

MARCIN STUDNICKI**WIESŁAW MĄDRY**

Katedra Doświadczalnictwa i Bioinformatyki, SGGW w Warszawie

Metodyka statystyczna pobierania próby do tworzenia kolekcji podstawowej roślinnych zasobów genowych: przegląd dorobku*

Sampling methodology to establish a core collection of plant genetic resources: an overview of research

Kolekcje podstawowe są podzbiorem obiektów wybranych z aktualnie zgromadzonej kolekcji (kolekcji wyjściowej) zasobów genowych, tak aby reprezentowały, z minimalną liczbą duplikatów, one różnorodność genetyczną w kolekcji wyjściowej. Tworzenie tego typu kolekcji ma na celu redukcję kolekcji wyjściowej do rozsądnej liczby obiektów co ułatwi systematyczną i pełną ocenę zmienności genetycznej w kolekcji dla wielu cech genotypowych oraz markerów molekularnych. Kolekcje podstawowe odgrywają ważną rolę w zarządzaniu i wykorzystaniu kolekcji zasobów genowych w badaniach i programach hodowli roślin. Opracowano wiele metod służących do tworzenia kolekcji podstawowych z już istniejących kolekcji roślinnych zasobów genowych. Ważnym aspektem w trakcie tworzenia kolekcji podstawowej jest dobór odpowiedniej metody pobierania próby. Metody pobierania próby są powszechnie stosowane do wyboru próby, która tworzą reprezentatywne kolekcje podstawowe z kolekcji wyjściowej.

Słowa kluczowe: kolekcje podstawowe, kolekcje zasobów genowych, metody pobierania próby

A core collection is a sample of an entire crop germplasm collection, selected to adequately represent, with a minimum of redundancies, the genetic diversity in the entire collection. The purpose of forming plant core collections is generally to reduce the entire collection to a manageable size that facilitates easier systematic and rigorous characterization and evaluation of the genetic diversity in that collection for numerous phenotypic descriptors and for molecular attributes. These activities have the key importance for effective maintaining, managing and sustainable utilization of plant genetic resources for research and crop breeding programs. Many methods have been developed to construct core collections from the entire collections. In the establishing of core collections, the specifying of an appropriate sampling strategy is critical. A sampling strategy is the methodology of selection such a sample that adequately represents the variation or diversity in a population from which has been drawn.

Key words: germplasm collection, core collections, sampling methods

* Mniejsza praca wykonana była w ramach projektu promotorskiego numer N N310 066339, przyznanego przez Ministerstwo Nauki i Szkolnictwa Wyższego.

WSTĘP

Głównymi celami działalności specjalistów pracujących dla kolekcji roślinnych zasobów genowych (zwanymi także bankami genów) są: zabezpieczenie, charakteryzowanie i ocena różnorodności genetycznej i fenotypowej oraz tworzenie łatwego dostępu do tych zasobów w celu ich wszechstronnego badania i efektywnego wykorzystania przez naukowców, hodowców i innych użytkowników obecnie i w przyszłości (van Hintum, 1999, 2000; Upadhyaya i in., 2006a, 2009, 2010; van de Wouw i in., 2010 a,b). Wiele kolekcji roślinnych zasobów genowych zawiera tak liczne obiekty, że utrudnia to realizację podstawowych celów, dla jakich je utworzono. Duża liczba obiektów w kolekcji zasobów genowych przyczynia się do trudności organizacyjnych, wzrostu kosztów utrzymania oraz słabego wykorzystaniem zgromadzonej różnorodności (van Hintum i in. 2000; Franco i in., 2006; Jansen i van Hintum, 2007; Yan i in., 2007; Upadhyaya i in., 2009). W celu rozwiązania tych problemów Frankel (1984), Frankel i Brown (1984) oraz Brown (1989 a,b) zaproponowali koncepcję tworzenia kolekcji podstawowych (ang. core collections). Kolekcja podstawowa dobrze reprezentuje różnorodność genetyczną dla fenotypowych cech ilościowych w zgromadzonej dotychczas (wyjściowej) kolekcji zasobów genowych, czyli jest ona reprezentatywna pod względem różnorodności genetycznej dla kolekcji wyjściowej.

W scenariuszu tworzenia kolekcji podstawowej bardzo ważna jest odpowiednia statystyczna metoda pobierania próby, czyli próbkowania (van Hintum i in., 1999, 2000; Xu i in., 2006; Li i in., 2007). Metoda pobierania próby jest sposobem wyboru podzbioru z populacji wyjściowej, który adekwatnie reprezentuje zmienność genetyczną w tej populacji. Dotychczas opracowano i zastosowano wiele metod pobierania próby z wyjściowej kolekcji roślinnych zasobów genowych, wykorzystywanych do tworzenia kolekcji podstawowych. Obejmują one pobieranie próby całkowicie losowej, metody pobierania warstwowej próby losowej i nielosowej (Brown, 1989a,b; Charmet i Balfourier, 1995; Grenier i in., 2000; Hu i in., 2000; Li i in., 2002; Upadhyaya i in., 2003, 2007, 2009; Xu i in., 2006) oraz wysublimowane inne metody pobierania próby wykorzystywane do tworzenia kolekcji podstawowych (Marita i in., 2000; Raamsdonk i Wijnker, 2000; Chandra i in., 2002; Jansen i van Hintum, 2007; Kim i in., 2007).

Dotychczas przeprowadzone badania efektywności metod pobierania próby do tworzenia kolekcji podstawowej, reprezentatywnej pod względem różnorodności genetycznej dla cech fenotypowych w kolekcji wyjściowej, nie wskazują jednoznacznie, które z metod można uznać za najefektywniejsze (optymalne) w realizacji postawionego celu (Brown i Spillane, 1999; van Hintum i in., 2000; Hu i in., 2000; Franco i in., 2005; Li i in., 2007). Dotyczy to także metod pobierania próby według innych kryteriów różnorodności genetycznej, zwłaszcza markerów DNA. Zatem, w niniejszej pracy podjęto próbę zebrania oraz opracowania literatury światowej i krajowej, przedstawiającej wyniki badań teoretycznych i empirycznych nad tworzeniem, charakterystyką i oceną porównawczą efektywności wielu metod statystycznych pobierania próby, stanowiącej kolekcję podstawową roślinnych zasobów genowych.

Podstawowe zagadnienia kolekcji zasobów genowych

Głównym celem tworzenia kolekcji roślinnych zasobów genowych jest zabezpieczenie różnorodności genetycznej, istniejącej aktualnie w obrębie gatunków roślin uprawnych i innych gatunków z nimi spokrewnionych, a także gatunków ważnych z przyrodniczego punktu widzenia, przed utratą w wyniku zastępowania lokalnych odmian przez nowoczesne odmiany uprawiane w wysoko nakładowych systemach produkcji (Hausmann i in., 2004; van de Wouw i in., 2010 a,b). Utrata różnorodności genetycznej następuje w wyniku występowania zjawiska erozji genetycznej (Ramanatha Rao i Hodgkin, 2002; van de Wouw i in., 2010 a,b). Erozja genetyczna definiowana jest jako postępujące zmniejszenie się liczby gatunków lub odmian roślin uprawnych z obszarów ich dotychczasowej uprawy. Kolekcje roślinnych zasobów genowych odgrywają coraz większą rolę w zapewnieniu żywności oraz przyczyniają się do rozwoju gospodarczego (FAO, 2010). Problemem w zapewnieniu żywności na odpowiednim poziomie ilościowym i jakościowym dla mieszkańców Ziemi, staje się przystosowanie roślin do obecnych oraz przyszłych zmian klimatycznych. Zatem, wykorzystanie roślinnych zasobów genowych w pracach hodowlanych może przyczynić się do efektywnej hodowli odmian przystosowanych do zmieniających się warunków środowiska poprzez znaczące zwiększenie ich tolerancji na stropy biotyczne i abiotyczne (Robertson i in., 1996; Ramanatha Rao i Hodgkin, 2002; van Hintum i in., 1999, 2000).

Konwencja o Różnorodności Biologicznej (Dz. U. nr 184 poz. 1532 r. 1992) ogłoszona i przyjęta podczas międzynarodowej konferencji Środowisko i Rozwój, znanej jako Szczyt Ziemi, która odbyła się w Rio de Janeiro w 1992 roku, została ratyfikowana przez Polskę w 1995 roku. Konwencja ta stanowi podstawę prawa międzynarodowego w zakresie zasobów genowych i różnorodności biologicznej, potwierdza duże znaczenie różnorodności biologicznej, stanowiącej podstawy rolnictwa i wyżywienia na świecie. Nakłada na państwa sygnatariuszy moralny obowiązek zrównoważonego gospodarowania różnorodnością genetyczną ziemi (Bulińska-Radomska, 2008; FAO, 2010). Od 2006 roku w ramach agendy FAO działa instytucja Bioversity International, zajmująca się koordynowaniem i wspieraniem międzynarodowej współpracy w zakresie zasobów genowych roślin uprawnych.

Kolekcje roślinnych zasobów genowych zawierają materiał biologiczny (materiał siewny, rośliny, albo też części roślin i kultury tkankowe *in vitro*) zebrany z dzikich populacji, odmian lokalnych, współczesnych odmian i materiałów hodowlanych, stanowiące obiekty genetyczne w tych kolekcjach (Ramanatha Rao i Hodgkin, 2002; Jansen i van Hintum, 2007). Obiekty zgromadzone w kolekcjach zasobów genowych mogą być chronione za pomocą dwóch metod — ochronę *ex situ* oraz ochronę *in situ* (Ramanatha Rao i Hodgkin, 2002). Ochrona *ex situ* umożliwia zachowanie różnorodności biologicznej roślin uprawnych poza miejscem ich naturalnego występowania. Polega ona na przeniesieniu pojedynczych roślin lub ich grup (obiektów genetycznych) z siedlisk naturalnych do innych, zwykle sztucznych warunków. Odbywa się to za pomocą następujących metod: przechowywanie diaspor generatywnych i wegetatywnych, przechowywania kultur *in vitro*, tworzenie kolekcji polowych i ogrodów botanicznych (Ramanatha Rao i Hodgkin, 2002). To postępowanie prowadzi do tworzenia kolekcji

zasobów genowych (van Hintum i in., 1999, 2000). Ochrona *in situ* oznacza zachowanie naturalnych ekosystemów występowania dziko rosnących roślin użytkowych, miejscowych populacji, odmian lokalnych i dzikich gatunków spokrewnionych z roślinami uprawnymi. Prowadzona jest on poprzez tworzenie rezerwatów przyrody i parków narodowych, zabezpieczenie różnorodności w gospodarstwie rolnym (ang. on-farm conservation) czy też w sadach i ogrodach przydomowych. W 2007 roku obszar objęty ochroną *ex situ* na świecie był równy 17,5 milionom km² (FAO, 2010).

Ochrona *ex situ* zasobów genowych jest najpowszechniejszym sposobem zabezpieczenia bioróżnorodności roślin uprawnych w kolekcjach zasobów genowych. Na całym świecie w takich kolekcjach zgromadzono około 7,4 miliona obiektów. Obecnie na świecie istnieje około 1750 banków genów, z czego 130 zawiera więcej niż 10000 obiektów. Natomiast, ogrodów botanicznych chroniących różnorodność roślin uprawnych jest ponad 2500, zgromadzono w nich około 80000 gatunków roślin (FAO, 2010). Ważnym osiągnięciem ostatnich lat jest wybudowanie Międzynarodowego Skarbcza Nasiennego w Svalbard, Norwegia (ang. The Svalbard Global Seed Vault) przez rząd Norwegii oraz prywatną fundację Global Crop Diversity Trust. Został on wybudowany w okolicach miejscowości Longyearbyen na wyspie Spitsbergen, należącej do Norwegii. Jest to skarbiec zabezpieczający nasiona oraz inny materiał biologiczny obiektów pochodzących z różnych kolekcji roślinnych zasobów genowych, przed gwałtownymi zmianami klimatu, katastrofami naturalnymi, czy konfliktami politycznymi, społecznymi lub zbrojnymi, w szczególności konfliktem atomowym.

Do największych na świecie narodowych kolekcji zasobów genowych roślin uprawnych można zaliczyć kolekcję Stanów Zjednoczonych Ameryki, która liczy ponad 500 000 obiektów. Kolejną kolekcją w rankingu według wielkości jest chińska kolekcja zasobów genowych, w której zgromadzono prawie 400 000 obiektów. Kolekcje zasobów genowych roślin uprawnych gromadzone w Indiach i Federacji Rosyjskiej liczą ponad 300 000 obiektów. Natomiast, do największych kolekcji prowadzonych przez międzynarodowe instytuty badawcze można zaliczyć kolekcję prowadzoną w Międzynarodowym Centrum Doskonalenia Kukurydzy i Pszenicy, CIMMYT (ang. International Wheat and Maize Improvement Center), w której zabezpieczonych jest 174 000 obiektów. Drugą, co do wielkości kolekcja tego typu, gromadzona w Międzynarodowym Instytucie Badawczym Rolnictwa na Terenach Suchych, ICARDA (ang. International Center for Agricultural Research in the Dry Areas) zawiera 133 000 obiektów (Robertson i in., 1996; FAO, 2010).

Polski Narodowy Program Ochrony Zasobów Genowych Roślin Użytkowych, który musieliśmy przygotować i wprowadzić w życie, jako państwo sygnatariusz Konwencji o Różnorodności Biologicznej, jest finansowany przez Ministerstwo Rolnictwa i Rozwoju Wsi. Program ten realizują 3 wyższe uczelnie, 7 branżowych instytutów, 4 stacji hodowli roślin oraz Ogród Botaniczny Centrum Zachowania Różnorodności Biologicznej Polskiej Akademii Nauk w Powsinie (Bulińska-Radomska i in., 2010). Program koordynowany jest przez Krajowe Centrum Roślinnych Zasobów Genowych IHAR oraz Radę do spraw Ochrony Zasobów Genowych, działającą przy tym Centrum. Polska jest członkiem Europejskiego Programu Ochrony Zasobów Genowych ECPGR (ang. European Cooperative Programme for Plant Genetic Resources), który został utworzony w 1980

roku. Należą do niego 43 kraje. Program ma na celu rozwój współpracy pomiędzy europejskimi kolekcjami, wyznacza standardy działania i zarządza ogólnoeuropejską bazą danych o zgromadzonych zasobach genowych roślin uprawnych AEGIS (ang. An European Genebank Integrated System). W programie uczestniczy 38 krajów europejskich. Działa on za pośrednictwem 9 rozbudowanych sieci, w tym 6 dotyczących grup roślin (zbóż, roślin pastewnych, roślin oleistych i wysokobiałkowych, warzyw i przypraw) oraz 3 sieci tematycznych. W ich kręgu zainteresowania znajduje się dokumentacja i sieć informatyczna, oraz współpraca międzynarodowa. Polska jako uczestnik Europejskiego Programu Ochrony Zasobów Genowych jest odpowiedzialna za koordynację międzynarodowych kolekcji i organizację Europejskiej Bazy Danych dla roślin z rodzaju *Secale*, *Lupinus*, *Dactylis*, *Festuca*. W polskich kolekcjach roślinnych zasobów genowych zgromadzonych jest około 75 000 obiektów. Struktura gatunkowa tych zasobów jest następująca: 35,6% obiektów zbóż, 22,3% obiektów traw, 12,2% obiektów warzyw, 9,5% obiektów motylkowych drobnonasiennych (Bulińska-Radomska i in., 2008). W Krajowym Centrum Roślinnych Zasobów Genowych IHAR znajduje się długoterminowa przechowalnia nasion, która może pomieścić 100 000 obiektów.

