej wiarygodności z restrycją REML (ang. Restricted Maximum Likelihood). Oceny typu BLUP dla efektów obiektowych  $g_i$  każdej cechy posłużyły do wyznaczenia ocen średnich genotypowych, według wzoru:

$$\hat{m}_i = \hat{m} + \hat{g}_i$$

gdzie:

$\hat{m}_i$  jest nieobciążoną oceną średniej genotypowej dla danej cechy (oceną średniej cechy w populacji lat) dotyczącej  $i$ -tego obiektu,  $\hat{m}$  jest oceną średniej ogólnej,  $\hat{g}_i$  jest oceną BLUP losowego efektu głównego, dotyczącego  $i$ -tego obiektu.

Tak wyznaczone średnie genotypowe,  $\hat{m}_i$ , dla każdej cechy zostały wykorzystane do oceny średnich i zmienności cech w kolekcji roboczej, stanowiącej kolekcję wyjściową oraz do dalszych prac nad utworzeniem kolekcji podstawowych.

#### Opis technik statystycznych do tworzenia kolekcji podstawowych

Techniki te były kombinacjami metod lub kryteriów, stosowanych w każdym z trzech etapów tworzenia kolekcji podstawowej. Dla pierwszego etapu, polegającego na grupowaniu obiektów jednorodnych (tworzenie warstw w próbie reprezentatywnej) zastosujemy dwie metody hierarchicznej analizy skupień, tj. metodę UPGMA - średniej odległości między obiektami w skupieniach (ang. unweighted pair group method with arithmetic mean) oraz metodę Warda. Metoda UPGMA była efektywnie stosowana między innymi przez Ghamkhar i wsp. (2008), Hu i wsp. (2000), Xu i wsp. (2006) i Li i wsp. (2004). Metoda Warda jest oparta na statystycznym kryterium minimalizacji sumy kwadratów odchyleń wewnątrz grup (ang. error sum of squares — ESS). Stosowana była do tworzenia kolekcji podstawowych w wielu pracach (van Hintum i in., 2000; Crossa i Franco, 2004; Crossa i in., 2005; Hu i in., 2000; Xu i in., 2006; Li i in., 2004; Balakrishnan i in., 2000; Mahalakshmi i in., 2007; Padilla i in., 2007). W obu metodach analizy skupień oparto się na macierzy kwadratów odległości euklidesowej (Crossa i Franco, 2004). Natomiast, przecięcie dendrogramu następowało w miejscu, które odpowiadało liczbie grup (skupień), wyjaśniającej dominującą część (około 70%) całkowitej sumy kwadratów odchyleń między obiektami (Casler i van Santen, 2000; Mahalakshmi i in., 2007).

Drugi etap polega na wyznaczaniu liczby obiektów z każdej grupy (warstwy), wchodzących w skład kolekcji podstawowej. Zastosujemy w nim dwa kryteria, tj. kryterium proporcjonalności (P) i logarytmicznej proporcjonalności (L), które zaproponował Brown (1989). Według kryterium proporcjonalności z poszczególnych grup bierze się liczbę obiektów wchodzących do kolekcji podstawowych, która jest taką samą frakcją (częścią) liczby obiektów w grupie, jak frakcja obiektów z całej kolekcji wyjściowej, przechodzącej do kolekcji podstawowej. Zatem, im większa jest grupa jednorodna, tym więcej obiektów z niej przejdzie do tworzonej kolekcji podstawowej. Według kryterium

logarytmicznej proporcjonalności liczba obiektów z każdej grupy, przeznaczonych do kolekcji podstawowej jest taką samą frakcją logarytmu liczby obiektów w grupie, jak frakcja obiektów z całej kolekcji wyjściowej. Zatem, przy użyciu kryterium logarytmicznej proporcjonalności grupy mniej liczne są liczniej reprezentowane w kolekcji podstawowej, niż przy użyciu kryterium proporcjonalności (Brown, 1989; van Hintum i in., 2000).

