ISSN 1020-3362

Plant Genetic Resources Newsletter

Bulletin de Ressources Phytogénétiques

Noticiario de Recursos Fitogenéticos

No. 119 Supplement, 1999

Food and Agriculture Organization of the United Nations and the International Plant Genetic Resources Institute
Organisation des Nations Unies pour l'alimentation et l'agriculture et l'institut international des ressources phytogénétiques
Organización de las Naciones Unidas para la Agricultura y la Alimentación y el Instituto Internacional de Recursos Fitogenéticos

IPGRI's Editorial Office - Bureau de rédaction - Oficina de Redacción:

Managing Editor
Plant Genetic Resources Newsletter
IPGRI
Via delle Sette Chiese 142
00145 Rome, Italy

Tel.: +39-0651892233
Email: p.stapleton@cgiar.org, or ipgri@cgiar.org
Fax: +39-065750309
Web: http://www.cgiar.org/ipgri/publicat/

The designations employed, and the presentation of material in the periodical, and in maps which appear herein, do not imply the expression of any opinion whatsoever on the part of IPGRI or FAO concerning the legal status of any country, territory, city or area or its authorities, or concerning the delimitation of its frontiers or boundaries. Similarly, the views expressed are those of the authors and do not necessarily reflect the views of IPGRI or FAO.

Les appellations employées dans cette publication et la présentation des données et cartes qui y figurent n'impliquent de la part de l'IPGRI et de la FAO aucune prise de position quant au statut juridique des pays, territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites. Les opinions exprimées sont celles des auteurs et ne reflètent pas nécessairement celles de l'IPGRI ou de la FAO.

Las denominaciones empleadas, y la forma en que aparecen presentados los datos en esta publicación, no implican, de parte del IPGRI o la FAO, juicio alguno sobre la condición jurídica de países, territorios, ciudades o zonas, o de sus autoridades, ni respecto de la delimitación de sus fronteras o límites. Asimismo, las opiniones expresadas son las de sus autores y no reflejan necesariamente la opinión del IPGRI o la FAO.

Cover: Harvesting wheat in Guatemala (photo by G. Hettel, CIMMYT).

Couverture: Récolte du blé à Guatemala (photo G. Hettel, CIMMYT).

Portada: Cosecha de trigo en Guatemala (foto tomada por G. Hettel, CIMMYT).

© IPGRI/FAO 1999

Genetics - for genetic resources

The XVIIIth International Congress of Genetics was held in Beijing, China, in August 1998 under the auspices of the International Genetics Federation, It was organized by the Chinese Academy of Sciences and the Genetics Society of China under the slogan "Genetics - Better Life for All" and attended by approximately 2000 scientists from around the world.

In welcoming the delegates the President of the Congress, Prof C.C. Tan noted:

"The world - in particular the developing world - is experiencing dropping standards of living, rapid population growth, severe food shortages, and environmental deterioration. To those working in the field, it is apparent that new genetic theories and technologies may help solve or alleviate some of these global problems."

Given the importance of the wise conservation and use of plant genetic resources for improving food security, helping to eradicate poverty and protecting the environment, it was entirely appropriate that IPGRI should organize a symposium at the XVIIIth International Genetics Congress. The aim of the symposium was to explore some of the developments in the genetic aspects of the conservation and use of plant genetic resources.

The very wide subject range of the papers that were presented at the symposium illustrated the many different ways in which new genetic knowledge is contributing to improving conservation and use. This is especially true of the newer molecular genetic technologies that are beginning to provide us with a direct insight into the nature and organization of diversity at the DNA level. Several of the papers presented at the symposium showed how important molecular methods are becoming to understanding patterns of diversity and to improving our ability to use diversity to improve production.

The papers presented at the symposium were of high quality and provided important insights on recent developments in the genetics of conservation. For these reasons, IPGRI proposed that they should be published in a supplementary issue of the *Plant Genetic Resources Newsletter*. Both authors and journal accepted this proposal and the papers published in this special issue were subject to the *Newsletter*'s usual process of peer review. Publishing in this medium has the advantage of making the information in the papers available to a wide audience of those actively involved in plant genetic resources work.

Toby Hodgkin
Symposium organizer and Supplement editor
IPGRI

ARTICLE Plant Genetic Resources Newsletter, 1999, No. 119 Supp.: 2-9

Cytoplasmic diversity in the *Triticum-Aegilops* complex: its conservation and use in wheat breeding

Koichiro Tsunewaki
Fukui Prefectural University, 4-1-1 Kenjyojima, Matsuoka, Yoshida-gun, Fukui 910-1195, Japan.
Tel: +81 0776 61-6000; Fax: +81 0776 61-6011

Summary

Cytoplasmic diversity in the *Triticum-Aegilops* complex: its conservation and use in wheat breeding

The magnitude of cytoplasmic diversity among *Triticum* and *Aegilops* species in terms of molecular and phenotypic variabilities has been demonstrated. DNA diversity per nucleotide pair is about the same among the nuclear, chloroplast and mitochondrial genomes. Genetic variation in most characters, except fertility, produced by an entire range of cytoplasmic variation in the *Triticum*/*Aegilops* complex is only 5 to 30% of the variation found among present-day common wheat genotypes. A new strategy for the preservation and use of cytoplasmic diversity, in the form of extracted organellar DNA, is proposed. Also, some studies aimed at utilizing cytoplasmic variation for wheat improvement are described. Production of high-yielding alloplasmic wheat for enlarging the cytoplasmic variability of wheat is relatively easy by fertile alloplasmon transfer.

Key words: *Aegilops*, alloplasmon, cytoplasmic diversity, cytoplasm substitution, molecular conservation, nucleotide diversity, organellar DNA, wheat, wheat breeding

Résumé

Diversité cytoplasmique dans le complexe *Triticum-Aegilops* : sa conservation et son utilisation dans la sélection du blé

L'importance de la diversité cytoplasmique chez les espèces de *Triticum* et d'*Aegilops* en termes de variabilité moléculaire et phénotypique est bien connue. La diversité de l'ADN, exprimée par paire de nucléotides, est approximativement la même pour le génome nucléaire, chloroplastique et mitochondrial. Pour la plupart des caractères, excepté la fertilité, la variation génétique produite par toute une gamme de variations cytoplasmiques dans le complexe *Triticum*/*Aegilops* s'élève à 5 - 30 % seulement de la variation observée entre les génotypes de blé courants actuels. Une nouvelle stratégie pour la conservation et l'utilisation de la diversité cytoplasmique, sous la forme d'ADN extrait des organites est présentée. Certaines études axées sur l'utilisation de la variation cytoplasmique dans l'amélioration du blé sont également décrites. Il est relativement facile de produire du blé alloplasmique à haut rendement par transfert d'alloplasme fertile, afin d'accroître la variabilité cytoplasmique.