Jednym z głównych przyczyn występowania problemów i trudności w trakcie prowadzenia, zarządzania, oceny i wykorzystania kolekcji roślinnych zasobów genowych jest duża i ciągle rosnąca liczba obiektów. Duża liczba obiektów utrudnia realizację podstawowych celów, dla jakich powstają takie kolekcje. Duża liczba obiektów w roślinnych kolekcjach zasobów genowych wywołuje problemy w zachowaniu na odpowiednim poziomie żywotności zgromadzonego materiału biologicznego, co przyczynia się także do trudności w namnożeniu odpowiedniej ilości i odpowiedniej jakości materiału siewnego, która może zostać przekazana innym badaczom i hodowcom. Trudnością w gromadzeniu zasobów genowych są spowodowane dużą liczbą obiektów, co powoduje, że około 30% lub nawet więcej nowo gromadzonych obiektów jest duplikatami już istniejących w kolekcjach (FAO, 2010). Identyfikacja takich obiektów jest trudna, zwłaszcza w dużych kolekcjach, więc na wszelki wypadek wszystkie nowo pozyskane obiekty przyłączane są do kolekcji zasobów genowych, bez ich wcześniejszej identyfikacji i oceny. Duża liczba obiektów w kolekcjach zasobów genowych utrudnia wykonanie pełnej i właściwej oceny ich różnorodności genetycznej. Pełna i właściwa ocena różnorodności ogromnej liczby obiektów zarówno pod względem cech fenotypowych i genotypowych, jest często niewykonalna ze względów finansowych oraz organizacyjnych (odpowiedniej liczby ludzi czy wystarczająco wyposażonego zaplecza technicznego). Z ostatniego raportu dotyczącego stanu zasobów genowych na świecie (FAO, 2010) wynika, że średnio 64% obiektów scharakteryzowanych jest pod względem cech morfologicznych i agronomicznych, a tylko 14 procent obiektów jest ocenionych za pomocą markerów molekularnych i izoenzymów. Natomiast, 51% obiektów pochodzących z kolekcji zasobów genowych roślin oleistych zostało ocenionych za pomocą metod biochemicznych (markery DNA i izoenzymy), a wśród roślin ozdobnych czy roślin przyprawowych żaden z obiektów nie został oceniony pod względem różnorodności genetycznej za pomocą tych metod (FAO, 2010).

Stopień wykorzystania roślinnych zasobów genowych w Polsce jest na bardzo niskim poziomie, liczba średnio rocznie dystrybuowanych obiektów z polskich kolekcji roślin uprawnych wynosi około 5700 wysłanych (przekazanych) do odbiorców w kraju jak i zagranicą (Bulińska-Radomska i in., 2008; FAO, 2010). Około 70% tych obiektów przekazano do przeprowadzenia różnego typu prac badawczych, 20% obiektów wysłano hodowcom, a 10% obiektów przekazano innym kolekcjom na świecie w ramach wymiany. W ostatnich latach obserwuje się spadek zainteresowania hodowców z różnych części świata wykorzystaniem zgromadzonej w kolekcjach różnorodności genetycznej, w prowadzonych przez nich pracach hodowlanych (FAO, 2010). Wiele krajów oraz organizacji międzynarodowych tworzy specjalne programy wspierające hodowlę roślin uprawnych, mające na celu promowanie i informowanie o zasobach genowych oraz o możliwości ich wykorzystania.

Idea tworzenia kolekcji podstawowych

Aby sprostać wyzwaniom związanym z dużą i ciągle rosnącą liczbą obiektów w kolekcjach zasobów genowych roślin uprawnych, Frankel (1984), Frankel i Brown (1984) i Brown (1989 a,b) zaproponowali ograniczenie liczebności obiektów w istniejących kolekcjach poprzez tworzenie kolekcji podstawowej (ang. core collections) dla każdej z nich. Aktualnie istniejąca duża (liczna) kolekcja, z której powstaje kolekcja podstawowa nazywana jest kolekcją wyjściową lub całkowitą (ang. initial collection, entire collection, whole collection). Kolekcja wyjściowa może być traktowana jako populacja skończona lub nieskończona (Vencovsky i Crossa, 1999, 2003).

Kolekcja podstawowa jest podzbiorem obiektów o ograniczonej liczebności, wybranych tak z kolekcji wyjściowej, aby odzwierciedlały one (reprezentowały, przenosiły) dobrze jej różnorodność genetyczną, czyli powinna być ona pozbawiona jak największej liczby obiektów podobnych genetycznie, czyli duplikatów, które byłyby nadmierne (redundancyjne) dla zachowania różnorodności kolekcji wyjściowej (Frankel, 1984; Frankel i Brown, 1984; Brown, 1989 b, 1995; van Hintum i in., 2000; Chung i in., 2009). Pozostałe obiekty z kolekcji wyjściowej, nie włączone do kolekcji podstawowej mogą tworzyć kolekcję rezerwową (ang. reserve collection) — (Brown, 1995). Obiekty z takiej kolekcji są przekazywane do długoterminowego przechowywania, które nie wymaga częstego reprodukcji nasion oraz nie generuje dużych kosztów. Najlepiej w tym celu sprawdza się przechowywanie w warunkach kriogenicznych (van Hintum, 2000). Ocena różnorodności oraz charakterystyka obiektów zgromadzonych w kolekcji rezerwowej jest ograniczona do minimum (Skroch i in., 1998).

Kolekcje podstawowe odgrywają ważną rolę w zarządzaniu i wykorzystaniu kolekcji zasobów genowych. Stanowią jedno z podstawowych i powszechnie stosowanych narzędzi kuratora, wspomagających i usprawniających jego codzienną pracę (Brown, 1995; Skroch i in., 1998; van Hintum i in., 2000). Utworzenie kolekcji podstawowych może przyczynić się do identyfikacji braków (luk) w różnorodności biologicznej wśród obiektów zgromadzonych w aktualnie istniejącej kolekcji zasobów genowych roślin uprawnych (Upadhyaya i in., 2010; Gowda i in., 2011). Kolekcje podstawowe w trakcie dołączania nowych obiektów do kolekcji, mogą pomóc kuratorowi w podjęciu decyzji o włączeniu lub nie, takiego obiektu. Jeżeli nowy obiekt jest podobny do obiektów znajdujących się w

kolekcji podstawowej, kurator może nie włączać go do kolekcji zasobów genowych. Natomiast, jeśli nowy obiekt wyraźnie odbiega od tych, które już znajdują się w kolekcji podstawowej, to będzie on wzbogacał różnorodność w istniejącej kolekcji i dlatego warto go włączyć do kolekcji (Brown, 1995; van Hintum i in., 2000).

Materiał biologiczny (nasiona kultury *in vitro*) obiektów zgromadzonych w kolekcji podstawowej posiada wysoki priorytet w trakcie ich przechowywania. Taki materiał powinien być częściej poddawany rutynowej ocenie żywotności, niż obiekty zgromadzone w kolekcji wyjściowej. Kolekcje podstawowe są także odpowiednie do badania nowych rozwiązań przechowywania zgromadzonych zasobów roślin uprawnych, które później można bez obaw zastosować w przechowywanie aktualnie istniejącej kolekcji (Brown, 1995, 1999). Ze względu na mniejszą liczebność obiektów w kolekcjach podstawowych, realna jest wszechstronna ocena zmienności genetycznej, która jest pośrednio także oceną kolekcji wyjściowej (Brown, 1995, 1999; van Hintum i in., 2000). Ocena zmienności genetycznej obiektów w kolekcjach zasobów genowych powinna być wykonywana na podstawie dwóch podstawowych kryteriów, tj. cech fenotypowych oraz genetycznych, przeważnie na poziomie markerów DNA lub markerów biochemicznych (Schoen i Brown, 1993; Chavarriaga-Aguirre i in., 1999; van Hintum, 1999; Grenier i in., 2000a,b; van Hintum i in., 2000; Balfourier in., 2007; Franco i in., 2005, 2006, 2010; Wang i in., 2006; van de Wouw i in., 2010b; Pagnotta i in., 2011).

Ocena zmienności genetycznej pod względem cech fenotypowych (nazywana czasami zmiennością fenotypową — Jahufer i in., 1997; Casler i van Santen, 2000; Upadhyaya i in., 2003, 2007 a; Ntundu i in., 2006; Badea i in., 2008) w kolekcjach roślinnych zasobów genowych powinna obejmować ważne ilościowe rolnicze oraz biologiczne cechy rozpatrywanego gatunku. Taka ocena obiektów zgromadzonych w kolekcji jest przeprowadzana zazwyczaj na podstawie danych z polowych obserwacji cech w czasie wegetacji (Hartung, 2006). Jednakże, aby ocena zmienności genotypowej dla ilościowych cech agronomicznych była wiarygodna na potrzeby rolnictwa, powinna być przeprowadzona w wielu środowiskach oraz powtórzona w latach (Franco i in., 1999, 2003; Upadhyaya i in., 2007 a,b; Gowda i in., 2011). Jest to możliwe tylko dla mniejszej liczby obiektów zwłaszcza tych należących do kolekcji podstawowej.

Zmienność genetyczna na poziomie markerów DNA jest oceniana za pomocą różnych, coraz to nowszych, technik inżynierii genetycznej (Chavarriaga-Aguirre i in., 1999; Kölliker i in. 1999; Ghamkhar i in., 2005; Hao i in., 2006; Franco i in., 2006; Balfourier i in., 2007; Badea i in., 2008; Peng i in., 2008; Zeng i in., 2008; van de Wouw i in., 2010 b; Xie i in., 2010; Olukolu i in., 2011). Stosowane są także zintegrowane podejścia metodyczne do oceny różnorodności zasobów genowych, które uwzględniają łącznie fenotypowe deskrytory jakościowych i ilościowych cech morfologicznych oraz agronomicznych, a także markery DNA dobrze reprezentujące genom danego gatunku roślin (Franco i in., 2001; Wang in., 2006; Badea i in., 2008; Olukolu i in., 2011). Są one najbardziej skuteczne przy wnioskowaniu o przydatności zasobów genowych do programów hodowlanych (Franco i in., 2010; van de Wouw i in., 2010b; Pagnotta i in., 2011). Taka wielostronna ocena różnorodności i charakterystyka jest w pełni wykonalna zarówno pod względem finansowym jak i organizacyjnym, tylko na ograniczonej liczbie

obiektów w kolekcji podstawowej. Obiekty zgromadzone w kolekcjach podstawowych powinny być najpierw oceniane za pomocą nowoczesnych technik inżynierii genetycznej, które są często bardzo kosztowne i czasochłonne, a więc trudne do wykonania dla dużej kolekcji zasobów genowych rozpatrywanego gatunku.

Ocenę różnorodności genetycznej pod względem markerów DNA lub biochemicznych wykonano między innymi w kolekcji podstawowej sorga cukrowego (*Sorghum bicolor* (L.) Moench) — (Grenier i in., 2000b), koniczyny łąkowej (*Trifolium pratense* L.) — (Mosjidis i Klingler 2006), pszenicy zwyczajnej (*Triticum aestivum* L.) — (Hao i in., 2006; Balfourier i in., 2007) oraz manioku jadalnego (*Manihot esculenta* Crantz) — (Chavarriaga-Aguirre i in., 1999). Natomiast ocenę różnorodności genetycznej pod względem cech fenotypowych wykonano w wielu kolekcjach podstawowych, między innymi w kolekcji podstawowej koniczyny białej (*Trifolium repens* L.) — (Jahufer i in., 1997), sorga cukrowego (*Sorghum bicolor* (L.) Moench) — (Grenier i in., 2000 a, Upadhyaya i in., 2010), orzecha ziemnego (*Arachis hypogaea* L.) — (Upadhyaya i in. 2003), kapusty warzywnej (*Brassica oleracea* L.) — (Santos i Dias, 2004), życicy trwałej (*Lolium perenne* L.) — (Schmidt 2005a,b), nikli indyjskiej (*Cajanus cajan* (L.) Millsp.) — (Upadhyaya i in., 2007 a), ciecierzycy pospolitej (*Cicer arietinum* L.) — (Upadhyaya i in. 2007 b) oraz melona (*Cucumis melo* L.) — (Luan i in., 2008).

Wyniki oceny różnorodności genotypowej na poziomie fenotypowym i markerów genowych w kolekcjach podstawowych powinny być łatwo dostępne dla hodowców i innych zainteresowanych badaczy. W tym celu tworzy się internetowe bazy danych, dostępne dla zainteresowanych specjalistów. Znakomitym przykładem są bazy danych ICRISAT (ang. The International Crops Research Institute for the Semi-Arid-Tropics) dla utworzonych w tym instytucie kolekcji podstawowych oraz mini kolekcji podstawowych. Są one ogólnodostępne pod adresem <http://www.icrisat.org/gene-bank-activities.htm> (weryfikacja 25.04.2012). W tych bazach danych, oprócz podstawowych informacji o obiektach w kolekcjach podstawowych (dane paszportowe, pochodzenie geograficzne, kwalifikacja systematyczna, itp.), prezentowane są także wyniki oceny różnorodności cech fenotypowych, wykonanej na podstawie polowych doświadczeń lub badań obserwacyjnych oraz wyniki oceny różnorodności genotypowej za pomocą markerów molekularnych DNA (Upadhyaya i Ortiz, 2001; Upadhyaya i in., 2002; Upadhyaya i in., 2007a; Upadhyaya i in., 2007b; Gowda i in., 2011). Podobne bazy danych utworzono dla wielkiej kolekcji zasobów genowych roślin strączkowych w ICARDA (Robertson i in. 1996).