W trzecim etapie przeprowadzony jest wybór obiektów z poszczególnych grup, o wyznaczonej liczebności (w drugim etapie), do tworzonej kolekcji podstawowej. W tej pracy stosowano dwie metody wyboru obiektów z grup (ang. sampling strategies), tj. metodę całkowicie losowego wyboru (van Hintum i in., 2000) oraz metodę preferowanego wyboru obiektów, w której badacz tworzący kolekcję podstawową subiektywnie decyduje jakie obiekty w grupach mają być wybrane. W klasycznym podejściu do metody preferowanego wyboru istnieje kilka kryteriów, pozwalających podjąć decyzję, który obiekt zostanie przekazany do kolekcji podstawowej, np.: wybierane są takie obiekty, które posiadają wartości skrajne (maksymalne lub minimalne) dla jednej lub kilku ocenianych cech (Li i in., 2004; van Hintum i in., 2000; Xu i in., 2006). Aby zobiektywizować ten sposób wyboru obiektów do kolekcji podstawowej, zaproponowaliśmy jako kryterium wyboru wartość pierwszej zmiennej kanonicznej obliczoną dla obiektów w kolekcji wyjściowej. Analiza zmiennych kanonicznych pozwala zidentyfikować cechy o dużym znaczeniu w wielocechowym różnicowaniu grup (Krzanowski, 1993). Do kolekcji podstawowej włączono obiekty o skrajnych (maksymalnych i minimalnych) wartościach pierwszej zmiennej kanonicznej.

W wyniku skojarzenia każdej z dwóch metod lub kryteriów, stosowanych w poszczególnych etapach tworzenia kolekcji podstawowej, uzyskano 8 technik statystycznych, których efektywność badano dla każdej z pięciu wielkości kolekcji podstawowej, tj. 20 obiektów (frakcja 10,2% kolekcji wyjściowej), 30 obiektów (frakcja 15,3% kolekcji wyjściowej), 40 obiektów (frakcja 20,4% kolekcji wyjściowej), 50 obiektów (frakcja 25,51% kolekcji wyjściowej) oraz 60 obiektów (frakcja 30,61% kolekcji wyjściowej). Zatem, zastosowano 40 postępowań (procedur) przy tworzeniu kolekcji podstawowych, powstałych jako kombinacje 5 wymienionych wielkości kolekcji podstawowych, dwóch metod wielocechowej analizy skupień (metoda UPMGA, metoda Warda), dwóch kryteriów wyznaczenia liczby obiektów z poszczególnych grup, wchodzących w skład kolekcji podstawowej (kryterium proporcjonalności – P, kryterium logarytmicznej proporcjonalności – L), a także dwóch metod wyboru obiektów z każdej grupy (wyboru losowego, wyboru preferowanego opartego na pierwszej zmiennej kanonicznej). Za pomocą każdej procedury utworzono jedną kolekcję podstawową.

#### Ocena efektywności 40 procedur przy tworzeniu kolekcji podstawowych

Ocenę efektywności i skuteczności badanych 40 procedur do tworzenia kolekcji podstawowych oparto na 2 kryteriach, stosowanych w wielu pracach (Hu i in., 2000; Li i in., 2004; Xu i in., 2006). Pierwszym z nich był procent istotnie zróżnicowanych średnich – MD%. Kryterium to opisuje procent liczby cech dla których stwierdzono istotnie zróżnicowanie średnich pomiędzy kolekcją wyjściową a kolekcją podstawową. Kolekcja podstawowa lepiej przenosi zmienność kolekcji wyjściowej, gdy wartość tego parametru jest bliższa zeru. Kryterium nie powinno przekraczać 20% cech istotnie zróżnicowanych,

aby uznać kolekcje podstawową za dobrze reprezentującą wielocechową zmienność aktualnie zgromadzonych obiektów (Hu i in., 2000). W celu zbadania istotności różnicy średnich pomiędzy kolekcją wyjściową a kolekcją podstawową, zastosowano test t-Studenta (Hu i in., 2000; Li i in., 2004; Xu i in., 2006), w przypadku nie stwierdzenia istotnego zróżnicowania wariancji dla cech w kolekcji wyjściowej i kolekcji podstawowej (spełnienia jednego z warunków do stosowania testu t-Studenta). W przeciwnym przypadku, zastosowano test t-Studenta z aproksymacją Cochra i Coxa (SAS Institute Inc. 2004).