Resumen

La diversidad citoplasmática en el complejo *Triticum-Aegilops*: su conservación y uso en la mejora del trigo

Se ha demostrado la gran diversidad citoplasmática entre las especies de *Triticum* y *Aegilops* en cuanto a variabilidad fenotípica y molecular. La diversidad del ADN por cada pareja de nucleótidos es casi la misma entre los genomas del mitocondrio, del cloroplastio y del núcleo. La mutación genética de muchas características, con excepción de la fertilidad, generada por una entera gama de mutaciones citoplásmicas en el complejo *Triticum*/*Aegilops* representa sólo del 5 al 30% de la mutación individuada entre los genotipos actuales de trigo común. Se propone una nueva estrategia para la conservación y el uso de la diversidad citoplásmica, en forma de ADN organelar extraído y se describen algunos estudios orientados a la utilización de la mutación citoplásmica para la mejora del trigo. La producción de trigo aloplásmico altamente productivo para la extención de la variabilidad citoplásmica se demuestra relativamente fácil gracias a la transferencia de aloplasma fértil.

Introduction

This article is based on my presentation at the symposium "Conservation and use of plant genetic resources", held during the 18th International Congress of Genetics, Beijing, August 10-15, 1998, sponsored by IPGRI, Rome. Four topics are covered, all of which relate to cytoplasmic variability in the *Triticum* (wheat)/*Aegilops* complex: (i) cytoplasmic diversity shown by restriction fragment length polymorphism (RFLP) analyses of chloroplast (ct) and mitochondrial (mt) DNA, (ii) cytoplasmic diversity evaluated by phenotypic variability observed among 12 sets of alloplasmic wheat lines, (iii) a strategy for the molecular conservation of cytoplasmic diversity, and (iv) the use of cytoplasmic variability in improving wheat.

In this article, an alloplasmic wheat line is indicated by its abbreviation, the name of its cytoplasm donor being given in parentheses (often with species name) followed by that of the nucleus donor (with species or cultivar name) and a hyphen between them. For example, (*caudata*)-'Chinese Spring' is an alloplasmic line of *T. aestivum* 'Chinese Spring' and has the cytoplasm of *Ae. caudata*.

Cytoplasmic diversity estimated by RFLP of organellar DNA

Early studies of molecular diversity of cytoplasm in the *Triticum-Aegilops* complex used electrophoretic analysis of endonuclease-treated chloroplast (ct) or mitochondrial (mt) DNA isolated from intact chloroplast or mitochondria, respectively, here referred to as simple RFLP analysis (Vedel *et al.* 1978, 1981; Ogihara and Tsunewaki 1982, 1988; Bowman *et al.* 1983; Terachi and Tsunewaki 1986). This method provided information on the diversity of the entire chloroplast or mitochondrial genome, but required a large amount of purified ct or mtDNA, limiting its application to a small number of samples. In contrast, Southern hybridization analysis, in which small amounts of the total DNA treated with the restriction enzyme were electrophoresed and hybridized with ct or mtDNA probes after Southern transfer to a membrane, usually gave simple band patterns and allowed accurate identification of the homologous bands (Breiman 1987; Graur *et al.* 1989; Miyashita *et al.* 1994; Ogihara *et al.* 1994).

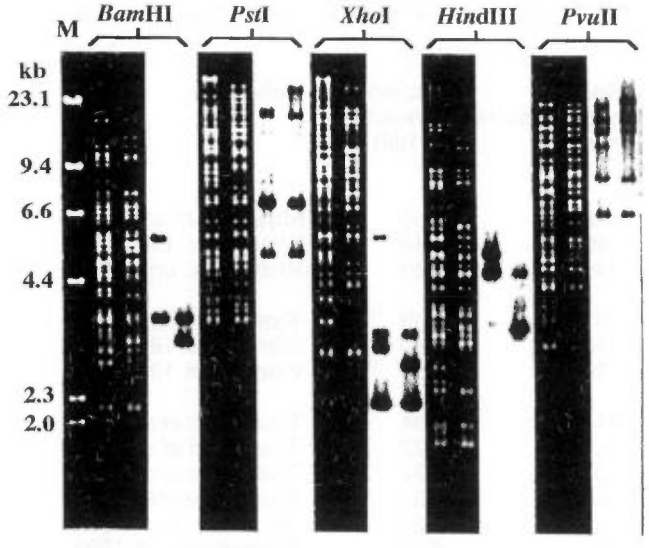

Fig. 1. Simple RFLP and Southern hybridization patterns of restriction endonuclease-treated mtDNA of *T. dicoccoides* and *T. araraticum* (Mori *et al.* 1988). The respective patterns on the black and white backgrounds were obtained by simple RFLP and Southern hybridization analysis. The restriction enzymes used are shown at the top of the corresponding lanes. The probe in Southern hybridization was a wheat 18S-5S rRNA gene clone. MtDNA of *T. dicoccoides* (accession, dcd) and *T. araraticum* (accession, arr2) respectively are on the left- and right-hand sides of each pair of lanes. M: Molecular size markers.

Figure 1 shows the band patterns obtained by simple RFLP and Southern hybridization analyses of mtDNA of *T. dicoccoides* and *T. araraticum* treated with five different 6-base cutters (Mori *et al.* 1988). Patterns with a black background were obtained by simple RFLP analysis, and patterns with a white background by Southern hybridization in which a clone of the wheat 18S-5S rRNA mitochondrial gene was the probe. The two species are paired in all the patterns with *T. dicoccoides* on the left and *T. araraticum* on the right. They showed different simple RFLP patterns. Due to the large number of fragments generated from each mtDNA sample, however, it was difficult to identify fragments that showed differences. In contrast, Southern hybridization patterns revealed distinct differences between the 18S-5S rDNA regions of the two species.

Both the ct and mtDNA diversities among *Triticum/Aegilops* species were studied by the two methods described. Based on the RFLP data obtained, the genetic distances were calculated as 1-psb (the proportion of shared bands) for all possible pairs of the chloroplast or mitochondrial genomes. Figure 2 shows a phylogenetic tree of the mitochondrial genomes in 47 *Triticum/Aegilops* accessions constructed by the UPGMA method that was applied to RFLP data obtained by Southern hybridization with 21 probe-enzyme combinations (Wang *et al.* 1999).

A distance of less than 0.1 was obtained between closely related accessions, e.g. between those of *Ae. umbellulata, Ae. biuncialis* (2 accessions), *Ae. triuncialis* and *Ae. triaristata* (3 accessions). Moderately related accessions showed a distance of ca. 0.2-0.3, e.g. the *Ae. bicornis, Ae. searsii, Ae. kotschyi,* (2 accessions), *Ae. variabilis, Ae. sharonensis* and *Ae. longissima* accessions. Distances of ca. 0.5 or more were observed between distantly related accessions, e.g. between the accessions of emmer-common wheat and *Ae. speltoides*-timopheevi groups. Three groups (einkorn wheat, polyploid wheats including *Ae. speltoides* and all other species) were separated from each other by distances of ca. 0.7.

Nucleotide diversity is that fraction of a nucleotide pair at a given position that differs in a comparison of two DNA molecules. It is estimated from the genetic distance after Nei (1987). Table 1 shows the nucleotide diversities estimated between the chloroplast and mitochondrial genomes of different *Triticum/Aegilops* species (Terachi and Tsunewaki 1992; Miyashita *et al.* 1994; Wang *et al.* 1997, 1999) compared with those of the nuclear genome estimated between different wheat groups (Tsunewaki *et al.* 1993).