Utworzenie kolekcji podstawowej może znacząco przyczynić się do zwiększenia stopnia wykorzystania zasobów genowych (Frankel i Brown, 1984; Robertson i in., 1996; van Hintum, 2000; Upadhyaya i Ortiz, 2001; Gowda i in., 2011). Ograniczona liczba obiektów, jaka znajduje się w kolekcji podstawowej, może być efektywnie przebadana pod względem ich adaptacji do docelowych warunków środowiskowych (Brown 1995; van Hintum, 2000; Upadhyaya i Ortiz, 2001; Upadhyaya i in., 2007 a,b; Gowda i in., 2011). Hodowca może włączyć do swojego programu większość obiektów z kolekcji podstawowej, co nie było by możliwe z obiektami znajdującymi się w kolekcji wyjściowej. Hodowcy oraz inni badacze mają znacznie większe ułatwienie w poszukiwaniu informacji

o obiektach w kolekcji podstawowej, niż w dużej kolekcji wyjściowej, a także łatwiej jest wybrać do swoich potrzeb hodowlanych obiekty wszechstronnie scharakteryzowane i ocenione (Robertson i in. 1996; van de Wouw i in. 2010 a; Upadhyaya i in., 2007 b, 2009, 2011 c).

Światowy plan działania na rzecz zachowania i zrównoważonego wykorzystania roślinnych zasobów genowych (FAO, 1996) zaleca tworzenie kolekcji podstawowych jako jednego z głównych działań, mających na celu bardziej efektywne wykorzystanie zgromadzonych i stale poszerzanych banków genów. Kolekcje podstawowe zostały zaakceptowane jako efektywne narzędzie zabezpieczenia oraz wykorzystania zgromadzonej różnorodności biologicznej. Plan postuluje, aby w pierwszej kolejności tworzyć kolekcje podstawowe dla gatunków roślin uprawnych o dużym znaczeniu gospodarczym i przyrodniczym dla regionu, kraju czy świata. Informacje o istniejących i nowo powstałych kolekcjach podstawowych powinny być intensywnie rozpowszechniane wśród kuratorów, hodowców oraz innych potencjalnych użytkowników. Dla realizacji planów ochrony zasobów genowych w poszczególnych krajach lub w obrębie gatunków roślin uprawnych, powinny zostać zapewnione z różnych źródeł międzynarodowych, państwowych i prywatnych, odpowiednie środki finansowe oraz wsparcie techniczne.

W celu upowszechnienia wiedzy i informacji o kolekcjach podstawowych wśród naukowców i kuratorów kolekcji zasobów genowych roślin uprawnych, Międzynarodowy Instytut Roślinnych Zasobów Genowych (ang. International Plant Genetic Resources Institute), a dzisiejszy Bioversity International (agenda FAO) opracował odpowiedni biuletyn techniczny (van Hintum i in., 2000). Opracowanie to przedstawia ówczesną wiedzę i doświadczenia z zakresu tworzenia, zarządzania oraz wykorzystania kolekcji podstawowych w obrębie roślinnych banków genów. Biuletyn prezentuje powszechnie stosowane procedury, metody i techniki stosowane do tworzenia kolekcji podstawowych, stanowiąc pewnego rodzaju wytyczne i instrukcje dla osób odpowiedzialnych za zarządzanie istniejącymi kolekcjami.

Kolekcje podstawowe będą zawsze znacznie mniej liczne niż kolekcje wyjściowe, z których zostały one wyłonione. Wielkość (liczebność) kolekcji podstawowej jest określana przez frakcję próby, czyli procent obiektów, które z kolekcji wyjściowej zostaną włączone do kolekcji podstawowej. Nie ma jednoznacznych norm dla wartości frakcji próby, która byłaby uznana za optymalną dla większości kolekcji roślinnych zasobów genowych. Brown (1989 b) zaproponował, aby kolekcje podstawowe nie przekraczały 10% kolekcji wyjściowej, a ich liczebność nie przekraczała 2000 obiektów. Było by odwrotnym do zamierzonego celu, gdyby kolekcja podstawowa przekraczała te wartości dla frakcji oraz liczebności obiektów — zbyt duża kolekcja podstawowa mogłaby być narażona na te same problemy i trudności, co kolekcja wyjściowa (Brown, 1995; Brown i Spillane, 1999). Frakcja większości utworzonych kolekcji podstawowych mieści się w zakresie od 5 do 20% obiektów z kolekcji wyjściowej (Brown, 1989 a; Brown i Spillane, 1999; van Hintum, 1999; van Hintum i in., 2000; Franco i in., 2005, 2006). Z kolekcji zasobów genowych liczących setki obiektów utworzono kolekcje podstawowe, nieprzekraczające 25% frakcji próby, natomiast z kolekcji wyjściowych, liczących tysiące

obiektów powstały kolekcje podstawowe nie przekraczające 5% frakcji próby (Brown, 1995; Brown i Spillane, 1999).

Upadhyaya i Ortiz (2001) zaproponowali tworzenie jeszcze bardziej zredukowanych kolekcji podstawowych z już utworzonych kolekcji podstawowych, zwanych mini kolekcjami podstawowych (ang. mini core collections). Kolekcje tego typu radykalnie ograniczają liczbę obiektów, ale jednocześnie zachowują większość różnorodności genetycznej kolekcji wyjściowej danego gatunku. Dotychczas powstałe mini kolekcje podstawowe zawierają najczęściej około 1% obiektów w kolekcji wyjściowej. Powstały one między innymi dla zasobów genowych ciecierzycy pospolitej (*Cicer arietinum* L.) — (Upadhyaya i Ortiz, 2001), sorga cukrowego (*Sorghum bicolor* (L.) Moench) — (Upadhyaya i in. 2009), rosplenicy perłowej (*Pennisetum glaucum* (L.) R. Br.) — (Upadhyaya i in. 2011c) oraz orzecha ziemnego (*Arachis hypogaea* L.) — (Upadhyaya i in. 2002; Holbrook i Dong, 2005).

Dotychczas kolekcje podstawowe utworzono w obrębie krajowych lub międzynarodowych kolekcji wyjściowych dla wielu ważnych gospodarczo roślin uprawnych o znaczeniu regionalnym lub światowym (van Hintum, 2000; FAO, 2010). W drugim raporcie FAO o zasobach genowych na świecie (FAO 2010) podkreślono znaczenie kolekcji podstawowych dla poprawy efektywności i skuteczności oceny różnorodności biologicznej oraz dla efektywniejszego wykorzystania zgromadzonych zasobów. Wiele krajów oraz instytucji międzynarodowych utworzyło kolekcje podstawowe w obrębie swoich zasobów genowych lub planują one utworzenie kolekcji podstawowych w najbliższym czasie. W wymienionym raporcie FAO zachęca się także do tworzenia kolekcji podstawowej z kolekcji wyjściowej, stanowiącej zbiór wszystkich zasobów genowych danego gatunku roślin uprawnych, zgromadzonych na świecie (FAO, 2010).

Jak dotychczas, w obrębie roślinnych kolekcji zasobów genowych prowadzonych w Polsce, powstała jedna kolekcja podstawowa dla życicy trwałej (*Lolium perenne* L.) z kolekcji zasobów genowych tego gatunku, zgromadzonej w Ogrodzie Botanicznym Instytutu Hodowli i Aklimatyzacji Roślin w Bydgoszczy (Schmidt, 2005 a, 2005 b). Prowadzone są także prace nad utworzeniem kolekcji podstawowej dla pszenżyta jarego (*×Triticosecale* Wittm.) w Instytucie Genetyki, Hodowli i Biotechnologii Roślin Uniwersytetu Przyrodniczego w Lublinie (Kociuba i in., 2010; Studnicki i in., 2010 a, 2010 b). Trwają także zaawansowane prace przygotowawcze do tworzenia kolekcji podstawowych innych gatunków roślin uprawnych, które są koordynowane przez Krajowe Centrum Roślinnych Zasobów Genowych IHAR (Bulińska-Radomska i in., 2008 oraz dane niepublikowane uzyskane od autorki).

Badacze przystępujący do tworzenia kolekcji podstawowej muszą podjąć trzy fundamentalne decyzje. Pierwszą z nich jest określenie wielkości kolekcji podstawowej, jako frakcji próby z kolekcji wyjściowej. Druga decyzja dotyczy wyboru kryterium (obserwowalnych zmiennych genetycznych lub fenotypowych) do oceny genetycznej różnorodności obiektów w kolekcji wyjściowej. Trzecia decyzja dotyczy wyboru (lub opracowania oryginalnej) statystycznej metody pobierania próby z kolekcji wyjściowej (van Hintum i in., 2000; Ghamkhar i in. 2008; van de Wouw i in., 2010 b).

Metody pobierania próby do tworzenia kolekcji podstawowych

Większość dotychczas opracowanych metod wyboru kolekcji podstawowych korzysta z teorii metody reprezentacyjnej (ang. sampling method) — (Neyman 1934; Cochran 1977; Thompson, 2002). Opracowano wiele metod (technik) pobierania próby, służących do tworzenia kolekcji podstawowych z już istniejących kolekcji roślinnych zasobów genowych. Metody pobierania próby, wykorzystywane do tworzenia kolekcji podstawowej, można podzielić na dwie podstawowe grupy, tj. (a) metoda całkowicie losowego wyboru próby (wybór próby losowej, czyli próby prostej) — (ang. simple random sampling) oraz (b) metody wyboru próby warstwowej (ang. stratified sampling methods) (Brown, 1989 b; Spagnoletti-Zeuli i Qualset 1993; Hu i in. 2000; van Hintum i in., 2000; Franco i in., 2005, 2006).

Metoda całkowicie losowego wyboru próby jest najprostszym sposobem tworzenia kolekcji podstawowej. Stosując tablice liczb losowych lub generatory liczb losowych (ang. random number generator) losujemy bez zwracania numery przypisane obiektom (ang. accession numbers) w kolekcji wyjściowej, które zostaną włączone do kolekcji podstawowej (van Hintum, 1999; van Hintum i in., 2000). Metoda opisana powyżej nie wymaga przeprowadzenia wcześniej oceny zmienności genetycznej obiektów, zgromadzonych w kolekcji wyjściowej. Obiekty w kolekcji wyjściowej mają równe szanse znalezienia się w kolekcji podstawowej, tzn. obiekty wybierane są do kolekcji podstawowej z takim samym prawdopodobieństwem, niezależnie od różnorodności obiektów. Metody te nie wymagają od kuratora specjalistycznej wiedzy merytorycznej o strukturze różnorodności genetycznej oraz geograficznej w kolekcji wyjściowej oraz użycia skomplikowanych narzędzi statystycznych i informatycznych (van Hintum, 2000).

Metody wyboru próby warstwowej, stosowane przy tworzeniu kolekcji podstawowej, polegają w pierwszej kolejności na podziale obiektów w kolekcji wyjściowej na grupy jednorodne (warstwy) pod względem określonych kryteriów. Jako kryteria podziału obiektów na warstwy stosowane są dane ekologiczno-geograficzne dla obiektów (Malosetti i Abadie, 2001; Chandra i in., 2002; Logozzo i in., 2007; Balfourier i in., 2007; Ghamkhar i in., 2008), odległości genetyczne między obiektami na podstawie ich wartości genotypowych dla fenotypowych cech ilościowych (van Raamsdonk i Wijnker, 2000; Crossa i Franco, 2004; Franco i in., 2005; Mahalakshmi i in., 2007; Khan i in., 2009) lub skategoryzowanych (Upadhyaya, 2003; Dwivedi i in., 2005; Upadhyaya i in., 2006 b), albo też odległości genetyczne między obiektami na podstawie ich izoenzymów lub markerów molekularnych DNA, wybranych dla genomu rozpatrywanego gatunku (McKhann i in., 2004; Ronfort i in., 2006; Franco i in., 2006; Balfourier i in., 2007). Następnie, z każdej warstwy wybierana jest losowo lub metodycznie (quasi-losowo) określona liczba obiektów do kolekcji podstawowej (van Hintum i in., 2000; Franco i in., 2005, 2006; Ghamkhar i in., 2008).

POBIERANIE PRÓBY WARSTWOWEJ DO TWORZENIA KOLEKCJI PODSTAWOWEJ

Większość ze stosowanych metod pobierania próby do tworzenia kolekcji podstawowych wykorzystuje koncepcję próby warstwowej oraz jej modyfikacje (Harch i in.,

1995; van Hintum i in., 2000; Franco i in., 2005, 2006; Ghamkhar i in., 2008). Podstawy teorii próby warstwowej stworzył profesor Jerzy Sława-Neyman w latach trzydziestych XX wieku (Neyman 1934; Franco i in. 2005, 2006). Metody pobierania próby warstwowej są kombinacjami określonych metod w obrębie każdej z trzech niżej zaprezentowanych rodzajów procedur (technik) statystycznych (van Hintum, 1999; van Hintum i in., 2000; Li i in., 2004; Liu i in., 2009):

- warstwowanie (grupowanie) obiektów w kolekcji wyjściowej,
- alokacja obiektów, czyli określenie liczb wybieranych obiektów z każdej warstwy,
- wybór obiektów z poszczególnych warstw.

Warstwowanie obiektów w kolekcji wyjściowej

W trakcie pobierania próby warstwowej do tworzenia kolekcji podstawowej w pierwszej kolejności przeprowadzane jest warstwowanie obiektów w kolekcji wyjściowej. Warstwowanie obiektów można przeprowadzić, zarówno za pomocą metod statystycznych, jak i bez ich udziału. Niektóre kolekcje podstawowe zostały utworzone metodami próby warstwowej, w których warstwowanie obiektów w kolekcji wyjściowej przeprowadzono na podstawie odpowiednich, merytorycznych kryteriów *a priori*, z wykluczeniem zastosowania metod statystycznych. W celu utworzenia kolekcji podstawowej fasoli zwyczajnej (*Phaseolus vulgaris* L.) z kolekcji tego gatunku, prowadzonej na Półwyspie Iberyjskim, obiekty pogrupowano pod względem typów nasion (Rodino i in., 2003). W trakcie tworzenia kolekcji podstawowej sorga cukrowego (*Sorghum bicolor* (L.) Moench) w celu podziału kolekcji wyjściowej na warstwy wykonano klasyfikację obiektów według kryteriów taksonomicznych (Grenier i in., 2000 a). Często spotykanym sposobem warstwowania obiektów w kolekcji wyjściowej jest ich podział względem kraju pochodzenia, czy też według regionu pochodzenia obiektów (Spagnoletti Zeuli i Qualset, 1993; Malosetti i Abadie, 2001; Chandra i in., 2002; Upadhyaya i in., 2009; Logozzo i in., 2007). W pracy Li i in. (2002) zaproponowano kilka sposobów grupowania obiektów bez użycia metod statystycznych przy tworzeniu kolekcji podstawowej lokalnych odmian ryżu z prowincji Yunnan. Wykonano warstwowanie obiektów na podstawie ich barwy ziaren oraz pochodzenia geograficznego.