Drugim zastosowanym kryterium do oceny efektywności proponowanych procedur tworzenia kolekcji podstawowej, był procent istotnie zróżnicowanych wariancji — VD%, opisujący procent liczby badanych cech dla których stwierdzono istotnie zróżnicowane wariancje pomiędzy kolekcją wyjściową a wyznaczoną kolekcją podstawową. Im niższa jest wartość tego kryterium (bliższa zero), tym kolekcja podstawowa lepiej reprezentuje wielocechową zmienność aktualnie zgromadzoną kolekcję zasobów genowych roślin uprawnych. Istotność różnicy wariancji pomiędzy kolekcją podstawową a wyjściową, zbadano za pomocą testu F (Hu i in. 2000, Li i in. 2004, SAS Institute Inc. 2004, Xu i in. 2006).

Aby obiektywnie wskazać, która z metod tworzenia kolekcji podstawowych, stosowanych w różnych etapach, istotnie różnicuje skuteczność (reprezentatywność kolekcji) zastosowano czteroczynnikową analizę wariancji. Badanymi czynnikami były trzy metody (lub kryteria) statystyczne oraz wielkość kolekcji podstawowej, zaś zmiennymi zależnymi były dwa stosowane kryteria efektywności procedur, tj. MD% i VD%.

Wszystkie obliczenia numeryczne wykonano za pomocą pakietu statystycznego SAS 9.1 (Littell i in., 2006; SAS Institute Inc., 2004). Wykorzystano następujące procedury statystyczne: MIXED — do estymacji efektów obiektowych  $g$ ; typu BLUP; CLUSTER — do przeprowadzenia analizy skupień metodą Warda i UPGMA; CANDISC — do wyznaczenia pierwszej zmiennej kanonicznej; TTEST — do wyznaczenia MD% oraz VD%, jako kryterium oceny efektywności procedur do tworzenia kolekcji podstawowej.

## WYNIKI I DYSKUSJA

W tabeli 1 zaprezentowano charakterystykę zmienności kolekcji roboczej pszenicy jarej (kolekcja wyjściowa) pod względem 10 ocenianych cech, na podstawie ocen średnich genotypowych otrzymanych za pomocą estymatorów typu BLUP dla wszystkich 196 obiektów. Największą zmiennością obiektową charakteryzowała się liczba dni do kłoszenia (współczynnik zmienności CV był równy 20,14%) oraz plon ziarna (współczynnik zmienności CV był równy 8,08%). Najmniejszą zmiennością charakteryzowały się następujące cechy: liczba dni do dojrzałości (współczynnik zmienności CV był równy 1,81%), zawartość białka w ziarnie (współczynnik zmienności CV był równy 3,56%). Wartość średnią plonu ziarna z poletka oszacowano na poziomie 67,1 dt/ha, natomiast kształtował się on w zakresie od 44,10 dt/ha do 79,16 dt/ha. Średnia wysokość 196 obiektów pszenicy jarej wynosiła 95,54 cm, minimalną zaobserwowaną wartość to 70,35 cm, natomiast maksymalną wartość to 111,53 cm. Kolekcja robocza pszenicy jarej

charakteryzowała się dużą zmiennością badanych cech, może się okazać doskonałym źródłem komponentów rodzicielskich w trakcie hodowli nowych odmian. Zwłaszcza dla takich cech jak długość okresu wegetacji i plon ziarna z poletka. Przyczyniając się jednocześnie do postępu biologicznego i hodowlanego w obrębie tego gatunku.