Southern hybridization of DNA amplified by polymerase chain elongation reaction (PCR) showed mainly RFLP in the within-gene regions, whereas for total DNA it detected RFLP in both the within-gene and flanking regions. As for the

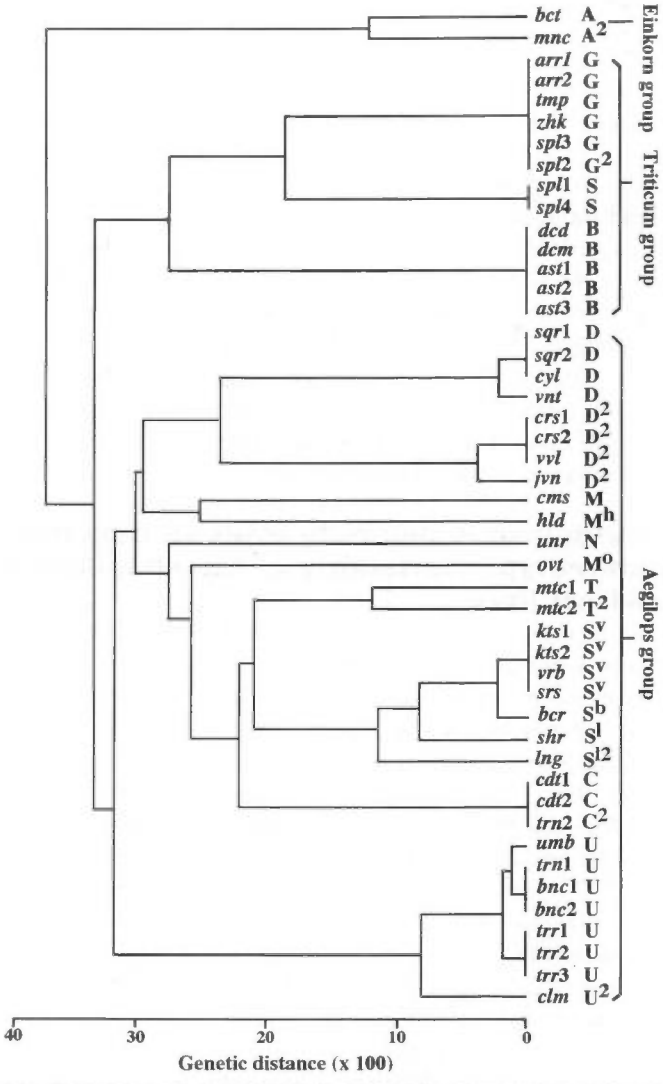

Fig. 2. Phylogenetic tree of the mitochondrial genomes of 47 *Triticum/Aegilops* accessions, constructed from the genetic distances by the UPGMA method (Wang *et al.* 1999).

within-gene regions, the variability of the mitochondrial genomes (ca. 3 x 10^{-3}) was only half that of the chloroplast genomes (ca. 8 x 10^{-3}). This difference reflects the higher evolutionary stability of mitochondrial genes as compared with chloroplast genes. The variability of the within-genes and their flanking regions inclusively, however, was four times higher in the mitochondrial (ca. 6 x 10^{-2}) than in the chloroplast (ca. 1.5 x 10^{-2}) genomes. This difference indicates a higher rate of genome reorganization in the mitochondria. In both the chloroplast and mitochondrial genomes, the variability of the gene-flanking regions was much higher than that of the within-gene regions.

Table 1. Nucleotide diversity estimated for different genomes based on the average genetic distance

Source of DNAs compared	No. of comparisons made	Nucleotide diversity (x 100)	Reference
Chloroplast genome			
Within *Ae. speltoides*	378	0.60	Miyashita *et al.* 1994†
Between *Triticum/Aegilops* spp.	630	0.80	Wang *et al.* 1997†
Between *Triticum/Aegilops* spp.	630	1.50	Wang *et al.* unpubl.
Mitochondrial genome			
Between *Triticum/Aegilops* spp.	153	2.89	Terachi and Tsunewaki 1992
Between *Triticum/Aegilops* spp.	630	0.30	Wang *et al.* 1997†
Between *Triticum/Aegilops* spp.	630	6.30	Wang *et al.* 1999
Nuclear genome			
Einkorn vs. emmer wheat	84	6.84	Tsunewaki *et al.* 1993
Einkorn vs. timopheevi wheat	15	6.23	Tsunewaki *et al.* 1993
Einkorn vs. common wheat	33	7.32	Tsunewaki *et al.* 1993
Emmer vs. timopheevi wheat	768	4.82	Tsunewaki *et al.* 1993
Emmer vs. common wheat	308	2.26	Tsunewaki *et al.* 1993
Timopheevi vs. common wheat	55	4.64	Tsunewaki *et al.* 1993

† Southern hybridization of PCR-amplified DNA regions. All others were Southern hybridization of the total DNA.

Tsunewaki *et al.* (1993) also estimated the nucleotide diversity of the nuclear genome in various wheat groups and found similar magnitudes of diversity, 2.3 - 7.3 x 10^{-2} (Table 1). Because these values indicate the variability of both the within-gene and flanking regions, it is concluded that DNA diversity per nucleotide pair is almost the same among the three genomes: nuclear, chloroplast and mitochondrial. It should be mentioned, however, that the number of nucleotide pairs in a single nuclear genome is about 20 000 times that in both the chloroplast and mitochondrial genomes in wheat and its relatives.

Cytoplasmic diversity evaluated by the phenotypic variability of the alloplasmic wheat lines

Nuclear genes are transmitted to offspring from both parents, whereas organellar genes are transferred exclusively from the female parent in most plants, including wheat and its relatives. Because of these different modes of inheritance of nuclear and organellar genes, alloplasmic lines having the same nuclear genome and different cytoplasmic genomes (or plasmons) can be produced by repeated backcrosses as follows. When the F_1 hybrid between strain A (♀) and C (♂) and its offspring are backcrossed repeatedly with the pollen of C, a line is obtained which has the nuclear genome of C and organellar genomes of A that is called an alloplasmic line of C. In the same way, repeated backcrosses of the F_1 hybrid between strain B (♀) and C (♂) and its offspring with the pollen of the C parent gives an alloplasmic line of C that has the nuclear genome of C and organellar genomes of B. A comparison of the three lines (two alloplasmic and a euplasmic line of C strain) as to phenotypic characters, reveals the genetic differences between the plasmons of three strains, A, B and C.

Forty-six accessions of *Triticum/Aegilops* species were used as alloplasmon donors to common wheat (Table 2). These included 20 accessions of 14 diploid species, 18 accessions of 14 tetraploid species and eight accessions of six hexaploid species. Twelve common wheat genotypes representing *T. aestivum* (eight accessions), *T. compactum*, *T. sphaerococcum*, *T. spelta* and *T. macha* (single accession each for the last four species) were used as the recipients of all 46 alloplasmons. Most alloplasmic lines had reached the B_{10} or later backcross generation by 1993 (Tsunewaki 1993).

In 1993, 22 characters of all 552 alloplasmic (12 nuclear genotypes x 46 alloplasmons) and 12 euplasmic lines were investigated in four replications, and the results obtained analyzed statistically (Wang *et al.* unpubl.). Table 3 shows the average and range of variation for the euplasmic and alloplasmic lines for six of the 22 characters investigated. The average of the alloplasmic lines did not differ much from that of euplasmic lines for any of the characters, except seed fertility which was about half the value of the euplasmic lines. The range of variation was much greater in the alloplasmic than the euplasmic lines for all the characters. Four points deserve special mention; some alloplasmons caused great delay in heading, nearly complete inviability as indicated by very short height and small dry matter weight, complete sterility, and a lowered seed germination rate caused by premature sprouting.