Do podziału obiektów z kolekcji wyjściowej za pomocą metod statystycznych, stosuje się analizę skupień (ang. *cluster analysis*), wykonywaną za pomocą różnych algorytmów aglomeracji (Crossa i in., 1995; van Hintum i in., 2000; Crossa i Franco, 2004). Przed przystąpieniem do analizy skupień należy określić odległości genetyczne (ang. *genetic distance*) pomiędzy obiektami, które stanowią miarę ich wzajemnego niepodobieństwa genetycznego (Krzanowski, 1988; Johnson i Wichern, 2002; Mohammadi i Prasanna, 2003; Reif i in., 2005). Analiza skupień pozwala na grupowanie obiektów podobnych genetycznie do siebie w grupy (warstwy) w taki sposób, aby odległości genetyczne między obiektami w różnych warstwach były możliwie jak największe, a odległości między obiektami należącymi do tej samej grupy były minimalne (Krzanowski, 1988; Johnson i Wichern, 2002; Mohammadi i Prasanna, 2003; Crossa i Franco, 2004; Franco i in., 2005, 2006). Analizę skupień obiektów w kolekcji wyjściowej można przeprowadzić na podstawie odległości genetycznych dla wartości cech fenotypowych (Jahufer i in., 1997; Franco i in., 1998; Casler i van Santen, 2000; Upadhyaya i in., 2003, 2007 a; Ntundu i in.,

2006), markerów DNA, izoenzymów (Ghamkhar i in., 2005; Franco i in. 2006; Balfourier i in., 2007; Xie i in., 2010) lub łącznych danych z obu wymienionych odległości (Badea i in., 2008; Franco i in., 2010; van de Wouw i in., 2010 b; Olukolu i in., 2011; Pagnotta i in., 2011). W analizie skupień obiektów w kolekcjach zasobów genowych stosuje się różne rodzaje odległości (Reif i in., 2005). Dla danych z obserwacji markerów molekularnych DNA stosowane są przeważnie odległości Nei'a lub Jaccarda (Mohammadi i Prasanna, 2003; Reif i in., 2005). Franco i in. (2006) do klasyfikacji obiektów w kolekcji wyjściowej na podstawie markerów molekularnych DNA, zastosowali dwie inne odległości, tj. zmodyfikowaną odległość Rogera (Mosjidis i Klingler, 2006) oraz odległość Cavalli-Sforza. Natomiast, w analizie skupień obiektów w kolekcjach zasobów genowych, kiedy różnorodność genetyczną ocenia się za pomocą eksperymentalnych danych dla cech fenotypowych, stosowana jest przede wszystkim odległość Euklidesowa lub kwadrat tej odległości (Diwan i in., 1995; van Hintum i in., 2000; Crossa i Franco, 2004; Mahalakshmi i in., 2007; Khan i in., 2009), albo też odległość Mahalonobisa (Hu i in., 2000; Li i in. 2004; Xu i in., 2006). Natomiast, przy klasyfikacji kolekcji, scharakteryzowanych genetycznie za pomocą fenotypowych cech skategoryzowanych (jakościowych) i ilościowych, zaleca się zwykle stosowanie odległości Gowera (Gower, 1971; Crossa i Franco 2004; Franco i in. 2005, 2006, 2010).

Metody pobierania próby warstwowej przy tworzeniu kolekcji podstawowej uwzględniają najczęściej dwie metody analizy skupień, tj. metodę UPGMA (ang. Unweighted Pair Group Method with Arithmetic Mean) i metodę Warda (van Hintum i in., 2000; Mohammadi i Prasanna 2003; Franco i in., 2006).

Metoda UPGMA, zwana po polsku metodą średnich odległości między obiektami w skupieniach, jest hierarchiczną techniką analizy skupień. Jest ona jedną z najprostszych metod analizy skupień, powszechnie stosowaną w trakcie tworzenia drzewek filogenetycznych i jest częścią wielu narzędzi taksonomicznych, genetycznych i bioinformatycznych. Metoda ta określa odległość między dwiema grupami, jako średnią odległość pomiędzy wszystkimi parami obiektów należących do rozpatrywanych grup (Johnson i Wichern, 2002). Chavarriaga-Aguirre (1999), Mosjidis i in. (2006), Hao i in. (2006), Luan i in. (2008), Escribano i in. (2008); Ghamkhar i in. (2008) oraz Khan i in. (2009) zastosowali w trakcie tworzenia kolekcji podstawowych metodę UPGMA do warstwowania obiektów w kolekcjach wyjściowych.

W metodzie Warda (1963) w czasie aglomeracji wykorzystywane jest podejście analizy wariancji do oszacowania odległości między grupami (skupieniami). Jest ono oparte na minimalizacji wariancji wewnątrzgrupowej, liczonej przez sumę kwadratów odchyłeń od średnich w grupach (ang. error sum of squares — ESS). W każdym kroku aglomeracji obiektów za pomocą metody Warda następuje połączenie (fuzja) takich dwóch grup, które tworzą nową grupę, posiadają najniższą wartość sumy kwadratów odchyłeń spośród wszystkich możliwych do utworzenia grup. To znaczy, że metoda Warda polega na takiej fuzji skupień na każdym kroku aglomeracji, która zapewnia najmniejszy przyrost wariancji wewnątrzgrupowej. Ta metoda odznacza się własnością podziału obiektów na dużą liczbę grup o małej liczebności, a powstałe grupy są jednorodne i mają zbliżoną liczebność (Crossa i Franco, 2004). Upadhyaya i Ortiz (2001), Upadhyaya i in. (2002), Dwivedi i in.

(2005), Holbrook i Dong (2005), Amalraj i in. (2006), Mahalakshmi i in. (2007), Upadhyaya i in. (2007a,b, 2009) zastosowali metodę Warda do grupowania obiektów w kolekcji wyjściowej zasobów genowych roślin uprawnych przy pobieraniu próby stanowiącej kolekcję podstawową. Franco i in. (1998, 1999, 2005, 2010) opracowali zmodyfikowaną metodę Warda (ang. Ward-Modified Location Method, Ward-MLM) i zastosowali ją do warstwowania kolekcji wyjściowej. Metoda ta umożliwia wykorzystanie obserwacji fenotypowych zarówno dla cech ilościowych, jak i jakościowych, albo wspólne obserwacje cech fenotypowych, markerów molekularnych i (lub) izoenzymów (Franco i in., 1998; Crossa i Franco, 2004). Do warstwowania obiektów w trakcie tworzenia kolekcji podstawowej, stosowano także metodę niehierarchicznej analizy skupień, zwaną metodą k-średnich (ang. k-means method) — (Grenier i in. 2001, Johnson i Wichern, 2002; Zewie i in., 2004; Amalraj i in., 2006). Charmet i Balfourier (1995) zastosowali geostatystyczną metodę (geographic contiguity constraint) podziału obiektów na warstwy w kolekcji wyjściowej życicy trwałej (*Lolium perenne* L.).

Alokacja obiektów

Po określeniu warstw (grup) wyznaczonych za pomocą analizy skupień w kolekcji wyjściowej, drugim etapem pobierania próby warstwowej jest określenie liczby obiektów z poszczególnych warstw, przekazywanych do kolekcji podstawowej. Liczby te określa się za pomocą wybranej metody alokacji obiektów (ang. allocation methods). Metody te stanowią drugi rodzaj technik (procedur) statystycznych, składających się na metody pobierania próby warstwowej. Opracowano wiele metod alokacji obiektów, które można podzielić na trzy następujące grupy (Cochran, 1977; Brown, 1989 a,b; van Hintum i in., 2000; Franco i in., 2005):

- metody alokacji oparte na liczebności obiektów w warstwach,
- metody alokacji oparte na ocenie różnorodności genetycznej obiektów w warstwach,
- metody alokacji oparte na subiektywnej ocenie i decyzji kuratora kolekcji.

Metody alokacji oparte na liczebności obiektów w warstwach

W tych metodach liczba obiektów, które przechodzą do kolekcji podstawowej z poszczególnych grup (warstw) kolekcji wyjściowej jest uzależniona od liczebności tych warstw. Brown (1989 a,b) zaproponował 3 metody alokacji obiektów, oparte na liczebności warstw, tj. alokacja stałej liczby obiektów z warstw, alokacja liczby obiektów z warstw proporcjonalnej do ich liczebności w warstwach (alokacja proporcjonalna), alokacja proporcjonalno-logarytmiczna. Są to jedne z najczęściej stosowanych metod alokacji obiektów do tworzenia kolekcji podstawowych z wykorzystaniem metod pobierania próby warstwowej (van Hintum, 2000; Franco i in., 2005; van de Wouw i in., 2010 b).

Alokacja stałej liczby obiektów z warstw w prosty sposób określa liczbę obiektów, otrzymywaną w postaci ilorazu liczby obiektów, które muszą znaleźć się w kolekcji podstawowej (wynikające z frakcji próby) oraz liczby wyróżnionych warstw. Stała liczba obiektów z warstw nie powinna przekraczać liczebności najmniej licznej grupy. Można ją wyznaczyć za pomocą następującego wzoru:

$$n^{\text{Const}} = \frac{n}{g}$$

gdzie: n określa liczebność obiektów, które zostaną przekazane do kolekcji podstawowej, g jest liczbą warstw wyznaczonych za pomocą analizy skupień lub innych kryteriów.

Liczba obiektów wybieranych z warstwy w metodzie alokacji proporcjonalnej jest wyznaczana za pomocą następującej formuły:

$$n_t^{\text{Pro}} = n \times \frac{N_t}{\sum_{i=1}^g N_i}$$

gdzie: N_t określa liczebność obiektów sklasyfikowanych w t -tej warstwie.

W alokacji proporcjonalno-logarytmicznej liczba obiektów przechodzących do kolekcji podstawowej z warstwy jest proporcjonalna do logarytmu z liczebności danej warstwy i wyraża się następująco:

$$n_t^{\text{Log}} = n \times \frac{\log(N_t)}{\sum_{i=1}^g \log(N_i)}$$

Chandra i in. (2002) zaproponowali metodę pierwiastkowej alokacji obiektów (ang. *square root allocation*). Liczba obiektów z t -tej warstwy, które zostają przekazane do kolekcji podstawowej jest proporcjonalna do pierwiastka z liczebności danej warstwy i jest wyrażona za pomocą następującej formuły:

$$n_t^{\text{SR}} = n \times \frac{\sqrt{N_t}}{\sum_{i=1}^g \sqrt{N_i}}$$

Każda z zaprezentowanych wyżej metod alokacji obiektów daje odmienne wyniki przy tworzeniu kolekcji podstawowych. Metoda alokacji proporcjonalno-logarytmicznej w porównaniu do alokacji proporcjonalnej preferuje warstwy o mniejszej liczebności. Stosując metodę proporcjonalno-logarytmiczną z warstw o mniejszej liczebności do kolekcji podstawowej przekazanych zostanie więcej obiektów, niż przy użyciu metody proporcjonalnej. Natomiast z warstw bardzo licznych w trakcie pobierania próby z zastosowaniem metody proporcjonalno-logarytmicznej przejdzie do kolekcji podstawowej stosunkowo mniej obiektów, w porównaniu do zastosowania alokacji proporcjonalnej (Brown, 1989 b; van Hintum, 2000).

Metody alokacji oparte na ocenie różnorodności genetycznej obiektów w warstwach

Metody te opierają się na ocenie różnorodności genetycznej na podstawie obserwacji zarówno markerów DNA, jak cech fenotypowych pomiędzy obiektami w poszczególnych warstwach. Z warstw charakteryzujących się relatywnie dużą różnorodnością obiektów, metody te pobierają do kolekcji podstawowej więcej obiektów, niż z warstw charakteryzujących się mniejszą różnorodnością. Dotychczas opracowano kilka metod

alokacji opartych na ocenie różnorodności genetycznej obiektów w poszczególnych warstwach.

Typowymi metodami alokacji obiektów z kolekcji wyjściowych, ocenianych za pomocą markerów DNA oraz izoenzymów jest strategia H (van Hintum, 2000). Strategia H alokacji obiektów wykorzystuje wskaźnik różnorodności genetycznej h , który można obliczyć dzięki dostarczeniu szacunkowych informacji o częstości alleli w poszczególnych *loci*. Liczba obiektów, które przechodzą do kolekcji podstawowej z t -tej warstwy jest proporcjonalna do wskaźnika różnorodności genetycznej obiektów w tej warstwie, h_t , czyli:

$$n_t^H = n \times \frac{h_t}{\sum_{t=1}^g h_t}.$$

Dla kolekcji wyjściowej ocenionej za pomocą markerów DNA lub izoenzymów wskaźnikiem różnorodności h jest najczęściej indeks Nei (ang. Nei's gene diversity index) — (Nei, 1973; Reif i in., 2005). Strategię H alokacji obiektów można również zastosować dla kolekcji wyjściowych ocenionych za pomocą fenotypowych cech ilościowych. Mahajan i in. (1999) wskaźnik różnorodności genetycznej h został wyznaczony za pomocą indeksu różnorodności H' Shannona-Weavera (Krebs, 1989).

Franco i in. (2005) zaproponowali 3 metody alokacji obiektów (D_1 , D_2 oraz D_3), w których liczba obiektów przechodzących do kolekcji podstawowej jest zależna od średniej odległości Gowera (Gower, 1971) pomiędzy obiektami w poszczególnych warstwach. Można stosować także inne odległości na podstawie danych pochodzących zarówno z obserwacji cech fenotypowych (ilościowych i jakościowych), jak i z obserwacji przeprowadzonych za pomocą narzędzi inżynierii genetycznej (różne typy markerów DNA, izoenzymów) (Reif i in., 2005). W metodzie D_1 (Franco i in., 2005; Franco i in., 2006) liczba obiektów przekazanych do kolekcji podstawowej z danej warstwy jest proporcjonalna do średniej odległości pomiędzy obiektami w warstwie, czyli:

$$n_t^{D_1} = n \times \frac{d_t}{\sum_{t=1}^g d_t}$$

gdzie: d_t jest średnią wartością odległości pomiędzy obiektami w t -tej warstwie.

W metodzie D_2 liczba obiektów przekazanych do kolekcji podstawowej z warstwy jest proporcjonalna do iloczynu liczebności grupy i średniej odległości pomiędzy obiektami w warstwie, czyli:

$$n_t^{D_2} = n \times \frac{N_t \times d_t}{\sum_{t=1}^g N_t \times d_t}$$

Natomiast, w metodzie D_3 liczba obiektów przekazanych do kolekcji podstawowej z warstwy jest proporcjonalna do iloczynu logarytmu liczebności warstwy i średniej odległości pomiędzy obiektami w warstwie, czyli:

$$n_t^{D_3} = n \times \frac{\log(N_t) \times d_t}{\sum_{t=1}^g \log(N_t) \times d_t}$$

W pracy Franco i in. (2005) zastosowano odległość Gowera dla cech fenotypowych ilościowych i jakościowych przy warstwowaniu wyjściowej kolekcji zasobów genowych kukurydzy zwyczajnej (*Zea mays* L.). Natomiast, w pracy Franco i in. (2006) zastosowano różne odległości genetyczne oszacowane na podstawie markerów molekularnych DNA do alokacji obiektów za pomocą metod D_1 , D_2 i D_3 przy tworzeniu kolekcji podstawowych w obrębie zasobów genowych kukurydzy zwyczajnej.