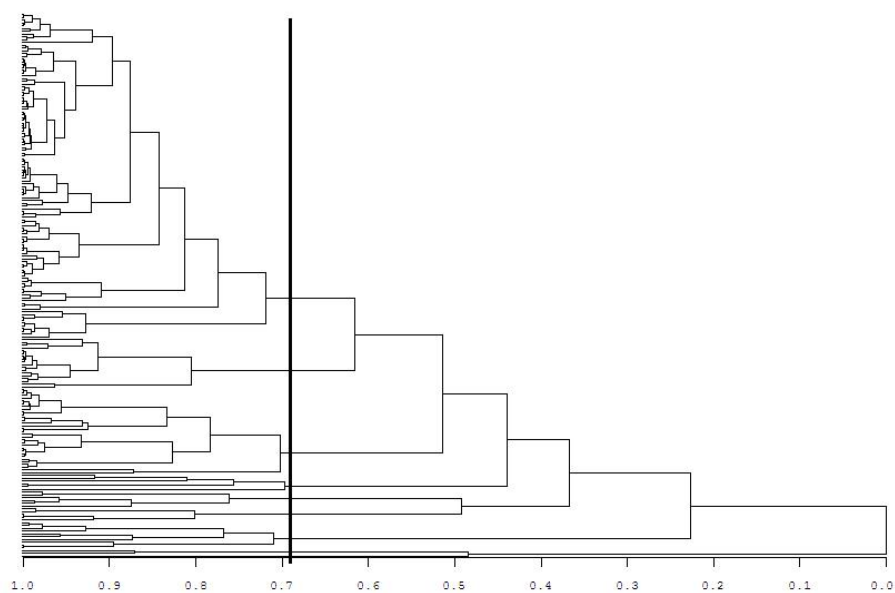
Tabela 1

**Charakterystyka zmienności kolekcji roboczej pszenicy jarej pod względem 10 cech**  
**Characteristics of variability of 10 traits in a spring wheat collection**

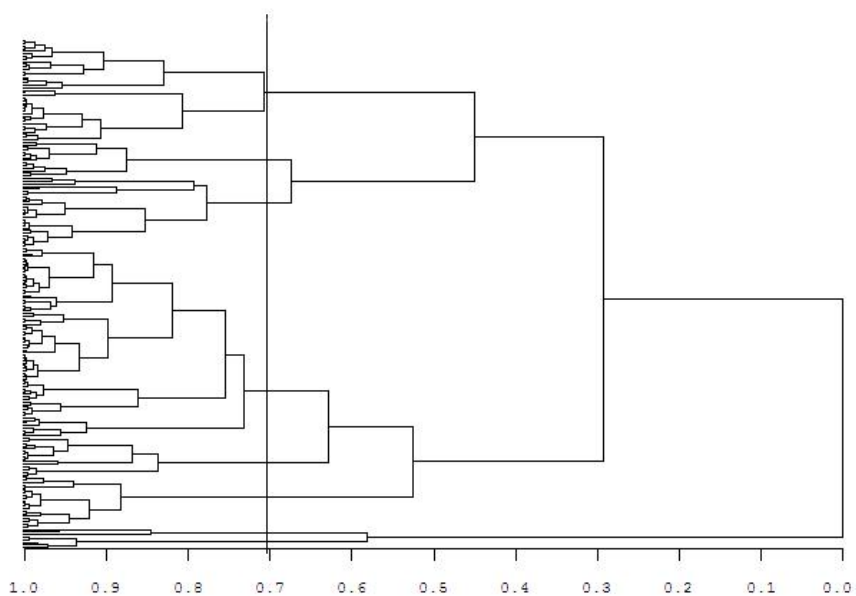
Cechy Traits	Średnia Mean	Odchylenie standardowe Standard deviation	Współczynnik Zmienności Coefficient of variability	Min	Max
Zawartość białko w ziarnie Protein content in grain	13,38	0,48	3,56	12,09	15,44
Liczba dni do dojrzałości Number of days to maturity	107,05	1,94	1,81	102,56	111,14
Liczba dni do kłoszenia Number of days to heading	13,38	2,69	20,14	6,94	19,93
Plon ziarna z poletka Grain yield	67,1	5,42	8,08	44,10	79,16
Wysokość rośliny Plant height	95,54	6,43	6,73	70,35	111,53
Liczba ziaren w kłosie Number of grains per spike	39,46	2,82	7,15	32,02	47,25
Masa ziaren w kłosie Weight of grains per spike	1,64	0,12	7,19	1,20	1,97
Masa 1000 ziaren 1000 grain weight	41,57	3,14	7,55	34,55	49,56
Porażenie chorobami liści Resistance to leaf diseases	5,02	0,39	7,69	3,82	6,01
Porażenie chorobami kłosa Resistance to spike diseases	6,1	0,44	7,18	4,62	6,94

W wyniku pierwszego etapu tworzenia kolekcji podstawowej, analizy skupień metodą UPGMA, 196 obiektów kolekcji wyjściowej zostało podzielonych na 9 grup jednorodnych. Przy takiej liczbie grup jednorodnych sumy kwadratów odchyłeń pomiędzy grupami wyjaśniły 70% całkowitej sumy kwadratów odchyłeń. Na rysunku 1 przedstawiono dendrogram podziału 196 obiektów pszenicy jarej, wraz z miejscem jego przecięcia, wykorzystując metodę UPGMA. Liczebność wyznaczonych grup kształtowała się w zakresie od 117 obiektów w grupie pierwszej, do 1 obiektu w grupie dziewiątej.