Phenotypic variation in seven characters of common wheat induced by 46 alloplasmons was compared with that of the 12 nuclear genotypes in terms of variance (Table 4). In 1977, Tsunewaki *et al.* (1978) had made a similar investigation of 288 alloplasmic lines (the same 12 nuclear genotypes x 24 alloplasmons, but the backcross generations of most lines did not reach B_{10} at that time) and the same 12 euplasmic lines. The findings of that investigation also are shown in Table 4.

Table 2. *Triticum* and *Aegilops* species (abbreviated species name, and number of accessions in parentheses) used as cytoplasm donors to alloplasmic lines of common wheat

Diploid	Tetraploid	Hexaploid
Triticum		
T. boeoticum (btc, 1)	*T. dicoccoides* (dcd, 1)	*T. aestivum* (ast, 2)
T. monococcum (mnc, 1)	*T. dicoccum* (dcm, 1)	*T. zhukovskyi* (zhk, 1)
	T. araraticum (arr, 2)	
	T. timopheevi (tmp, 1)	
Aegilops Sect. Polyeides		
Ae. umbellulata (umb, 1)	*Ae. triuncialis* (trn, 2)	*Ae. triaristata* (trr6x, 2)
	Ae. biuncialis (bnc, 2)	
	Ae. columnaris (clm, 1)	
	Ae. ovata (ovt, 1)	
	Ae. triaristata (trr4x, 1)	
	Ae. kotschyi (kts, 2)	
	Ae. variabilis (vrb, 1)	
Aegilops Sect. Cylindropyrum		
Ae. caudata (cdt, 2)	*Ae. cylindrica* (cyl, 1)	
Aegilops Sect. Comopyrum		
Ae. comosa (cms, 1)		
Ae. heldreichii (hld, 1)		
Ae. uniaristata (unr, 1)		
Aegilops Sect. Vertebrata		
Ae. squarrosa (sqr, 2)	*Ae. ventricosa* (vnt, 1)	*Ae. crassa* (crs6x, 1)
	Ae. crassa (crs4x, 1)	*Ae. juvenalis* (jvn, 1)
		Ae. vavilovii (vvl, 1)
Aegilops Sect. Sitopsis		
Ae. speltoides (spl, 4)		
Ae. bicornis (bcr, 1)		
Ae. sharonensis (shr, 1)		
Ae. longissima (lng, 1)		
Ae. searsii (srs, 1)		
Aegilops Sect. Amblyopyrum		
Ae. mutica (mtc, 2)		

Table 3. Averages and ranges of variation of six characters observed in 12 euplasmic and 552 alloplasmic lines of common wheat. Results of a 1993 investigation (Wang *et al.* unpubl.)

Character	Type of lines	Average	Range of variation
Heading date (no. of days; April 20 = 0)	Euplasmic	27.0	5.8 - 36.8
	Alloplasmic	31.3	3.8 - 66.0
Plant height (cm)	Euplasmic	126.9	88.3 - 54.1
	Alloplasmic	115.6	5.8 -157.3
No. of spikelets per ear	Euplasmic	21.9	15.2 - 24.6
	Alloplasmic	21.9	2.3 - 29.3
Dry weight of 2 culms, including grains (g)	Euplasmic	6.4	3.7 - 12.4
	Alloplasmic	5.0	0.2 - 12.1
Selfed-seed fertility (%)	Euplasmic	93.2	82.4 - 98.3
	Alloplasmic	48.2	0.0 - 98.5
Seed germination rate (%)	Euplasmic	92.2	85.5 - 97.3
	Alloplasmic	90.6	23.1 - 100.0

Analysis of variance was used to estimate the variance attributable to the nucleus, S_N^2; to the cytoplasm, S_C^2; and to their interaction, S_I^2. The sum of S_C^2 and S_I^2 gives the variance that is directly and indirectly attributable to the cytoplasm. Its ratio to the variance attributable to the nucleus was similar in the investigations of 1977 and 1993 (Table 4). Of the seven characters shown in this table, ear length was influenced least and selfed seed fertility was affected most by alien plasmons. Findings of the 1993 investigation indicated that the variance attributable to the 47 plasmons of *Triticum/Aegilops* species was only 5-31% of the variance attributable to the 12 nuclear genotypes of common wheat in all the characters, except selfed seed fertility. As for this character, the sum of the variances attributable to the cytoplasm was almost three-fold the variance for the nucleus. This character, mainly determined by male fertility (Tsunewaki *et al.* 1996), was extremely sensitive to alien cytoplasms.

Strategy for the conservation of cytoplasmic diversity

Two methods are available for the conservation of cytoplasmic diversity: conservation of living materials and preservation in the form of extracted ct and mtDNA. Conservation in the form of living materials is indispensable for the use of alloplasmons in wheat breeding until the development of an efficient method of transferring alien organellar genes. Meanwhile, both the alloplasmic wheat lines already created and potential alloplasmon donors to wheat need to be maintained. The method for their maintenance is the same as that for ordinary stocks and, therefore, is not discussed further in this article.

Conservation of cytoplasmic diversity can be achieved efficiently in the form of extracted DNA owing to the appropriate size of the chloroplast and mitochondrial genomes, although the use of preserved organellar DNA for actual wheat breeding is not practical at present and may be impossible. However, extracted organellar DNA can be used to screen for organellar DNA variation and should serve as the source of useful organellar genes in future.

I propose to achieve organellar DNA preservation and to use the preserved DNA by three steps. The first is the separate extraction and preservation of intact ct and mtDNA from representative variations of all species. To do this, intact chloroplasts and mitochondria must be fractionated separately, then the intact ct or mtDNA is extracted and stored under deep refrigeration for an indefinite period. The minimum fresh material required per accession would be ca. 100 g green leaves for ctDNA and ca. 500 g etiolated seedlings for mt DNA. MtDNA preparation

Table 4. Genetic variance attributable to the nucleus (12 wheat nuclear genotypes), cytoplasm (24 or 47 cytoplasms of *Triticum/Aegilops* species) and their interaction: estimated for the same seven characters investigated in the 1977 and 1993 studies (Tsunewaki *et al.* 1978; Wang *et al.* unpubl.)

Character	S_N^2	S_C^2	S_I^2	$S_C^2+S_I^2 / S_N^2$
1977 investigation (288 lines = 12 nuclei x 24 cytoplasms)				
Ear length	14.61	0.302	0.523	0.06
Heading date	97.03	10.36	3.69	0.14
Ear number	53.68	15.41	8.37	0.44
Plant height	403.9	86.2	45.7	0.33
Spikelet no./ear	15.69	1.52	1.34	0.18
Dry matter weight	593.2	447.8	168.0	1.04
Selfedseed fertility	102.3	935.1	478.9	13.82
1993 investigation (564 lines = 12 nuclei x 47 cytoplasms)				
Ear length	2088	99.2	6.3	0.05
Heading date	16467	1776	58	0.11
Ear number	17581	1827	158	0.11
Plant height	62423	14488	634	0.24
Spikelet no./ear	1536	372	16	0.25
Dry matter weight	216.0	64.6	1.8	0.31
Selfedseed fertility	23369	60203	2390	2.68

might prove difficult for some materials because of the large amount of etiolated material required per accession. These DNA samples would be used in the subsequent steps of conservation.