Metody alokacji obiektów oparte na subiektywnej ocenie i decyzji kuratora kolekcji

Subiektywne wyznaczenie liczby obiektów, które przechodzą do kolekcji podstawowej z każdej warstwy odbywa się na podstawie wiedzy zdobytej przez kuratora w trakcie prowadzenia i zarządzania kolekcją zasobów genowych danego gatunku rośliny uprawnej (van Hintum i in., 2000). Po przeprowadzeniu alokacji obiektów opartej na wielkości czy różnorodności poszczególnych warstw, kurator subiektywnie może zdecydować o zwiększeniu lub zmniejszeniu wyznaczonej liczby obiektów, które przechodzą do kolekcji podstawowej z warstw. Kurator może zdecydować o zmniejszeniu liczby obiektów przekazanych do kolekcji podstawowej z warstw posiadających obiekty o małym znaczeniu dla zachowania różnorodności genetycznej danego gatunku w banku genów lub/i o małym znaczeniu dla hodowli nowych odmian czy małej przydatności w badaniach naukowych. Dla warstw złożonych z obiektów o dużym znaczeniu dla zachowania różnorodności genetycznej gatunku w banku genów lub/i dużym znaczeniu w hodowli roślin, kurator może podjąć decyzję o zwiększeniu liczby alokowanych obiektów (Brown, 1995; van Hintum i in., 2000).

WYBÓR OBIEKTÓW Z POSZCZEGÓLNYCH WARSTW

Ostatnią procedurą statystyczną stosowaną w trakcie pobierania próby warstwowej jest wybór alokowanej liczby obiektów z każdej warstwy wyodrębnionej w kolekcji wyjściowej, które mają być włączone do kolekcji podstawowej. Najprostszym oraz najczęściej stosowaną metodą (sposobem) wyboru obiektów z wyróżnionych warstw kolekcji podstawowej jest wybór losowy (Cochran, 1977; Brown, 1989 a,b, 1995; van Hintum, 1999; van Hintum i in., 2000; Ghamkhar i in., 2008). Losowy wybór obiektów nie wymaga wiedzy o różnorodności genetycznej obiektów w poszczególnych warstwach.

Kolejną metodą wyboru obiektów z warstw jest subiektywny, preferowany wybór obiektów przez kuratora kolekcji (Thompson, 2002). Kurator decyduje, czy dany obiekt powinien znaleźć się w kolekcji podstawowej, a swoją decyzję podejmuje na podstawie różnych przesłanek (van Hintum, 2000; Ghamkhar i in. 2008; Chung i in., 2009). Na przykład, do kolekcji podstawowej kurator wybiera z warstw te obiekty, o których zgromadzono więcej informacji. Preferowane przez kuratora kolekcji zasobów genowych mogą być te obiekty, które już odegrały wcześniej ważną rolę w hodowli roślin, obiekty o

dużej reputacji i popularne wśród badaczy i hodowców. Kolejnym kryterium preferowanego wyboru obiektów z warstw, stosowanym przez kuratorów jest dostępność materiału biologicznego.

W trakcie wyboru próby warstwowej do tworzenia kolekcji podstawowej stosuje się także inne metody wyboru obiektów z poszczególnych warstw. Metody te wymagają zgromadzenia możliwie wszechstronnej informacji o różnorodności obiektów w kolekcji wyjściowej. Dla uzyskania takich informacji można przeprowadzić dla każdej warstwy wielocechową analizę danych, na przykład analizę składowych głównych, czy analizę skupień. Crossa i in. (1995) zaproponowali podział każdej warstwy za pomocą analizy skupień na taką liczbę podgrup (podwarstw), która jest równa liczbie obiektów alokowanych w warstwie, z przeznaczeniem ich do kolekcji podstawowej. Z każdej tak powstałej podwarstwy losowo wybierany jest jeden obiekt. Noirot i in. (1996) zaproponowali metodę wyboru obiektów z poszczególnych warstw opartą na wartościach składowych głównych (ang. principal component scoring, *PCS*). Jest to metoda iteracyjna, maksymalizująca wartość uogólnionej sumy kwadratów (ang. generalized sum of square, *GSS*), której celem jest wybranie alokowanej liczby obiektów z każdej warstwy w taki sposób, aby charakteryzowały się one największą różnorodnością pod względem wartości składowych głównych. Autorzy metody zastosowali ją do tworzenia kolekcji podstawowej na podstawie sztucznych danych dla cech ilościowych obiektów w kolekcji wyjściowej. Stwierdzili oni skuteczność metody *PCS* w tworzeniu kolekcji podstawowej o dużym stopniu reprezentatywności pod względem różnorodności genetycznej. Stosując metodę *PCS* utworzono kolekcję podstawową zasobów genowych trzciny cukrowej (*Saccharum officinarum* L.), opierając się na zmienności cech ilościowych w kolekcji wyjściowej (Balakrishnan i in., 2000) oraz sorga cukrowego (*Sorghum bicolor* (L.) Moench), opierając się na zmienności markerów molekularnych w kolekcji wyjściowej (Grenier i in., 2000 b). Studnicki i in. (2009) zastosował wartości pierwszej zmiennej kanonicznej, wyznaczonej dla obiektów w kolekcji wyjściowej scharakteryzowanych za pomocą cech ilościowych, jako kryterium wyboru obiektów z poszczególnych warstw. Do kolekcji podstawowej włączono obiekty o skrajnych (maksymalnych i minimalnych) wartościach pierwszej zmiennej kanonicznej.

MODYFIKACJE METOD PRÓBY WARSTWOWEJ DO TWORZENIA KOLEKCJI PODSTAWOWEJ

Do tworzenia kolekcji podstawowych nie stosuje się wyłącznie metod wykorzystujących klasyczne podejście próby warstwowej. Opracowano wiele metod tworzenia kolekcji podstawowych, będących modyfikacjami metod pobierania próby warstwowej. Jedną z nich jest strategia M (ang. maximization strategy lub M-strategy), zaproponowana przez Schoen i Brown (1993). Metoda ta pozwala wybrać takie obiekty, aby kolekcja podstawowa cechowała się możliwie maksymalną liczbą alleli w obrębie rozpatrywanych *loci* dla markerów molekularnych. W iteracyjny sposób wyszukuje się takiej kombinacji obiektów, aby zmaksymalizować bogactwo alleli w nowo utworzonej kolekcji podstawowej. Strategia M określa liczbę obiektów z poszczególnych warstw wybieranych

do kolekcji podstawowej oraz wskazuje konkretnie te obiekty. Zatem, jest ona połączeniem zarówno metody alokacji obiektów, jak i metody wyboru obiektów z warstw do kolekcji podstawowej. Strategia M jest najczęściej stosowaną metodą do tworzenia kolekcji podstawowej na podstawie danych dla markerów molekularnych u obiektów (Gouesnard i in., 2001, 2005; Franco i in., 2006; Kim i in., 2007). Strategię M zastosowano z powodzeniem przy tworzeniu wielu kolekcji podstawowych, między innymi dla koniczyny (*Trifolium* sp.) — (Ghamkhar i in., 2005), kukurydzy zwyczajnej (*Zea mays* L.) — (Gouesnard i in., 2005; Franco i in., 2006), rzodkiewnika pospolitego (*Arabidopsis thaliana* L.) — (McKhann i in., 2004; Ronfort i in., 2006) winorośli właściwej (*Vitis vinifera* L.) — (Barnaud i in. 2006) oraz pszenicy zwyczajnej (*Triticum aestivum* L.) — (Balfourier i in., 2007). Zastosowano tą strategię również w celu utworzenia mini kolekcji podstawowej dla kolekcji zasobów genowych ryżu siewnego (*Oryza sativa* L.) — (Agrama i in., 2009; Zhang i in., 2011). Opracowano także programy komputerowe, mające na celu aplikację strategii M do tworzenia kolekcji podstawowych. Pierwszym z nich jest program MSTRAT (Gouesnard i in. 2001), aktualną wersję można za darmo pobrać ze strony autora programu <http://www.ensam.inra.fr/gap/MSTRAT/> (weryfikacja 25.04.2012). Kolejną aplikacją do tworzenia kolekcji podstawowych, stosującą strategię M, jest program PowerCore (Kim i in., 2007). Wykorzystuje on algorytm heurystycznego wyszukiwania (ang. heuristic search algorithm) optymalnej kombinacji obiektów, które powinny znaleźć się w kolekcji podstawowej. Jest to ten sam algorytm, który stosuje się między innymi do rozwiązywania problemu komiwojażera (ang. traveling salesman problem). Algorytmy tego typu są powszechnie stosowane w różnych narzędziach bioinformatycznych, na przykład do przeszukiwania i odnajdywania sekwencji aminokwasowych.

Kolejną metodą tworzenia kolekcji podstawowej, stanowiącą modyfikację warstwowego pobierania próby, jest zaproponowana przez Jansena i van Hintuma (2007) metoda pobierania próby obiektów do kolekcji podstawowej oparta na odległościach genetycznych (ang. genetic distance sampling). Pierwszym krokiem w tej metodzie jest wyznaczenie macierzy odległości genetycznych dla obiektów w kolekcji wyjściowej. Pierwszy obiekt jest wybierany losowo ze zbioru wszystkich obiektów w kolekcji wyjściowej. W kolejnym etapie z kolekcji wyjściowej eliminowane są te obiekty, których odległości od obiektu wybranego do kolekcji podstawowej są mniejsze, niż pewna wartości r . Z takiej okrojonej listy obiektów wybierany jest losowo kolejny obiekt, operacja eliminacji obiektów znowu jest powtarzana. Obiekty z kolekcji wyjściowej wybierane są do kolekcji podstawowej dotąd, aż zredukowany zbiór obiektów w kolekcji wyjściowej zostaje pusty. Wielkość otrzymanej kolekcji podstawowej jest różna w zależności od wybranej wartości r . Im mniejsza jest wartość tego parametru, tym otrzymamy większą (liczniejszą) kolekcję podstawową. Zwiększenie wartości r przyczynia się do ograniczenia wielkości kolekcji podstawowej. Wielkość kolekcji podstawowej przy zastosowaniu tej metody nie jest powtarzalna; za każdym razem, gdy ją zastosujemy możemy otrzymać kolekcję podstawową o różnej liczebności, ale także o różnym składzie obiektów. W celu eliminacji tej niedogodności autorzy proponują zastosowanie swojej metody dla kolekcji wyjściowej, która wcześniej została poddana warstwowaniu. Zatem, dla każdej warstwy wykonywana jest osobno procedura opisana wyżej. Dzięki zastosowaniu proponowanej metody, z

warstw o dużej wartości przeciętnej wewnątrzgrupowej odległości genetycznej, wybieranych jest stosunkowo więcej obiektów do kolekcji podstawowej, niż z warstw charakteryzujących się małymi wartościami przeciętnej wewnątrzgrupowej odległości.

Metoda tworzenia kolekcji podstawowej zaproponowana przez Diwana i in. (1994) jest kolejną modyfikacją metod warstwowego wyboru próby. Modyfikacja ta polega na powtórnym warstwowaniu obiektów w każdej warstwie wyróżnionej w pierwszym kroku. Każda warstwa jest dzielona na tyle grup (podwarstw), ile obiektów ma przejść do kolekcji podstawowej (ile wynosi liczba alokowanych obiektów w tej warstwie). Z każdej tak utworzonej podwarstwy wybiera się losowo jeden obiekt, który zostaje przekazany do kolekcji podstawowej. Autorzy proponują również alternatywny sposób wyboru obiektów z każdej utworzonej podwarstwy. Wybór jednego obiektu opiera się na różnorodności obiektów zgromadzonych w poszczególnych podwarstwach (ang. Relative Diversity Method). To podejście zastosowano do utworzenia kolekcji podstawowej w obrębie krajowych zasobów genowych lucerny (*Medicago sp.*) w USA (Diwan i in., 1995) oraz kukurydzy zwyczajnej (*Zea mays L.*), w obrębie urugwajskiej kolekcji wyjściowej zasobów genowych tego gatunku (Malosetti i Abadie, 2001).

Kolejną metodą wykorzystywaną do tworzenia kolekcji podstawowej jest metoda krokowego przeprowadzania analizy skupień (ang. stepwise clustering), zaproponowana przez Hu i in. (2000). Wykorzystując analizę skupień za pomocą metody UPGMA lub metody Warda tworzony jest dendrogram. Przecięcie dendrogramu następuje w miejscu, przy którym tworzone są grupy zawierające maksymalnie dwa obiekty. Do kolejnej analizy skupień przekazywane są obiekty wybrane z grup zawierających dwa obiekty. Z grup zawierających jeden obiekt, bezpośrednio jest on przekazywany do kolejnej analizy skupień. Kolejna analiza skupień wykonywana jest na podstawie nowo wyznaczonej macierzy odległości dla wybranych lub przekazanych z poprzedniej analizy skupień obiektów. Analizę skupień przeprowadza się do momentu, gdy wielkość utworzonej kolekcji podstawowej nie przekracza 20–30% kolekcji wyjściowej. Autorzy krokowej analizy skupień zaproponowali trzy metody wyboru obiektów z każdej warstwy zawierającej dwa obiekty, które przechodzą do kolejnej analizy skupień. Pierwszą z nich jest losowy wybór obiektów z podwarstw zawierających dwa obiekty, natomiast z grup zawierających jeden obiekt, przechodzi on do kolejnego etapu analizy skupień. Drugim zaproponowanym sposobem wyboru obiektów z grupy jest subiektywny wybór obiektów (ang. preferred sampling), składających się z dwóch obiektów przez kuratora kolekcji wyjściowej. Do dalszych etapów w trakcie tworzenia kolekcji podstawowych przez kuratora mogą być wybierane te obiekty, które charakteryzują się minimalnymi lub maksymalnymi wartościami dla rozpatrywanej cechy lub cech. Trzeci zaproponowany sposób wyboru obiektów z dwu obiektowej grupy jest oparty na porównaniu różnorodności; do kolejnego etapu analizy skupień wybierane są obiekty cechujące się większą różnorodnością. Jeszcze inną metodą wyboru obiektów do ponownego przeprowadzenia analizy skupień z grupy zawierającej dwa obiekty zaproponowali Wang i in. (2007). Według nich, do następnego etapu analizy skupień wybierane są te obiekty, które cechują się najniższymi wartościami odległości w grupach z dwoma obiektami (ang. least distance strategy). Metodę krokowej analizy skupień zastosowano między innymi to utworzenia

kolekcji podstawowej bawełny (*Gossypium* L.) — (Hu i in., 2000; Xu i in., 2006; Wang i in., 2007; Wang i in. 2008), ryżu siewnego (*Oryza sativa* L.) — (Li i in., 2004) oraz bakłażanu — psianki podłużnej (*Solanum melongena*) — (Weihai i in., 2008).