Natomiast analiza skupień metodą Warda pozwoliła wyznaczyć w kolekcji wyjściowej pszenicy jarej 8 grup jednorodnych. Podział na taką liczbę grup pozwoliła międzygrupowej sumie kwadratów odchyłeń dla wszystkich ocenianych cech wyjaśnić 71% całkowitej sumy kwadratów odchyłeń. Proces aglomeracji z wykorzystaniem tej metody 196 obiektów pszenicy jarej został zaprezentowany w postaci dendrogramu na rysunku 2. Największa liczba obiektów zostało zgromadzonych w grupie pierwszej — 73, najmniejszą liczbę obiektów stwierdzono w grupie ósmej — 3 obiekty pszenicy jarej.



**Rys. 1. Dendrogram uzyskany za pomocą metody UPMGA analizy skupień dla 196 obiektów pszenicy jarej**  
**Fig. 1. Dendrogram generated using UPMGA cluster analysis for 196 accessions**



**Rys. 2. Dendrogram uzyskany za pomocą metody Warda analizy skupień dla 196 obiektów pszenicy jarej**  
**Fig. 2. Dendrogram generated using Ward's cluster analysis for 196 accessions**



W tabeli 2 zaprezentowano wyniki analizy wariancji dla współczynnika istotnie zróżnicowanych średnich MD%. Stwierdzono istotny wpływ współdziałania pomiędzy rodzajem wykorzystywanej analizy skupień a sposobem wyboru obiektów ze skupień. Dla metody UPGMA lepszą (posiadała mniejszą wartość dla tego kryterium) okazała się metoda perforowanego wyboru obiektów do kolekcji podstawowej — średnia wartość badanego parametru wynosiła 2%. Natomiast dla metody Warda analizy skupień bardziej efektywną okazała się metoda losowego doboru obiektów ze skupień — średnia wartość badanego parametru 2%. Na podstawie tego parametru oceny efektywności technik służących do tworzenia kolekcji podstawowych, nie udało się stwierdzić innych zależności. Dla wszystkich utworzonych kolekcji podstawowych pszenicy jarej, procent istotnie zróżnicowanych średnich kształtował się w zakresie od 0 do 20%. Spełniając warunek opisany przez Hu i wsp. (2000) świadczący o dużej efektywności utworzonych kolekcji podstawowej i dobrze zachowujące zmienność biologiczną obiektów zgromadzonych w kolekcji wyjściowej. W dwudziestu czterech z pośród czterdziestu powstałych kolekcjach podstawowych nie stwierdzono istotnie różnych średnich, współczynnik MD% był równy zero. Parametr MD% w małym stopniu różnicował metody do tworzenia kolekcji podstawowych. Okazał się mało przydatny w badaniu efektywności technik statystycznych wykorzystywanych w trakcie naszych badań.

Tabela 2

**Analiza wariancji dla procentu istotnie zróżnicowanych średnich MD%**  
**Analysis of variance for the mean percentage difference MD%**

Źródło zmienności Source of variation	Liczba stopni swobody Degrees of freedom	Średnie kwadraty Mean squares	F-emp
Analiza skupień — A Cluster analysis — A	1	40	1,1
Liczba obiektów z grup — B Number of entries per group — B	1	90	2,47
Sposób wyboru z grup — C Sampling strategy — C	1	40	1,1
Wielkość kolekcji podstawowej — D Size of core collection — D	4	18,75	0,51
A×B	1	10	0,27
A×C	1	360	9,87*
B×C	1	10	0,27
A×D	4	33,75	0,93
B×D	4	46,25	1,27
C×D	4	58,75	1,61
Błąd — Error	17	36,47	

\*Istotne przy  $\alpha = 0,01$ ; Significant at  $\alpha = 0.01$

W wyniku przeprowadzonej analizy wariancji dla współczynnika istotnie zróżnicowanych wariancji VD% (tab. 3), stwierdzono istotny wpływ sposobu wyboru liczby obiektów z liczby obiektów ze skupień. Bardziej efektywnym postępowaniem okazała się metoda proporcjonalna do wyznaczania liczby obiektów z grup, średnia wartość dla parametru VD% wynosiła 10,5% cech. Dla metody logarytmicznej, kryterium to oszacowano na poziomie 29%.