The second step is the preparation of complete sets of the genome sequence-based contig clones of the wheat ct and mtDNA. To prepare these, complete sequencing of wheat chloroplast and mitochondrial genomes is required. Through the collaboration of seven laboratories in Japan, including mine, we completed the sequencing of the entire wheat chloroplast genome in 1998 using the common wheat 'Chinese Spring', (Fig. 3, Ogihara *et al.* 1998). The 'Chinese Spring' chloroplast genome was found to have 134,523 bp and carry ca. 142 genes, of which 24 were duplicated in the inverted repeats.

The wheat chloroplast genome can be divided into 30 sections of ca. 4.5 kb each and pairs of PCR primers are designed for the individual sections (ref. Fig. 3). Using these primer pairs, DNA fragments corresponding to all the sections can be separately amplified. The 30 PCR-amplified DNA clones obtained cover the entire wheat chloroplast genome. This contig set of wheat ctDNA clones is used as the probe for detecting RFLP variation in the entire chloroplast genomes of various accessions; not only those of the *Triticum / Aegilops* species but also those of such related genera as *Secale*, *Agropyron* and *Hordeum*. For this purpose, they should be stocked at an appropriate institution that will be responsible for their distribution to other researchers.

As for the mitochondrial genome, complete sequencing of the wheat mitochondrial genome has yet to be done. It is important that a research group carries out this work as until this has been done the second and third steps in mitochondrial genome preservation for the *Triticum / Aegilops* complex cannot be carried out.

The third step is the preparation of PCR-amplified clones for all chloroplast and mitochondrial genes of wheat, again based on the complete sequences of ct and mtDNA. Pairs of PCR primers must be designed for the individual genes and their coding regions separately amplified. The PCR-amplified ct and mtDNA clones that correspond to all the wheat organellar genes will serve as probes for the isolation of organellar genes or to detect variation in their transcripts in the various accessions of wheat and its relatives. These wheat organellar gene clones should be preserved in an appropriate institution and provided to other researchers upon request.

Although the complete nucleotide sequence of wheat mtDNA is not yet available, some wheat mitochondrial genes have been isolated and sequenced (Bonen *et al.* 1991; Bonnard and Grienenberger 1995; Vaitilingom *et al.* 1998). For those genes PCR-amplified DNA for their coding regions should be prepared for the purpose described above.

Use of cytoplasmic variability in wheat breeding

Systematic investigations of genetic effects on various wheat characters of the alloplasmons present in the *Triticum / Aegilops* complex suggested the following possibilities for the practical use of alloplasmons in improving wheat (Tsunewaki 1988): (i) male sterility for hybrid wheat breeding, (ii) delayed heading for breeding wheat with a prolonged growth period, (iii) parthenogenesis for haploid breeding, and (iv) enlarging cytoplasmic variability in wheat by the creation of new genotypes adaptive for fertile alloplasmons. The third point has been treated in detail elsewhere by Tsunewaki and Mukai (1990), therefore, I discuss the other points briefly.

Use of cytoplasmic male sterility for hybrid wheat

Hybrid wheat breeding using the *T. timopheevi* cytoplasm has yet to be realized in spite of the intensive efforts made in many institutions, including private seed companies, for nearly 40 years (Wilson 1968). Subsequently, use of chemical hybridizing agents for male sterility induction has been tested (e.g. Brears *et al.* 1987). In 1997, an F_1 hybrid 'Jinhua 1' was released to farmers in China (Zhang 1998). I pointed out three new sources of male sterile cytoplasm in 1988. These are a D^2 type cytoplasm of *Ae. crassa*, an S^v type cytoplasm of *Ae. kotschyi* and an N type cytoplasm of *Ae. uniaristata*, which should be considered as substitutes for the G type cytoplasm of *T. timopheevi* (Tsunewaki 1988). Since then some progress has been made toward the use of D^2, S^v and N type cytoplasms.

Murai made a series of investigations on the use of the D^2 type cytoplasm (Murai *et al.* 1991). In certain Japanese cultivars, such as 'Norin 26', this type of cytoplasm induced complete pistillody of all three stamens, resulting in male sterility when the daylength at the time of floral initiation was longer than 15 h. The F_1 seeds were produced at a high latitude on Hokkaido Island between (*crassa*)-'Norin 26', which became male sterile there, and a

few other cultivars as pollinator. These seeds were sown in an experimental field at a lower latitude on Honshu Island and the agronomic characters of the F_1 hybrids investigated. Some F_1s gave better yields than the check cultivar, but this work was discontinued when Murai moved to another institute in 1993.

As for the use of the S^v type cytoplasm, there are two possibilities based on findings that it caused complete male sterility in two spelt-type wheats and deep male sterility in a wheat line carrying a 1BL-1RS translocation chromosome that consists of the long arm of chromosome 1B of wheat and the short arm of the 1R chromosome of rye. I attempted to utilize the complete male sterility induced in a spelt wheat because it has the normal complement of wheat chromosomes. Unfortunately, all my attempts to cotransfer the male sterile gene(s) of the spelt wheat and the S^v type cytoplasm of *Ae. kotschyi* to ordinary wheat cultivars failed. The reason for this failure is still not clear.

The other approach was to utilize the male sterility induced by the interaction between the S^v type cytoplasm and a 1BL-1RS chromosome. This approach was adopted by He Peiru of the Northwestern Agricultural University, Shaanxi Province, China, as well as by two groups in Japan. He successfully bred a spring-type F_1 cultivar 'Xinong 901' that outyielded by 16-28% the local check cultivar in a 3-year trial. It has been under cultivation by farmers since 1995 (He, pers. comm.). The two Japanese groups have yet to succeed in producing any commercial F_1 hybrid (Ikeguchi *et al.* 1994; Ushiyama *et al.* 1996).

As for the use of the N type cytoplasm, like the S^v type cytoplasms, this induces nearly complete male sterility in interaction with a 1BL-1RS chromosome. Except for the work of Zhu *et al.* (unpubl.) at the Northwestern Agricultural University, China, little has been done so far with this cytoplasm. They showed that the N and S^v type cytoplasms are almost the same as a source of cytoplasmic male sterility. Their findings indicate that fertility restoration is slightly more difficult for the N type as compared to the S^vtype cytoplasm, but seed quality is a little better. Further investigation of the usefulness of this cytoplasm as the source of cytoplasmic male sterility for hybrid wheat breeding would be worthwhile.