OCENA EFEKTYWNOŚCI METOD POBIERANIA PRÓBY W TWORZENIU KOLEKCJI PODSTAWOWYCH

Ze względu na dużą różnorodność metod pobierania próby, stosowanych do tworzenia kolekcji podstawowych, ważnym celem badawczym staje się empiryczna ocena efektywności proponowanych metod. Zadaniem takiej oceny jest wskazanie dla przykładowych kolekcji wyjściowych, która z proponowanych metod lub grupa metod pozwala na utworzenie kolekcji podstawowych w najwyższym stopniu reprezentatywnych pod względem różnorodności genetycznej, to znaczy przenoszących (zachowujących) możliwie największą część różnorodności genetycznej obiektów zgromadzonych w kolekcji wyjściowej, pozbywając się jednocześnie duplikatów (obiektów genetycznie podobnych do siebie).

W dotychczasowych pracach nad oceną efektywności metod pobierania próby w tworzeniu kolekcji podstawowej można spotkać dwa podejścia. Pierwsze z nich polega na ocenie reprezentatywności w klasycznym rozumieniu statystycznym (reprezentatywności pod względem zmienności genetycznej) prób, jako kolekcji podstawowych tworzonych przez badane metody. Drugie podejście opiera się na badaniu reprezentatywności w stosunku do różnorodności genetycznej obiektów (reprezentatywności pod względem różnorodności genetycznej) prób, jako kolekcji podstawowych tworzonych przez badane metody. Pierwsze z tych podejść uznaje daną metodę pobierania próby za efektywną w tworzeniu kolekcji podstawowej, jeśli generowane przez tą metodę kolekcje podstawowe reprezentują dobrze rozkład prawdopodobieństwa częstości alleli w rozpatrywanych *loci* i/lub cech fenotypowych dla obiektów w kolekcji wyjściowej. To znaczy, że postać tego rozkładu w próbce i populacji wyjściowej jest podobna, wtedy też średnia, rozstęp, wariancja i korelacje cech fenotypowych w obu rodzajach kolekcji są podobne. Oznacza to, że metoda pobierania próby przenosi rozkład prawdopodobieństwa atrybutów genetycznych lub fenotypowych kolekcji wyjściowej do kolekcji podstawowej (Hu i in., 2000; Li i in., 2004; Li i in. 2005; Xu i in., 2006; Wang i in., 2007; Upadhyaya i in., 2007; Wang i in., 2008; Weihai i in., 2008). To podejście do oceny efektywności metod pobierania próby nie jest w pełni zgodne z pierwotną definicją kolekcji podstawowej zasobów genowych (Frankel, 1984; Frankel i Brown, 1984; Brown, 1989 a,b) oraz rozpatrywanej ideologicznie i metodycznie przez znaczącą liczbę badaczy (Diwan i in., 1995; Marita i in., 2000; van Hintum i in., 2000; van Raamsdonk i Wijnker, 2000; Malosetti i Abadie, 2001; Franco i in., 2005, 2006; Logozzo i in., 2007; Yan i in., 2007; Thachuk i in., 2009; Oliveira i in., 2010). Metody pobierania próby uznane w tym podejściu za efektywne, czyli tworzące kolekcje podstawowe dobrze reprezentujące zmienność kolekcji wyjściowej, nie muszą tworzyć dobrej kolekcji podstawowej, czyli przenoszącej z kolekcji wyjściowej możliwie dużą część jej różnorodności genetycznej (Marita i in., 2000; van Hintum i in., 2000; Franco i in., 2005, 2006; Thachuk i in., 2009).

W drugim podejściu do oceny efektywności metod pobierania próby, zastosowana metoda jest tym bardziej efektywna, im utworzone przez nią kolekcje podstawowe lepiej odzwierciedlają (reprezentują) różnorodność genetyczną kolekcji wyjściowej, pomijając w dużym stopniu obiekty podobne genetycznie (Marita i in., 2000; van Hintum i in. 2000; van Raamsdonk i Wijnker, 2000; Malosetti i Abadie, 2001; Franco i in., 2005, 2006; Logozzo i in., 2007; Thachuk i in., 2009; Yan i in., 2007). Efektywność metod pobierania próby w tworzeniu kolekcji podstawowych jest oceniana przez stopień ich reprezentatywności pod względem różnorodności genetycznej w kolekcji wyjściowej. Aby uznać daną metodę pobierania próby za efektywną w tworzeniu kolekcji podstawowych, to tworzone kolekcje podstawowe powinny spełniać następujące warunki: 1) odzwierciedlać jak największą część różnorodności genetycznej fenotypowych cech ilościowych w kolekcji wyjściowej, czyli powinny być złożone z obiektów bardziej zróżnicowanych od siebie w porównaniu do obiektów w kolekcji wyjściowej, a jednocześnie nie powinny odznaczać się przesunięciem średnich rozpatrywanych cech względem wartości tego parametru w kolekcji wyjściowej, 2) zawierać ograniczoną liczbę obiektów wielocechowo podobnych do siebie (Frankel, 1984; Brown 1989a,b, 1995; Marita i in., 2000; van Hintum i in., 2000; Franco i in., 2005; Jansen i van Hintum, 2007; Oliveira i in., 2010).

Do ilościowej oceny efektywności metod pobierania próby, wyrażonej w kategoriach reprezentatywności tworzonej kolekcji podstawowej pod względem różnorodności genotypowej cech ilościowych, opracowano oraz zastosowano wiele parametrów statystycznych. Wiele z tych parametrów może służyć do oceny efektywności metod pobierania próby zarówno w kategoriach reprezentatywności kolekcji podstawowej (próby) w sensie statystycznym, jak i jej reprezentatywności pod względem różnorodności genetycznej. Tylko odpowiednia interpretacja wyników, otrzymanych dla poszczególnych parametrów decyduje o tym, czy oceniają one efektywność metod pobierania próby reprezentatywnej pod względem jednego z wymienionych podejść (Hu i in., 2000; Li i in. 2004, 2007; Xu i in., 2006; Wang i in., 2007; Oliveira i in., 2010).

Pierwszym parametrem, stosowanym do oceny efektywności metod próbkowania w tworzeniu kolekcji podstawowej, jest procent istotnie zróżnicowanych średnich (MSD%), określony za pomocą następującego wzoru (Hu i in., 2000; Li i in., 2004; Xu i in., 2006; Wang i in., 2007; Wang i in., 2008):

$$MSD\% = \frac{S_m}{p} \times 100\%$$

gdzie: S_m jest liczbą cech dla których stwierdzono istotną różnicę średnich pomiędzy kolekcją wyjściową, a kolekcją podstawową za pomocą testu t-Studenta przy poziomie istotności $\alpha=0,05$, p jest liczbą rozpatrywanych cech. W celu zbadania istotności różnicy średnich w obu rodzajach kolekcji stosuje się 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, stosuje się test t-Studenta z aproksymacją Cochra i Coxa (SAS Institute Inc. 2004; Studnicki i in., 2009). Warto podkreślić, że taka metodyka nie jest poprawna statystycznie, ponieważ kolekcja

wyjściowa i wyprowadzona z niej kolekcja podstawowa (jako próba) nie są populacjami niezależnymi. Drugim parametrem, stosowanym do oceny efektywności metod próbkowania w tworzeniu kolekcji podstawowej, jest procent istotnie zróżnicowanych wariacji (VSD%), określony za pomocą następującego wzoru (Hu i in., 2000):

$$VSD\% = \frac{S_v}{p} \times 100\%$$

gdzie: S_v jest liczbą cech dla których stwierdzono istotną różnicę wariacji pomiędzy kolekcją wyjściową, a kolekcją podstawową za pomocą testu F przy poziomie istotności $\alpha=0,05$, p jest liczbą rozpatrywanych cech. Istotność różnicy wariacji pomiędzy kolekcją podstawową, a wyjściową badano za pomocą testu F (Hu i in., 2000; Li i in., 2004; SAS Institute Inc., 2004; Xu i in., 2006). Natomiast, w pracach Upadhyaya i Ortiz (2001) oraz Upadhyaya i in. (2006b, 2007a,b) do wykrywania istotnych różnic wariacji zastosowano test Levene (1960). Krytyczna uwaga o poprawności statystycznej odnosi się także do dwóch ostatnich parametrów (MSD% i VSD%), stosowanych do oceny efektywności metod próbkowania w tworzeniu kolekcji podstawowej, ponieważ są one oparte także na testowaniu istotności różnic średnich i wariacji w kolekcji wyjściowej i podstawowej.

W pracy Kim i in. (2007) zaproponowano parametry oceny efektywności metod pobierania próby, w których różnicę średniej i wariacji pomiędzy kolekcją wyjściową, a kolekcją podstawową określano w sposób ilościowy, a nie za pomocą testów statystycznych, informujących o frakcji istotnych różnic tych parametrów statystycznych. Pierwszym parametrem, zaproponowanym przez Kim i in. (2007), jest przeciętna bezwzględna różnica średnich pomiędzy kolekcją podstawową, a wyjściową, obliczona poprzez rozpatrywane cechy i relatywnie w stosunku do średniej w kolekcji wyjściowej (MD%). Parametr MD% jest określony za pomocą następującego wzoru:

$$MD\% = \frac{\sum_{\tau=1}^p \frac{|\bar{x}_{C\tau} - \bar{x}_{E\tau}|}{\bar{x}_{E\tau}}}{p} \times 100\%$$

gdzie: p jest liczbą rozpatrywanych cech, $\bar{x}_{C\tau}$ jest średnią wartością dla τ -tej cechy w kolekcji podstawowej ($\tau=1,2,\dots,p$), $\bar{x}_{E\tau}$ jest średnią wartością dla τ -tej cechy w kolekcji wyjściowej.

Drugim parametrem, zaproponowanym przez Kim i in. (2007), jest przeciętna bezwzględna różnica wariacji pomiędzy kolekcją podstawową, a wyjściową, obliczona poprzez rozpatrywane cechy i relatywnie w stosunku do wariacji w kolekcji wyjściowej, VD%. Parametr VD% jest określony za pomocą następującego wzoru:

$$VD\% = \frac{\sum_{\tau=1}^p \frac{|\hat{\sigma}_{C\tau}^2 - \hat{\sigma}_{E\tau}^2|}{\hat{\sigma}_{E\tau}^2}}{p} \times 100\%$$

gdzie: p jest liczbą rozpatrywanych cech, $\hat{\sigma}_{C\tau}^2$ jest wariancją dla τ -tej cechy w kolekcji podstawowej, $\hat{\sigma}_{E\tau}^2$ jest wariancją dla τ -tej cechy w kolekcji wyjściowej.

Jeżeli wartość obu parametrów MSD% i VSD%, albo MD% i VD%, jest bliższa zera, tym kolekcja podstawowa lepiej reprezentuje zmienność genetyczną dla rozpatrywanych cech ilościowych (jest bardziej reprezentatywna w sensie statystycznym) w wyjściowej kolekcji zasobów genowych roślin uprawnych (Hu i in., 2000; Li i in., 2004; Kim i in., 2007). Według Hu i in. (2000) wartości parametrów MSD% i VSD% nie powinny przekraczać 20% cech istotnie zróżnicowanych pod względem średnich i wariancji w obu rodzajach kolekcji, aby można uznać kolekcję podstawową za dobrze reprezentującą zmienność genetyczną w kolekcji wyjściowej i stwierdzić dobrą efektywność w kategoriach pierwszego podejścia (odzworowania rozkładu prawdopodobieństw zmienności genetycznej w kolekcji wyjściowej), odpowiadającej metody pobierania próby w tworzeniu kolekcji podstawowej. Natomiast, kolekcja podstawowa jest tym bardziej reprezentatywna pod względem różnorodności genetycznej, a odpowiadająca metoda pobierania próby ma większą efektywność w kategoriach drugiego podejścia, jeżeli wartość parametru VSD% lub VD% jest relatywnie (w stosunku do prób z innych metod) wyższa, a jednocześnie wartość parametru MSD% lub MD% jest bliska zero (van Hintum i in., 2000; Franco i in., 2005; Liu i in., 2009; Wang i in., 2007; Oliveira i in., 2010).

Do oceny efektywności metod pobierania próby stosowana jest także przeciętny dla badanych cech iloraz rozstępu w kolekcji podstawowej i wyjściowej, RR% (Hu i in., 2000; Franco i in. 2005; Kim i in., 2007), wyznaczony za pomocą następującej formuły:

$$RR\% = \frac{\sum_{\tau=1}^p \frac{R_{C\tau}}{R_{E\tau}}}{p} \times 100\%$$

gdzie: $R_{C\tau}$ jest wartością rozstępu dla τ -tej cechy w kolekcji podstawowej, $R_{E\tau}$ jest wartością rozstępu dla τ -tej cechy w kolekcji wyjściowej, p jest liczbą rozpatrywanych cech

Metoda pobierania próby jest tym bardziej efektywna pod względem reprezentatywności zmienności genetycznej tworzonych prób, jeżeli wartość tego parametru w wynikowej kolekcji podstawowej jest bliższa 100%. Parametr RR% określa tylko reprezentatywność próby w klasycznym sensie.

Kolejnym parametrem, stosowanym do oceny efektywności metod pobierania próby jest przeciętny, dla badanych cech, iloraz wartości współczynnika zmienności w kolekcji podstawowej i wyjściowej, VR% (Hu i in., 2000; Kim i in., 2007), wyznaczony za pomocą następującej formuły:

$$VR\% = \frac{\sum_{\tau=1}^p \frac{CV_{C\tau}}{CV_{E\tau}}}{p} \times 100\%$$

gdzie: $CV_{C\tau}$ jest wartością współczynnika zmienności dla τ -tej cechy w kolekcji podstawowej, $CV_{E\tau}$ jest współczynnikiem zmienności dla τ -tej cechy w kolekcji wyjściowej, p jest liczbą rozpatrywanych cech,

Metoda pobierania próby jest tym bardziej efektywna pod względem reprezentatywności zmienności, im wartość tego wskaźnika jest bliższa 100%. Natomiast, wartości tego parametru powyżej 100% wskazują na dobrą reprezentatywność próby pod względem różnorodności genetycznej i pozbycia się duplikatów w kolekcji podstawowej.