Tabela 3

**Analiza wariancji dla procentu istotnie zróżnicowanych wariancji VD%**  
**Analysis of variance for the variance difference percentage VD%**

Źródło zmienności Source of variation	Liczba stopni swobody Degrees of freedom	Średnie kwadraty Mean squares	F-emp
Analiza skupień — A Cluster analysis — A	1	2,5	0,02
Liczba obiektów z grup — B Number of entries per group - B	1	3422,5	26,98*
Sposób wyboru z grup — C Sampling strategy — C	1	3802,5	29,97*
Wielkość kolekcji podstawowej — D Size of core collection — D	4	108,75	0,86
A×B	1	422,5	3,33
A×C	1	62,5	0,49
B×C	1	2722,5	21,46*
A×D	4	158,75	1,25
B×D	4	566,25	4,46*
C×D	4	8,75	0,07
A×B×C	1	62,5	0,49
Błąd — Error	16	126,875	

\*Istotne przy  $\alpha = 0,01$ ; Significant at  $\alpha = 0.01$

W tabeli 4 zaprezentowano porównanie pomiędzy tymi dwiema zastosowanymi metodami wyznaczenia liczebności obiektów z poszczególnych grup. W kolekcjach podstawowych tworzonych z wykorzystaniem metody logarytmicznej, większe znaczenie (więcej obiektów z takiej grup zostało włączonych to kolekcji podstawowej) odgrywały obiekty pochodzące z grup mniej licznych. Obiekty w grupach o małej liczebności charakteryzują się bardzo dużą jednorodnością (są wielocechowo bardzo podobne do siebie). Takie postępowanie wpływa na zubożenie zmienności w powstałej kolekcji podstawowej. Do kolekcji podstawowej włączane są obiekty podobne, nie wnoszące nowej zmienności. Powstałe kolekcje podstawowe, w których do wyznaczania liczby obiektów z grup wykorzystano metodę logarytmiczną słabo reprezentują kolekcję wyjściową, o czym świadczy wysoka wartość parametru VD%.

Tabela 4

**Przykład wyznaczenia liczby obiektów z grup do kolekcji podstawowej z zastosowaniem metody proporcjonalnej i logarytmiczno-proporcjonalnej, jeśli należy wybrać 60 z 196 obiektów kolekcji roboczej pszenicy jarej. (analiza skupień — metoda Warda)**

**Allocation of entries from groups to a core collection according to the proportional and logarithmic method, if 60 entries are to be selected from 196 accessions (cluster analysis — Ward's method)**

Grupa Group	Liczebność No. of accessions	Metoda proporcjonalna Proportional method	Metoda logarytmiczna Logarithmic method
1	73	22	12
2	39	12	10
3	27	8	9
4	20	6	8
5	15	5	7
6	14	4	7
7	5	2	4
8	3	1	3

Istotny wpływ na zróżnicowanie kryterium VD% posiadał również sposobu wyboru obiektów z grup. Lepszą, bardziej efektywną była metoda losowego doboru obiektów do kolekcji podstawowej (wartość współczynnika VD% — 10%), od metody preferowanego doboru obiektów do kolekcji (wartość współczynnika VD% — 29,5%). Metoda preferowanego wyboru okazała się mniej efektywna, gorzej reprezentowała zmienność fenotypową kolekcji wyjściowej, podobnie jak w pracach Li i wsp. (2004), Xu i wsp. (2006). Zastosowanie obiektywnego kryterium, jakim była wartość dla pierwszej zmiennej kanonicznej, nie poprawiło efektywności rozpatrywanej techniki do tworzenia kolekcji podstawowej.