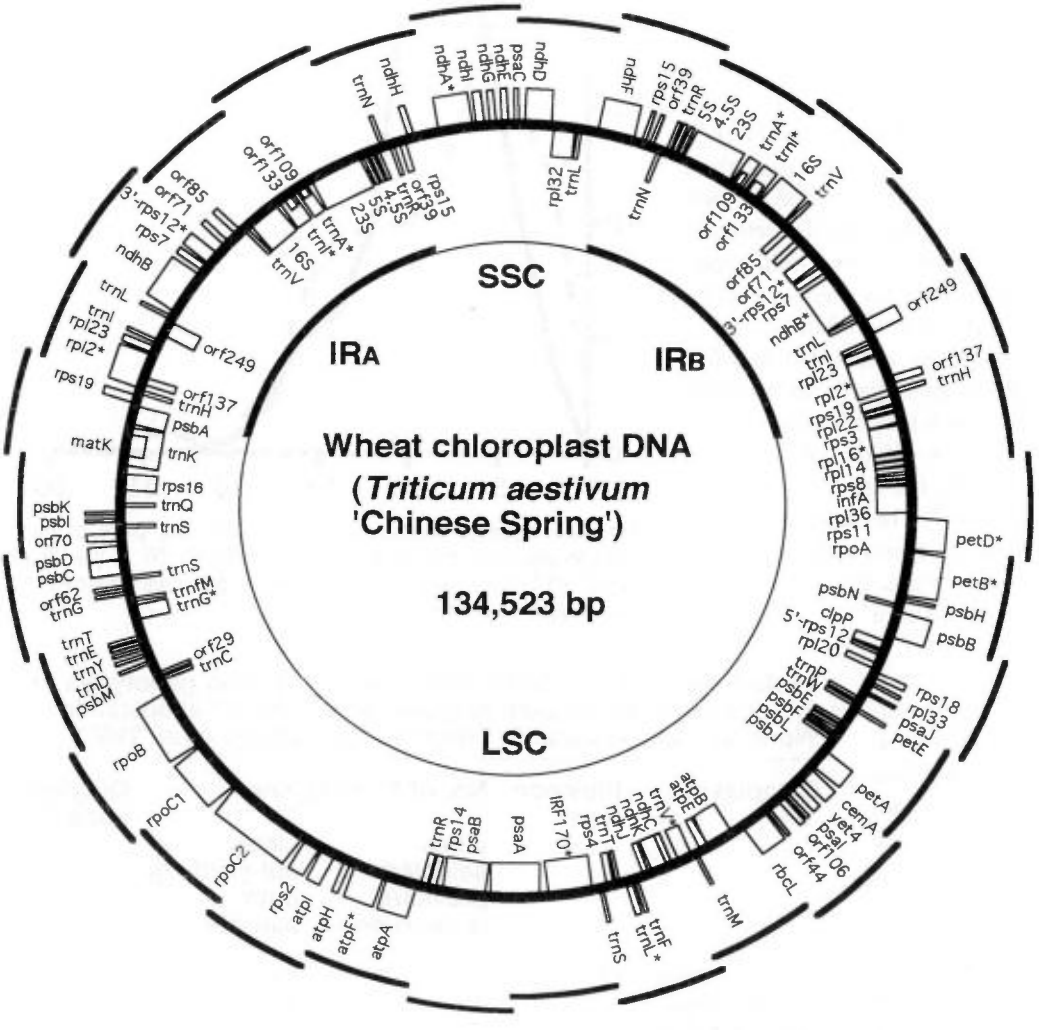

Fig. 3. A complete genetic map of the chloroplast genome of *Triticum aestivum* 'Chinese Spring' (Ogihara *et al.* 1998), and a postulated set of 30 contig clones of its ctDNA in the outer circle, each ca. 4.5 kb in size.

Delayed heading for breeding wheat lines with prolonged growth periods

At high latitudes, where summers are short and the day length very long, ordinary photoperiod-sensitive wheat cultivars do not have a growth period sufficient to produce high yields. One such area is found in Hokkaido Island, Japan. A possible solution is to breed cultivars with prolonged basic vegetative growth periods, but no nuclear genotypes suitable for this purpose have been found. The M° plasmon of *Ae. ovata* caused severe male sterility but no further serious effects other than delayed growth (Tsunewaki 1988). In addition, a dominant fertility-restoring gene (*Rfc1*) to this cytoplasm is available that originated from *Ae. caudata* and was translocated onto a wheat chromosome 1D (Kihara 1959;

Tahir and Tsunewaki 1971). Nagasawa and Tsunewaki (1983) introduced both the M° plasmon and *Rfc1* gene into four Hokkaido winter-wheat cultivars by repeated backcrosses. Fertility restoration showed clear monofactorial inheritance during the backcrosses. In the B_7S_3 generation, the lines fixed for the *Rfc1* gene were selected and tested for two years at the Kitami Agricultural Experiment Station, Hokkaido (Makita *et al.* unpubl.). As expected, both heading and maturity were delayed from 10 days to two weeks, but the grain yield per unit area was reduced 10-30%, mainly due to the reduction (20-40%) in the 1000 kernel weight. There was no difference in grain and flour appearance or noodle quality between the euplasmic and alloplasmic lines. These results indicate that a delay in heading is not necessarily favourable as expected for grain filling. For further improvement of these lines, the unfavourable genes of *Ae. caudata* that are linked to *Rfc1* must be eliminated by disrupting their linkages.

Enlarging cytoplasmic variability in wheat by creation of new genotypes adaptive to fertile alloplasmons

Cytoplasmic variability in a given crop species may be desirable because there is always the danger of new races of pathogens appearing that are virulent to cultivars with a particular type of cytoplasm, as happened in the late 1960s to hybrid corn with the T type male sterile cytoplasm. It is reasonable, however, to assume that for all crops their present-day nuclear genotypes are better adapted to their own cytoplasms than to any alien cytoplasms. Therefore, it is best to create new nuclear genotypes that collaborate better with fertile alloplasmons than with their own cytoplasms in order to broaden cytoplasmic variability.

To achieve this goal in wheat, we attempted to select new genotypes with high-yielding potential from the progenies of crosses between four fully fertile alloplasmic lines of 'Norin 26' and a euplasmic line of the most widely adapted cultivar 'Norin 61' among the Japanese common wheat cultivars (Tsunewaki *et al.* 1985; Yonezawa *et al.* 1987). The alloplasmons used were the D type cytoplasm of *Ae. squarrosa*, D^2 type cytoplasm of *Ae. crassa*, S type cytoplasm of *Ae. speltoides* and S^v type cytoplasm of *Ae. kotschyi*.

Segregation patterns in four pairs of the alloplasmic and euplasmic F_2 populations were investigated for several yield-relating characters. Figure 4 shows the frequency distributions of yield per plant in a pair of alloplasmic and euplasmic F_2 populations of reciprocal crosses; (*kotschyi*)-'Norin 26' x euplasmic 'Norin 61' and euplasmic 'Norin 61' x (*kotschyi*)-'Norin 26'. The euplasmic F_2 population from the latter cross had a nearly normal distribution, whereas the alloplasmic F_2 population with the S^v cytoplasm showed an excess of low-yielding F_2s as well as some transgressive high-yielding F_2s (Tsunewaki *et al.* 1985).

In the F_2 and F_3 generations of all the crosses, a limited number of individuals that exceeded both parents in one or another yield-related character were selected, and their F_3 or F_4 lines were selected based on their yields. All the lines with the D^2 cytoplasm were selected out at this stage. The number and percentage of the F_4 or F_5 lines that outyielded both parents are shown in Table 5 (Yonezawa *et al.* 1987). Clearly, the percentage of lines with yields higher than those of both parents was the same or greater in all three alloplasmic populations than in the euplasmic control, with the exception of the D^2 cytoplasm.