Kolejnym parametrem, stosowanym do oceny efektywności metod pobierania próby w kategoriach drugiego podejścia, jest średnia odległość wielocechowa dla obiektów w kolekcji podstawowej i wyjściowej (Franco i in., 2005, 2006). Im wartość średniej odległości w kolekcji podstawowej jest większa, od średniej odległości pomiędzy obiektami w kolekcji wyjściowej (w stosunku do innych metod pobierania próby), tym odpowiadająca jej metoda pobierania próby jest efektywniejsza w przenoszeniu możliwie największej części różnorodności genetycznej z kolekcji wyjściowej do kolekcji podstawowej oraz ograniczonego przenoszenia do niej duplikatów.

Powyżej zaprezentowane parametry statystyczne do oceny efektywności metod pobierania próby odnoszą się do fenotypowych cech ilościowych. Natomiast reprezentatywność kolekcji podstawowych w sensie statystycznym dla cech jakościowych przeprowadzana jest przez badanie zgodności rozkładu prawdopodobieństw tych cech w kolekcji podstawowej i wyjściowej. Takie badanie jest przeprowadzane za pomocą testu zgodności chi-kwadrat χ^2 (Upadhyaya, 2003; Dwivedi i in., 2005; Upadhyaya i in., 2006 b).

Dla cech ilościowych i jakościowych wykorzystywanych jako kryterium fenotypowe do oceny zmienności genetycznej w kolekcji wyjściowej i podstawowej, stosowany jest wskaźnik różnorodności Shannona-Weavera H' (Charmet i Balfourier, 1995; Balakrishnan i in., 2000; Grenier i in., 2001; Upadhyaya i Ortiz, 2001; Upadhyaya, 2003; Li i in. 2005; Mahalakshmi i in., 2007). Dla pojedynczej cechy wskaźnik różnorodności Shannona-Weavera H' jest określony za pomocą następującej formuły:

$$H' = \sum_{i=1}^S p_i \ln p_i$$

gdzie: S jest liczbą klas (przedziałów) w szeregu rozdzielczym zmienności badanej cechy, p_i jest stosunkiem liczebności obiektów w i -tej klasie (przedziale) do liczebności wszystkich obiektów kolekcji. Wartość wskaźnika H' w kolekcji podstawowej zbliżona do jego wartości w kolekcji wyjściowej dla wszystkich badanych cech, świadczy o dużej efektywności metody pobierania próby w kategoriach reprezentatywności rozkładu prawdopodobieństw obiektów w kolekcji wyjściowej. Natomiast, większa wartość wskaźnika różnorodności H' w kolekcji podstawowej w stosunku do wartości w kolekcji wyjściowej, świadczy o dobrej efektywności metody pobierania próby w kategoriach reprezentatywności różnorodności genetycznej obiektów w kolekcji wyjściowej.

LITERATURA

- Agrama H., Yan A., Wen Gui L., Fleet F., Robert C., Ming-Hsuan J., McClung M. 2009. Genetic Assessment of a Mini-Core Subset Developed from the USDA Rice Genebank. *Crop Science* 49: 1336 — 1346.
- Amalraj V.A., Balakrishnan R., Jebadhas A.W., Balasundaram N. 2006. Constituting a core collection of *Saccharum spontaneum* L. and comparison of three stratified random sampling procedures *Genetic Resources and Crop Evolution* 53: 1563 — 1572.
- 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.
- Balfourier F., Roussel V., Strelchenko P., Exbrayat-Vinson F., Sourdill P., Boutet G., Koenig J., Ravel C., Mitrofanova O., Beckert M., Charmet G. 2007. A worldwide bread wheat core collection arrayed in a 384-well plate. *Theoretical and Applied Genetics* 114: 1265 — 1275.
- Brown A.H.D. 1989 a. Core collections: a practical approach to genetic resources management. *Genome* 31: 818 — 824.
- Brown A. H. D. 1989 b. The case for core collections. In: *The Use of Plant Genetic Resources* (A.H.D. Brown, O. H. Frankel, D. R. Marshall and J. T. Williams, eds.). Cambridge University Press, Cambridge, UK.
- Brown A.H.D. 1995. The core collection at the crossroads. w: *Core Collections of Plant Genetic Resources* (T. Hodgkin, A.H.D. Brown, Th. J. L. van Hintum and E. A. V. Morales, eds.). John Wiley and Sons, UK.
- Brown A.H.D., Spillane C. 1999. Implementing core collections — principles, procedures, progress, problems and promise. w: Johnson, R.C. and T. Hodgkin. 1999. *Core collections for today and tomorrow*. International Plant Genetic Resources Institute, Rome, Italy.
- Bulińska-Radomska Z., Łapiński B., Arseniuk E. 2008. Plant genetic resources for food and agriculture in Poland – Second National Report. Plant Breeding and Acclimatization Institute, Radzików, Polska
- Chandra S., Huaman Z., Krishna S.H., Ortiz R. 2002. Optimal sampling strategy and core collection size of Andean tetraploid potato based on isozyme data — a simulation study. *Theoretical and Applied Genetics* 104: 1325 — 1334.
- Charmet G., Balfourier F. 1995. The use of geostatistics for sampling a core collection of perennial ryegrass population. *Genetic Resources and Crop Evolution* 42: 303 — 309.
- Chavarriga-Aguirre P., Maya M., Tohme J., Duque M. C., Iglesias C., Bonierbale M. W., Kresovich S., Kochert G. 1999. Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collection and to assess the usefulness of DNA-based markers to maintain germplasm collections. *Molecular Breeding* 5: 263 — 273.
- Chung H.K., Kim K.W., Chung J.W., Lee J.R., Lee S.Y., Dixit A., Kang H.K., Zhao W., McNally K.L., Hamilton R.S., Gwag J.G., Park Y.J. 2009. Development of a core set from a large rice collection using a modified heuristic algorithm to retain maximum diversity. *Journal of Integrative Plant Biology* 51: 1116 — 1125.
- Cochran W.G. 1977. *Sampling techniques*, 3rd ed., John Wiley and Sons, New York, USA
- Crossa J., DeLacy I.H., Taba S. 1995. The use of multivariate methods in developing a core collection. Pp. in *Core Collections of Plant Genetic Resources* (T. Hodgkin, A.H.D. Brown, Th.J.L. van Hintum and E.A.V. Morales, eds.). John Wiley and Sons, UK: 77 — 92.
- Crossa J., Franco J. 2004. Statistical methods for classifying genotypes. *Euphytica* 153: 19 — 37.
- Diwan N., Bauchan G.R., McIntosh M.S. 1994. A core collection for the United States annual *Medicago* germplasm collection. *Crop Sci.* 34: 279 — 285.
- Diwan N., McIntosh M.S., Bauchan G. R. 1995. Methods of developing a core collection of annual *Medicago* species. *Theoretical and Applied Genetics* 90: 755 — 761.
- Dwivedi S. L., Puppala N., Upadhyaya H.D., Manivannan N., Singh S. 2008. Developing a core collection of peanut specific to Valencia market type. *Crop Sci.* 48: 625 — 632.
- Dwivedi S. L., Upadhyaya H. D., Hegd D.M. 2005. Development of core collection using geographic information and morphological descriptors in safflower (*Carthamus tinctorius* L.) germplasm. *Genetic Resources and Crop Evolution* 52: 821 — 830.

- Escribano P., Viruel M., Hormaza J. 2008. Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (*Annona cherimola*, *Annonaceae*), an underutilised subtropical fruit tree species. *Annals of Applied Biology* 153: 25 — 32.
- FAO 1996. Global plan of action for the conservation and sustainable utilization of plant genetic resources for food and agriculture. FAO, Rome, Italy.
- FAO 2010. The second report on the state of the world's plant genetic resources for food and agriculture. FAO, Rome, Italy.
- Franco J., Crossa J., Desphande S. 2010. Hierarchical multiple-factor analysis for classifying genotypes based on phenotypic and genetic data. *Crop Sci.* 50: 105 — 117.
- Franco J., Crossa J., Ribout J.M., Betran J. 2001. A method for combining molecular markers and phenotypic attributes for classifying plant genotypes. *Theoretical and Applied Genetics* 103: 944 — 952.
- Franco J., Crossa J., Taba S., Shands H. 2003. A multivariate method for classifying cultivars and studying group x environment x trait interaction. *Crop Sci.* 43: 1249 — 1258.
- Franco J., Crossa J., Taba S., Shands H. 2005. A sampling strategy for conserving genetic diversity when forming core subsets. *Crop Sci.* 45: 1035 — 1044.
- Franco J., Crossa J., Villasenor J., Taba S., Eberhart S.A. 1998. Classifying genetic resources by categorical and continuous variables. *Crop Sci.* 38:1688 — 1696.
- Franco J., Crossa J., Villasenor J., Taba S., Eberhart S.A. 1999. A two-stage, three-way method for classifying genetic resources in multiple environments. *Crop Sci.* 39: 259 — 267.
- Franco J., Crossa J., Warburton M., Taba S. 2006. Sampling strategies for conserving maize diversity when forming core subsets using genetic markers. *Crop Sci.* 46: 854 — 864.
- Frankel O.H. 1984. Genetic perspectives of germplasm conservation. In: W. Arber, K. Llimensee, W.J. Peacock D. P. Starlinger, (eds.). *Genetic Manipulation: Impact on Man and Society*. Cambridge University Press, Cambridge: 161 — 170.
- Frankel O.H., Brown A.H.D. 1984. Plant genetic resources today: A critical appraisal. W: J.H.W. Holden and J.T. Williams, eds. *Crop Genetic Resources: Conservation and Evaluation*. Allen and Unwin, Winchester, Massachusetts, USA.
- Gauthier M.F., Lumaret R. 1999. Genetic introgression on between tetraploid *Dactylis glomerata* sp. *reichenbachii* and *glomerata* in the French Alps. Insight from morphological and isoenzyme variation, plant systematic and evolution. *Plant Syst. Evol.* 241: 219 — 234.
- Ghamkhar K., Snowball R., Bennett S.J. 2005. Improving the utilization of germplasm of *Trifolium spumosum* L. by the development of a core collection using ecogeographical and molecular techniques. p: 262. W: M.O. Humphreys (ed.) *Molecular breeding for the genetic improvement of forage crops and turf*. Wageningen Academic, Wageningen, Hollandia.
- 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.
- Gouesnard B., Bataillon T.M., Decoux G., Rozale C., Schoen D.J., David J.L. 2001. MSTRAT: An Algorithm for Building Germ Plasm Core Collections by Maximizing Allelic or Phenotypic Richness. *Journal of Heredity* 92: 93 — 94.
- Gouesnard B., Dallard J., Bertin P., Boyat A., and A. Charcosset. 2005. European maize landraces: Genetic diversity, core collection definition and methodology of use. *Maydica* 50: 225 — 234.
- Gowda C.L.L., Upadhyaya H.D., Dronavalli N., Singh S. 2011. Identification of large-seeded high-yielding stable kabuli chickpea germplasm lines for use in crop improvement. *Crop Sci.* 51: 198 — 209.
- Gower J.C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27: 857 — 874.
- Grenier C., Bramel-Cox P. J., Noirot M., Prasada Rao K. E., Hamon, P. 2000b. Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs. non-random sampling procedures B. Using molecular markers. *Theoretical and Applied Genetics* 101: 197 — 202.
- Grenier C., Bramel Cox P. J., Noirot M., Prasada Rao K. E., Hamon, P. 2000 a. Assessment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs. non-random sampling

- procedures A. Using morpho-agronomical and passport data. *Theoretical and Applied Genetics* 101: 190 — 196.
- Grenier C., Hamon P., Bramel.Cox P. J. 2001. Core Collection of Sorghum: II. Comparison of Three Random Sampling Strategies. *Crop Sci.* 41: 241 — 246.
- Hao C., Zhang X., Wang L., Dong Y., Shang X., Jia J. 2006. Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. *Molecular Breeding* 17: 69 — 77.
- Harch B.D., Brasford K.E., DeLacy I.H., Lawrence P.K., Cruickshank A. 1995. Patterns of diversity in fatty acid composition in the Australian groundnut germplasm collection. *Genetic Resources and Crop Evolution* 42: 243 — 256.
- Hartung K. 2006. Biometrical approaches for analysing gene bank evaluation data on barley (*Hordeum spec.*). Rozprawa doktorska, Uniwersytet w Hohenheim, Stuttgart, Niemcy.
- Hausmann B.I.G., Parzies H.K., Presterl T., Susic Z., Miedaner T. 2004. Plant genetic resources in crop improvement (Review). *Plant Genetic Resources: Characterization and utilization* 2: 3 — 21.
- Holbrook C.C., Dong W. 2005. Development and evaluation of a mini core collection for the U.S. peanut germplasm collection. *Crop Sci.* 45: 1540 — 1544.
- 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.
- Igartua E., Gracia M., Lasa J., Medina B., Molina-Cano J., Montoya J., Romagosa I. 1998. The Spanish barley core collection. *Genetic Resources and Crop Evolution* 45: 475 — 481.
- Islam M.R., Hamid A., Khaliq O.A., Ahmed J.U., Haque M.M., Karim M.A. 2007. Genetic variability in flooding tolerance of mungbean (*Vigna radiata* L. Wilczek) genotypes. *Euphytica* 156: 247 — 255.
- Jahufer M.Z.Z., Cooper M., Harch B.D. 1997. Pattern analysis of the diversity of morphological plant attributes and herbage yield in a world collection of white clover (*Trifolium repens* L.) germplasm characterised in a summer moisture stress environment of Australia. *Genetic Resources and Crop Evolution* 44: 289 — 300.
- Jansen J., van Hintum Th. 2007. Genetic distance sampling: a novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. *Theoretical and Applied Genetics* 114: 421 — 428.
- Johnson R.A., Wichern D.W. 2002. *Applied multivariate statistical analysis*. Prentice-Hall, Inc. Upper Saddle River, New York, USA.
- Johnson R.C. and Hodgkin T. 1999. Core collections for today and tomorrow. International Plant Genetic Resources Institute, Rome, Italy.
- Khan M. A., von Witzke-Ehbrecht S., Maass B. L., Becker H. C. 2009. Relationships among different geographical groups, agro-morphology, fatty acid composition and RAPD marker diversity in safflower (*Carthamus tinctorius*). *Genetic Resources and Crop Evolution* 6: 19 — 30.
- Kim K.W., Chung H.K., Cho G.T., Ma K.H., Chandrabalan D., Gwag J.G., Kim T.S., Cho E.G., Park Y.J. 2007. PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics* 23: 2155 — 2162.
- Kociuba W., Mađry W., Kramék A., Ukalski K., Studnicki M. 2010. Multivariate diversity of Polish winter triticale cultivars for spike and other traits. *Plant Breeding and Seed Science* 62:31 — 42
- Kölliker R., Stadelmann F.J., Reidy B., Nösberger J. 1999. Genetic variability of forage grass cultivars: A comparison of *Festuca pratensis* Huds., *Lolium perenne* L., and *Dactylis glomerata* L.. *Euphytica* 106: 261 — 270.
- Krebs C. 1989. *Ecological Methodology*. HarperCollins, New York, USA.
- Krzanowski W.J. 1988. *Principles of multivariate analysis: a user's perspective*. Oxford University Press, Oxford, UK.
- Levene H. 1960. Robust tests for equality of variances. In: Olkin, I. (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press, Stanford.
- 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.