Stwierdzono również istotną interakcję pomiędzy sposobem wyznaczenia liczby obiektów z poszczególnych skupień i metodami wyboru obiektów do kolekcji podstawowej. Losowy wybór obiektów okazał się korzystniejszy zarówno dla proporcjonalnego podejścia do wyznaczenia liczby obiektów z grup (średnia wartość współczynnika VD% — 9%) oraz metody logarytmicznej (średnia wartość współczynnika VD% — 11%). Istotny wpływ na kształtowanie się wartości VD% stwierdzono dla współdziałania pomiędzy sposobem wyznaczania liczby obiektów z poszczególnych skupień a wielkością kolekcji podstawowej. Najkorzystniejszą wielkością kolekcji podstawowej dla logarytmicznego podejścia do wyznaczenia liczby obiektów ze skupień okazała się liczba obiektów stanowiącej 10% kolekcji wyjściowej. Natomiast podejścia proporcjonalnego optymalną wielkością była kolekcja podstawowa stanowiąca 50 obiektów z kolekcji wyjściowej (25% aktualnie zgromadzonej kolekcji wyjściowej).

Wartości średnie współczynnika VD% dla obydwu zastosowanych metod analizy skupień nie różniły się istotnie. Na efektywność technik do tworzenia kolekcji podstawowej nie miała wpływu zastosowana metoda analizy skupień. Metoda Warda w porównaniu z metodą UPGMA tworzy grupy o zbliżonej do siebie oraz o stosunkowo małej liczebności. A dzięki zastosowaniu w trakcie procesu aglomeracji kryterium statystycznego opartego na minimalizacji sumy kwadratów odchylen wewnątrz grup, powstałe grupy są bardziej jednorodne. Z tych ze względów jest to metoda bardziej optymalna do zastosowania w trakcie tworzenia kolekcji podstawowych roślin uprawnych (Crossa i Franco, 2004; Crossa i in., 2000; Li i in., 2004). Powszechnie i często stosowaną metodą analizy skupień jest metoda UPGMA, uznana przez Xu i wsp. (2006) za najbardziej efektywną w trakcie tworzenia kolekcji podstawowej. W celu pogrupowania obiektów w kolekcji wyjściowej (aktualnie zgromadzonej) oprócz metod analizy skupień wykorzystuje się również podział formalny obiektów (van Hintum i in., 2000). Najczęściej obiekty grupowane są ze względu na rejon występowania czy pochodzenie botaniczne. Mogą być także klasyfikowane na podstawie cech jakościowy badanych roślin np.: barwy owoców, kwiatów czy kształtu liści.

Wielkość tworzonych kolekcji podstawowych nie miała wpływu na stopień reprezentatywności kolekcji wyjściowej. Najmniejsze kolekcje podstawowe liczące 20 obiektów (10% kolekcji wyjściowej) w były tak samo efektywne jak duże kolekcje liczące 60 obiektów (30% kolekcji wyjściowej). Duża redukcja liczebności (o 90%) kolekcji wyjściowej pszenicy jarej, w małym stopniu wpływała na zubożenie zmienność fenotypowej zgromadzonych dotychczas zasobów genowych pszenicy jarej. Decyzja o

wielkości tworzonej kolekcji podstawowych jest uzależniona od liczebności kolekcji wyjściowej i celu dla jakiego one powstają. Brown (1989) proponuje aby kolekcje podstawowe stanowiły około 10% dotychczas zgromadzonych zasobów genowych. Większość tworzonej kolekcji podstawowych stanowiło od 5% do 25% kolekcji wyjściowej, dla przykładu międzynarodowa kolekcja podstawowa jęczmienia (The International Barley Core Collection) stanowi 0,3% zgromadzonych na świecie zasobów tego gatunku (van Hintum i in., 2000).

#### WNIOSKI

1. Zachowanie (reprezentatywność) różnorodności fenotypowej w kolekcji podstawowej pszenicy jarej było niezależne od jej wielkości, już najmniejsze kolekcje podstawowe (około 10% kolekcji wyjściowej) dobrze reprezentowały zmienność fenotypową całej kolekcji pszenicy jarej.
2. Zmienność fenotypową kolekcji wyjściowej najlepiej reprezentowały kolekcje podstawowe, w których liczbę obiektów ze skupień wyznaczano zgodnie z zasadą proporcjonalności.
3. Lepszym sposobem wyboru obiektów ze skupień okazała się metoda losowa.
4. Przeprowadzone badania nie uzasadniają zalecania jednej konkretnej procedury, spośród badanych, wiele z nich ma podobną efektywność.