Acknowledgements

My attendance at the 18th International Congress of Genetics was made possible by generous financial support from

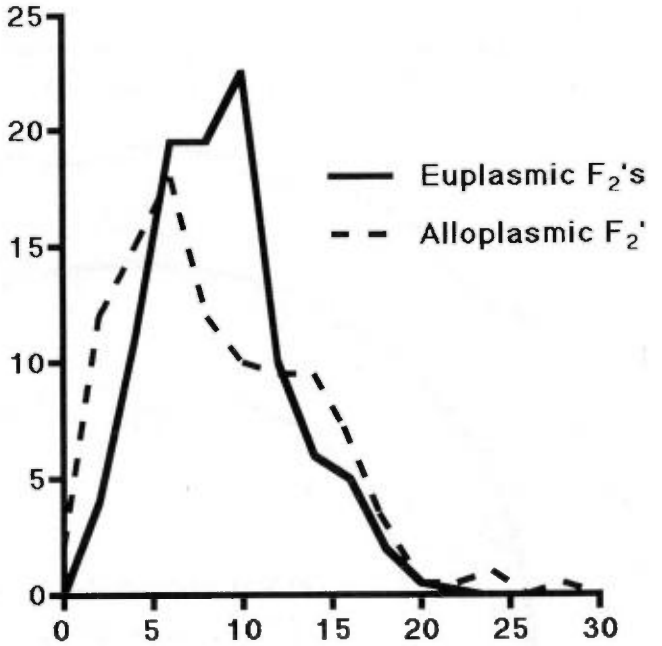

Fig. 4. Frequency distribution for the yield per plant of the F_2 populations from the cross, (*kotschyi*)-'Norin 26' x (*aestivum*)-'Norin 61' (solid line), and its reciprocal cross (dashed line) (Tsunewaki *et al.* 1985).

Table 5. Selection of fertile alloplasmon-adapted genotypes in the progenies of crosses between euplasmic or alloplasmic 'Norin 26' and euplasmic 'Norin 61' (Yonezawa *et al.* 1987)

Cytoplasm	Plasmon type	No. of F_4 or F_5 lines		Outyielded lines (%)
		Lines selected in previous generations	Lines out-yielding both parents	
Ae. squarrosa	D	50	18	36
Ae. kotschyi	S^v	48	31	65
Ae. speltoides	S	12	5	42
T. aestivum (euplasmic)	B	24	9	38

IPGRI, Rome, to which I express my sincerest appreciation. The findings presented in this article have been cited from published and unpublished works of my colleagues and co-workers, all of whom I thank for their kind collaboration.

References

Bonen, L., Y. Chapdelaine and P. Garber. 1991. Wheat mitochondrial protein-coding genes and their expression. Pp. 140-148 *in* Nuclear and Organellar Genomes of Wheat Species (T. Sasakuma and T. Kinoshita, eds.). Kihara Memorial Foundation, Yokohama.

Bonnard, G. and J.M. Grienenberger. 1995. A gene proposed to encode a transmembrane domain of an ABC transporter is expressed in wheat mitochondria. Mol. Gen. Genet. 246:91-99.

Bowman, C. M., G. Bonnard and T.A. Dyer. 1983. Chloroplast DNA variation between species of *Triticum* and *Aegilops*. Location of the variation on the chloroplast genome and its relevance to the inheritance and classification of the cytoplasm. Theor. Appl. Genet. 65:247-262.

Brears, T., A.G. Hydon and J. Bingham. 1987. An assessment of the feasibility of producing F_1 and F_2 hybrids for the UK. Pp. 1057-1062 *in* Proc. 7th Int. Wheat Genet. Symp. Cambridge, Vol. 2 (T.E. Miller and R.M.D. Koebner, eds.). Institute for Plant Science Research, Cambridge.

Breiman, A. 1987. Mitochondrial DNA diversity in the genera of *Triticum* and *Aegilops* revealed by Southern blot hybridization. Theor. Appl. Genet. 73:563-570.

Graur, D., M. Bogher and A. Breiman. 1989. Restriction endonuclease profiles of mitochondrial DNA and the origin of the B genome of bread wheat, *Triticum aestivum*. Heredity 62:335-342.

Ikeguchi, S., A. Hasegawa, Y. Oyamada, K. Toriyama and K. Tsunewaki. 1994. Basic studies on hybrid wheat breeding utilizing *Aegilops kotschyi* cytoplasm. Pp. 77-83 *in* Low Temperature Physiology and Breeding of Northern Crops (H. Saka, J. Abe, H. Yamaguchi and T. Kuwabara, eds.) Hokkaido Nat. Agr. Exp. Stat. Sapporo.

Kihara, H. 1959. Fertility and morphological variation in the substitution and restoration backcrosses of the hybrids *Triticum vulgare* x *Aegilops caudata*. Proc. 10th Int. Congr. Genet. Montreal 1:142-171.

Miyashita, N.T., N. Mori and K. Tsunewaki. 1994. Molecular variation in chloroplast DNA regions in ancestral species of wheat. Genetics 137:883-889.

Mori, N., T. Terachi and K. Tsunewaki. 1988. Organellar genome differentiation in wild tetraploid wheats, *Triticum dicoccoides* and *T. araraticum*. Pp. 109-114 *in* Proc. 7th Int. Wheat Genet. Symp. Cambridge, Vol. 1 (T.E. Miller and R.M.D. Koebner, eds.). Institute for Plant Science Research, Cambridge.

Murai, K., H. Hirohara and K. Tsunewaki. 1991. Two-line system for hybrid wheat breeding using photoperiod-sensitive cytoplasmic male sterility. Pp. 294-299 *in* Nuclear and Organellar Genomes of Wheat Species (T. Sasakuma and T. Kinoshita, eds.). Kihara Memorial Foundation, Yokohama.

Nagasawa, A. and K. Tsunewaki. 1983. Breeding and characterization of late type lines of Hokkaido common wheat cultivars with *Aegilops ovata* cytoplasm. Jpn. J. Breed. 33 (Suppl. 1):154-155, (in Japanese).

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Ogihara, Y. and K. Tsunewaki. 1982. Molecular basis of the genetic diversity of the cytoplasm in *Triticum* and *Aegilops*. I. Diversity of the chloroplast genome and its lineage revealed by the restriction pattern of ctDNA. Jpn. J. Genet. 57:371-396.

Ogihara, Y. and K. Tsunewaki. 1988. Diversity and evolution of chloroplast DNA in *Triticum* and *Aegilops* as revealed by restriction fragment analysis. Theor. Appl. Genet. 76:321-332.

Ogihara, Y., H. Shimizu, K. Hasegawa, H. Tsujimoto and T. Sasakuma. 1994. Chromosome assignment of four photosynthesis-related genes and their variability in wheat species. Theor. Appl. Genet. 88:383-394.