- Li Y., Shi Y., Cao Y., Wang T. 2005. Establishment of a core collection for maize germplasm preserved in Chinese National Genebank using geographic distribution and characterization data. *Genetic Resources and Crop Evolution* 51: 845 — 852.
- Li Y.X., Li T.H., Zhang H.L., Qi Y.W. 2007. Sampling strategy for a primary core collection of peach (*Prunus persica* (L.) Batsch.) germplasm. *European Journal of Horticultural Science* 72: 268 — 274.
- Li Z., Zhang H., Zeng Y., Yang Z., Shen S., Sun C., Wang X. 2002. Studies on sampling schemes for the establishment of core collection of rice landraces in Yunnan, China. *Genetic Resources and Crop Evolution* 49: 67 — 74.
- Liu X.L., Cai Q., Ma L., Wu C.W., Lu X., Ying X.M., Fan Y.H. 2009. Strategy of sampling for pre-core collection of sugarcane hybrid. *Acta Agronomica Sinica* 35: 1209 — 1216.
- Logozzo G., Donnoli R., Macaluso L., Papa R., Knüpffer H., Zeuli P. 2007. Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. *Genetic Resources and Crop Evolution* 54: 1763 — 1779.
- Luan F., Delannay I., Staub J. E. 2008. Chinese melon (*Cucumis melo* L.) diversity analyses provide strategies for germplasm curation, genetic improvement, and evidentiary support of domestication patterns. *Euphytica* 164: 445 — 461.
- Mahajan R.K., Bisht I.S., Gautam P.L. 1999. Sampling strategies for developing Indian sesame core collection. *Indian Journal of Plant Genetic Resources* 12: 1 — 9
- 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* 54: 1817 — 1825.
- Malosetti M., Abadie T. 2001. Sampling strategy to develop a core collection of Uruguayan maize landraces based on morphological traits. *Genetic Resources and Crop Evolution* 48: 381 — 390.
- Marita J. M., Rodriguez J. M., Nienhuis J. 2000. Development of an algorithm identifying maximally diverse core collections. *Genetic Resources and Crop Evolution* 47: 515 — 526.
- McKhann H.I., Camilleri C., Bera A., Bataillon T., David J.L., Reboud X., Le Corre V., Gut I.G., Brunel D. 2004. Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J.* 38: 193 — 202.
- Mohammadi S.A., Prasanna M. 2003. Analysis of genetic diversity in crop plants — salient statistical tools and considerations. *Crop Sci.* 43: 1235 — 248.
- Mosjidis J. A., Klingler K. A. 2006. Genetic Diversity in the Core Subset of the U.S. Red Clover Germplasm. *Crop Sci.* 46: 758 — 762.
- Nei M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* 70: 3321 — 3323.
- Neyman J. 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 97: 558 — 625.
- Noirot M., Hamon S., Anthony F. 1996. The principal component scoring: a new method of constituting a core collection using quantitative data. *Genetic Resources and Crop Evolution* 43: 1 — 6.
- Ntundu W.H., Shillah S.A., Marandu W.Y.F., Christiansen J.L. 2006. Morphological diversity of bambara groundnut [*Vigna subterranea* (L.) Verdc.] landraces in Tanzania. *Genetic Resources and Crop Evolution* 53: 367 — 378
- Oliveira M. F., Nelson R. L., Geraldi I. O., Cruz C. D., de Toledo J. F. F. 2010. Establishing a soybean germplasm core collection. *Field Crops Research* 119: 277 — 289.
- Olukolu B.A., Mayes S., Stadler F., Ng N.Q., Fawole I., Dominique D., Azam-Ali S.N., Abbott A.G., Kole C. 2011. Genetic diversity in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) as revealed by phenotypic descriptors and DArT marker analysis. *Genetic Resources and Crop Evolution* DOI: 10.1007/s10722-011-9686-5.
- Ramanatha Rao V., Hodgkin T. 2002. Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell, Tissue and Organ Culture* 68: 1 — 19.
- Reddy L. J., Upadhyaya H.D., Gowda C.L.L., Singh S. 2005. Development of core collection in pigeonpea [*Cajanus cajan* (L.) Millspaugh] using geographic and qualitative morphological descriptors. *Genetic Resources and Crop Evolution* 52: 1049 — 1056.

- Reif J.C., Melchinger A.E., Frisch M. 2005. Genetical and mathematical properties of similarity and dissimilarity coefficients applied to plant breeding and seed bank management. *Crop Sci.* 45: 1 — 7.
- Robertson L.D., Singh K.B., Erskine W., Abd El Moneim A.M. 1996. Useful genetic diversity in germplasm collections of food and forage legumes from West Asia and North Africa. *Genetic Resources and Crop Evolution* 43: 447 — 460.
- Rodino A., Santalla M., Ro A.D., Singh S. 2003. A core collection of common bean from the Iberian peninsula. *Euphytica* 131: 165 — 175.
- Ronfort J., Bataillon T., Santoni S., Delalande M., David J.L., Prospero J-M. 2006. Microsatellite diversity and broad scale geographic structure in a model legume: Building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula* L. *BMC Plant Biology* 6: 28.
- Santos M., Dias J., 2004. Evaluation of a core collection of *Brassica oleracea* accessions for resistance to white rust of crucifers (*Albugo candida*) at the cotyledon stage. *Genetic Resources and Crop Evolution* 51: 713 — 722.
- Schmidt J. 2005 a. The European *Lolium perenne* core collection in the Botanical Garden of the Plant Breeding and Acclimatization Institute, Bydgoszcz, Poland. In: Boller B., Willner E., Maggioni L., Lipman E. Report of a Working Group on Forages. Eighth meeting, 10–12 April 2003, Linz, Austria. International Plant Genetic Resources Institute, Rome, Italy.
- Schmidt J. 2005b. Variation of European ecotypes of perennial ryegrass (*Lolium perenne* L.) in Poland. *Plant Breeding and Seed Science* 51: 75 — 89.
- Schoen D.J., Brown A.H.D. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proceedings of the National Academy of Sciences* 90: 10623 — 10627.
- Skinner D.Z., Bauchan G.R., Auricht G., Hughes S. 1999. Developing a core collection from a large annual *Medicago* germplasm collection. W: Johnson, R.C., Hodgkin T. 1999. Core collections for today and tomorrow. International Plant Genetic Resources Institute, Rzym, Włochy.
- Skroch P.W., Nienhuis J., Beebe S., Tohme J., Pedraza F. 1998. Comparison of Mexican common bean (*Phaseolus vulgaris* L.) core and reserve germplasm collections. *Crop Sci.* 38: 488 — 496.
- Spagnoletti Zeuli P.L., Qualset C.O. 1993. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. *Theoretical and Applied Genetics* 87: 295 — 304.
- Studnicki M., Mądry W., Kociuba W. 2010a. 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 Sci.* 5: 127 — 135.
- Studnicki M., Mądry W., Kociuba W. 2010b. Efektywność metod pobierania próby w tworzeniu kolekcji podstawowej pszenżyta jarego przy użyciu danych fenotypowych. *Zeszyty Problemowe Postępów Nauk Rolniczych* 555: 409 — 418.
- Studnicki M., Mądry W., Śmiałowski T. 2009. Porównanie efektywności metod statystycznych tworzenia kolekcji podstawowej na przykładzie pszenicy jarej. *Biuletyn IHAR* 252: 105 — 117.
- Thachuk C., Crossa J., Franco J., Dreisigacker S., Warburton M., Davenport G.F. 2009. Core Hunter: an algorithm for sampling genetic resources based on multiple genetic measures. *BMC Bioinformatics* 10: 243.
- Thompson S.K. 2002. *Sampling*. 2nd ed. John Wiley & Sons, New York, USA.
- Upadhyaya H.D. 2003. Phenotypic diversity in groundnut (*Arachis hypogaea* L.) core collection assessed by morphological and agronomical evaluations. *Genetic Resources and Crop Evolution* 50: 539 — 550.
- Upadhyaya H.D., Dwivedi S.L., Gowda C., Singh S. 2007b. 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.
- Upadhyaya H.D., Bramel P.J., Ortiz R., Singh S. 2002. Developing a mini core of peanut for utilization of genetic resources. *Crop Sci.* 42: 2150 — 2156.
- Upadhyaya H. D., Dwivedi S. L., Nadaf H. L., Singh S. 2011 a. Phenotypic diversity and identification of wild *Arachis* accessions with useful agronomic and nutritional traits. *Euphytica* (w druku).
- Upadhyaya H. D., Gowda C., Pundir R., Reddy V. G., Singh, S. 2006 b. Development of core subset of finger millet germplasm using geographical origin and data on 14 quantitative traits *Genetic Resources and Crop Evolution*. 53: 679 — 685.

- Upadhyaya H.D., Gowda C.L.L., Buhariwalla H.K., Crouch J.H. 2006 a. Efficient use of crop germplasm resources: identifying useful germplasm for crop improvement through core and mini-core collections and molecular marker approaches. *Plant Genetic Resources: Characterization and Utilization* 4: 25 — 35.
- Upadhyaya H.D., Ortiz R. 2001. A mini core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement *Theoretical and Applied Genetics* 102: 1292 — 1298.
- Upadhyaya H.D., Ortiz R., Bramel P.J., Singh S. 2003. Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. *Genetic Resources and Crop Evolution* 50: 139 — 148.
- Upadhyaya H.D., Ravishankar C.R., Narasimhudu Y., Sarma N.D.R.K., Singh S.K., Varshney S.K., Reddy V.G., Singh S., Parzies H.K., Dwivedi S.L., Nadaf H.L., Sahrawat S., Gowda C.L.L. 2011b. Identification of trait-specific germplasm and developing a mini core collection for efficient use of foxtail millet genetic resources in crop improvement. *Field Crops Research* 124: 459 — 467.
- Upadhyaya H.D., Reddy K.N., Gowda C.L.L., Singh S. 2007a. Phenotypic diversity in the pigeonpea (*Cajanus cajan* L.) core collection. *Genetic Resources and Crop Evolution* 54: 1167 — 1184.
- Upadhyaya H.D., Reddy K.N., Gowda C.L.L., Singh S. 2011c. Development of pearl millet minicore collection for enhanced utilization of germplasm. *Crop Sci.* 51: 217 — 223.
- Upadhyaya H.D., Reddy L.J., Gowda C.L.L., Singh S. 2006. Identification of diverse groundnut germplasm: Sources of early maturity in a core collection. *Field Crops Research* 97: 261 — 271.
- Upadhyaya H.D., Sharma S., Ramulu B., Bhattacharjee R., Gowda C. L. L., Gopal R.V., Singh S. 2010. Variation for qualitative and quantitative traits and identification of trait-specific sources in new sorghum germplasm. *Crop and Pasture Science* 61: 609 — 618.
- van de Wouw M., Chris, K., van Hintum, T., van Treuren, R., Visser, B. 2010a. Genetic erosion in crops: concept, research results and challenges. *Plant Genetic Resources: Characterization and Utilization* 8: 1 — 15.
- van de Wouw M., van Hintum T., Kik C., van Treuren R., Visser B. 2010b. Genetic diversity trends in 20th century crop cultivars — a meta analysis. *Theoretical and Applied Genetics* 120: 1241 — 1252.
- van Hintum Th.J.L. 1999. The general methodology for creating a core collection. w: Johnson, R.C. and T. Hodgkin. 1999. Core collections for today and tomorrow. International Plant Genetic Resources Institute, Rome, Italy.
- van Hintum Th.J.L., Brown A.H.D., C. Spillane and T. Hodgkin. 2000. Core collections of plant genetic resources. IPGRI Technical Bulletin No. 3. International Plant Genetic Resources Institute, Rzym, Włochy.
- van Raamsdonk L., Wijnker J. 2000 The development of a new approach for establishing a core collection using multivariate analyses with tulip as case. *Genetic Resources and Crop Evolution* 47: 403 — 416.
- Vencovsky R., Crossa J. 1999. Variance effective population size under mixed self and random mating with applications to genetic conservation of species. *Crop Sci.* 39: 1282 — 1294.
- Vencovsky R., Crossa J. 2003. Measurements of representativeness used in genetic resources conservation and plant breeding. *Crop Sci.* 43: 6: 1912 — 1921.
- Wang J.C., Hu J., Huang X.X., Xu S.C. 2008. Assessment of different genetic distances in constructing cotton core subset by genotypic values. *Journal of Zhejiang University — Science B* 9:356 — 362
- Wang J.C., Hu J., Xu H. M., Zhang S. 2007. A strategy on constructing core collections by least distance stepwise sampling. *Theoretical and Applied Genetics* 115: 1 — 8.
- Wang Y., Zhang J., Sun H., Ning N., Yang L. 2011. Construction and evaluation of a primary core collection of apricot germplasm in China. *Scientia Horticulturae* 128: 311 — 319.
- Ward J. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 38: 236–244.
- Weihai M., Jinxin Y., Sihachakr D. 2008. Development of core subset for the collection of Chinese cultivated eggplants using morphological-based passport data. *Plant Genetic Resources: Characterization and Utilization* 6: 33 — 40.
- Xiurong Z., Yingzhong Z., Yong C., Xiangyun F., Qingyuan G., Mingde Z., Hodgkin T. 2000. Establishment of sesame germplasm core collection in China. *Genetic Resources and Crop Evolution* 47: 273 — 279.

- 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.
- Xu Y. 2010. *Molecular plant breeding*. CAB International, Wallingford, UK.
- Yan W.G., Ruter J.N., Bryant R.J., Bockelman H.E., Fjellstrom R.G., Chen M.H., Tai T.H., McClung A.M. 2007. Development and evaluation of a core subset of the USDA rice germplasm collection. *Crop Sci.* 47: 869 — 876.
- Zewdie Y., Tong N., Bosland P. 2004. Establishing a core collection of *Capsicum* using a cluster analysis with enlightened selection of accessions. *Genetic Resources and Crop Evolution* 51: 147 — 151.
- Zhang H., Zhang D., Wang M., Sun J., Qi Y., Li J., Wei X., Han L., Qiu Z., Tang S., Li Z. 2011. A core collection and mini core collection of *Oryza sativa* L. in China. *Theoretical and Applied Genetics* 122: 49 — 61.