#### LITERATURA

- Balakrishnan R., Nair N., Sreenivasan T. 2000. A method for establishing a core collection of *Saccharum officinarum* L. germplasm based on quantitative-morphological data. *Genetic Resources and Crop Evolution* 47: 1 — 9.
- Brown A. H. D. 1989. Core collections: a practical approach to genetic resources management. *Genome* 31:818 — 824.
- Casler M. D., van Santen E. 2000. Patterns of variation in a collection of meadow fescue accessions. *Crop Sci.* 40: 248 — 255.
- Crossa J., Franco J. 2004. Statistical methods for classifying genotypes. *Euphytica* 153: 19 — 37.
- Crossa J., Franco J., Taba S., Shands H. 2005. A sampling strategy for conserving genetic diversity when forming core subsets. *Crop Science* 45: 1035 — 1044.
- Frankel O. H. 1984. Genetic perspectives of germplasm conservation. *Genetic Manipulation: Impact on Man and Society*. Cambridge University Press, Cambridge.
- Frankel O. H., Brown A.H.D. 1984. Plant genetic resources today: A critical appraisal. *Crop Genetic Resources: Conservation and Evaluation*. Allen and Unwin, Massachusetts.
- Ghamkhar K., Snowball R., Wintle B. J., Brown A. H. D. 2008. Strategies for developing a core collection of bladder clover (*Trifolium spumosum* L.) using ecological and agro-morphological data. *Australian Journal of Agricultural Research* 59: 1103 — 1112.
- Hartung K., Piepho H. P. 2005. A threshold model for multiyear genebank data based on different rating scales. *Crop Sci.* 45:1045 — 1051.
- Hu J., Zhu J., Xu H. 2000. Methods of constructing core collections by stepwise clustering with three sampling strategies based on the genotypic values of crops. *Theoretical and Applied Genetics* 101: 264 — 268.
- Krzanowski W.J. 1993. Principles of multivariate analysis, a user's perspective. Clarendon Press, Oxford
- Li C. T., Shi C. H., Wu J. G., Xu H. M., Zhang H. Z., Ren Y. L. 2004. Methods of developing core collections based on the predicted genotypic value of rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 108: 1172 — 1176.

- Littell R. C., Milligan G. A., Stroup W. W., Wolfinger R. D., Schabenberger O. 2006. SAS for mixed models. SAS Institute Inc. Cary, NC.
- Mahalakshmi V., Ng Q., Atalobhor J., Ogunsola D., Lawson M., Ortiz R. 2007. Development of a West African yam *Dioscorea* spp. core collection. *Genetic Resources and Crop Evolution* 57: 1817 — 1825.
- Padilla G., Cartea M. E., Ordas A. 2007. Comparison of several clustering methods in grouping kale landraces. *J. Amer. Soc. Hort. Sci.* 132: 387 — 395.
- Piepho H. P., Möhring J. 2005. Best Linear Unbiased Prediction of cultivar effects for subdivided target regions. *Crop Sci.* 45:1151 — 1159.
- SAS Institute Inc. 2004. SAS OnlineDoc® 9.1.3 Cary, NC.
- Upadhyaya H., Dwivedi S., Gowda C., Singh S., 2007. Identification of diverse germplasm lines for agronomic traits in a chickpea (*Cicer arietinum* L.) core collection for use in crop improvement. *Field Crops Research* 100: 320 — 326.
- van Hintum T., Brown A., Spillane C., Hodgkin T. 2000. Core collections of plant genetic resources. PGRI Technical Bulletin No. 3. International Plant Genetic Resources Institute, Rome.
- Xu H., Mei Y., Hu J., Zhu J., Gong P. 2006. Sampling a core collection of Island cotton (*Gossypium barbadense* L.) based on the genotypic values of fiber traits. *Genetic Resources and Crop Evolution* 53: 515 — 521.