Ogihara, Y., K. Isono, T. Kojima, H. Tsudsuki, A. Endo, R. Murai, K. Murai, M. Hanaoka, T. Shiina, T. Terachi, S. Utsugi, M. Murata, N. Mori, S. Takumi, K. Ikeno, T. Gojobori, Y. Matsuoka, Y. Ohnishi, H. Tajiri and K. Tsunewaki. 1998. Complete sequence of wheat chloroplast DNA. Pp. 38-40 *in* Proc. 9th Int. Wheat Genet. Symp. Saskatoon, Vol. 3 (A.E. Slinkard, ed.). University Extension Press, Saskatoon, Saskatchewan, Canada.

Tahir, Ch. M. and K. Tsunewaki. 1971. Monosomic analysis of fertility-restoring genes in *Triticum aestivum* strain P168. Can. J. Genet. Cytol. 13:14-19.

Terachi, T. and K. Tsunewaki. 1986. The molecular basis of genetic diversity among cytoplasms of *Triticum* and *Aegilops*. 5. Mitochondrial genome diversity among *Aegilops* species having identical chloroplast genomes. Theor. Appl. Genet. 73:175-181.

Terachi, T. and K. Tsunewaki. 1992. The molecular basis of genetic diversity among cytoplasms of *Triticum* and *Aegilops*. VIII. Mitochondrial RFLP analyses using cloned genes as probes. Mol. Biol. Evol. 9:917-931.

Tsunewaki, K. 1988. Cytoplasmic variation in *Triticum* and *Aegilops*. Pp. 53-62 *in* Proc. 7th Int. Wheat Genet. Symp. Cambridge, Vol. 1 (T.E. Miller and R.M.D. Koebner, eds.). Institute of Plant Science Research, Cambridge.

Tsunewaki, K. 1993. Genome-plasmon interactions in wheat. Jpn. J. Genet. 68:1-34.

Tsunewaki, K. and Y. Mukai. 1990. Wheat haploids through the Salmon method. Pp. 460-478 *in* Biotechnology in Agriculture and Forestry, Vol. 13 (Y.P.S. Bajaj, ed.). Springer-Verlag, Berlin.

Tsunewaki, K., Y. Mukai and T.R. Endo. 1978. On the descent of the cytoplasms of polyploid species in *Triticum* and *Aegilops*. Pp. 261-272 *in* Proc. 5th Int. Wheat Genet. Symp. New Delhi, Vol. 1 (S. Ramanujam, ed.) Indian Society of Genetics Plant Breeding, New Delhi.

Tsunewaki, K., P. Spetsov and K. Yonezawa. 1985. Increasing genetic variability in common wheat by utilization of alien cytoplasms – Cytoplasmic effects on the performance and interplant variability of the F_1 and F_2 generations of the cross, *Triticum aestivum* cv. Norin 26 x cv. Norin 61. Jpn. J. Breed. 35:398-412.

Tsunewaki, K., Y.-G. Liu, S. Takumi, N. Mori, H. Nakamura, T. Terachi, U.J. Siregar and N.T. Miyashita. 1993. Use of RFLP analyses for wheat germplasm evaluation. Pp. 17-30 *in* Biodiversity and Wheat Improvement (A.B. Damania, ed.). John Wiley & Sons, Chichester.

Tsunewaki, K., G.-Z. Wang and Y. Matsuoka. 1996. Plasmon analysis of *Triticum* (wheat) and *Aegilops*. 1. Production of alloplasmic common wheats and their fertilities. Gene Genet. Syst. 71:293-311.

Ushiyama, T., K. Toriyama, K. Tsunewaki, S. Nonaka and T. Shimada. 1996. Seed production ability of male-sterile lines of common wheat induced by the interaction between an S^v type cytoplasm and a 1BL-1RS chromosome. Breed. Sci. 46:303-306.

Vaitilingom, M., M. Stupar, J.M. Grienenberger and J.M. Gualberto. 1998. A gene coding for an RPS2 protein is present in the mitochondrial genome of several cereals, but not in dicotyledons. Mol. Gen. Genet. 258:530-537.

Vedel, F., F. Quetier, F. Dosba and G. Doussinault. 1978. Study of wheat phylogeny by *Eco*RI analysis of chloroplastic and mitochondrial DNAs. Plant Sci. Lett. 13:97-102.

Vedel, F., F. Quetier, Y. Cauderon, F. Dosba and G. Doussinault. 1981. Studies on maternal inheritance in polyploid wheats with cytoplasmic DNAs as genetic markers. Theor. Appl. Genet. 59:239-245.

Wang, G.-Z., N. T. Miyashita and K. Tsunewaki. 1997. Plasmon analyses of *Triticum* (wheat) and *Aegilops*: PCR-single strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs. Proc. Natl. Acad. Sci. USA 94:14570-14577.

Wang, G.-Z., Y. Matsuoka and K. Tsunewaki. 1999. Evolutionary features of chondriome divergence in *Triticum* (wheat) and *Aegilops* shown by RFLP analysis of mitochondrial DNAs. Theor. Appl. Genet. (in press).

Wilson, J.A. 1968. Hybrid wheat developments with *Triticum timopheevi* Zhuk. derivatives. Pp. 423-430 *in* Proc. 3rd Int. Wheat Genet. Symp. (K.W. Finlay and K. W. Shepherd, eds.). Aust. Academy of Science.

Yonezawa, K., K. Mamiya, P. Spetsov and K. Tsunewaki. 1987. Increasing genetic variability in common wheat by utilizing alien cytoplasms – Characteristics of nearly fixed alloplasmic lines selected in the cross, alloplasmic Norin 26 x Norin 61. Japan. J. Breed. 37:159-169.

Zhang, A. 1998. Hybrid wheat breeding in the People's Republic of China. Pp. 381-383 *in* Proc. 9th Int. Wheat Genet. Symp Saskatoon, Vol. 2. (A.E. Slinkard, ed.). University Extension Press, Saskatoon, Saskatchewan, Canada.

Evaluation of genetic diversity in Chinese soyabean germplasm by AFLP

S.L. Zhu[1], L.M. Monti[2], A. Avitabile[3] and R. Rao[3]✉

[1] Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China
[2] Institute for Vegetables and Flour Breeding, CNR, Portici, Italy
[3] Department of Agronomy and Plant Genetics, University of Naples, Portici, Italy. Tel: +39-81-7885445; Fax: +39-81-7753579; Email: Rao@unina.it

Summary

Evaluation of genetic diversity in Chinese soyabean germplasm by AFLP

Amplified Fragment Length Polymorphism (AFLP) was used to evaluate genetic diversity among 61 Chinese soyabean accessions, including 41 cultivated (*Glycine max*) and 20 wild soyabean (*Glycine soja*), from four main cultivation regions. Eight AFLP primer pairs detected a total of 569 fragments with an average of 71 fragments per primer per accession. A large number of fragments (273), representing 48% of total fragments, were polymorphic, 172 in *G. max* and 249 in *G. soia*. The primer combinations differ widely in their ability to detect polymorphism. The range of polymorphic bands detected by individual primer combinations was from 16 to 51. The E-aag/M-ctc primer pair is particularly informative, detecting 51 polymorphic fragments, which differentiated all 61 accessions. As expected, the genetic diversity present in wild accessions is higher than that present in cultivated accessions, confirming that wild soyabean preserves a larger genetic variation. Cluster analysis based on an unweighted pair-group method using arithmetic average (UPGMA) clearly separated the two species as does present taxonomy and, to some extent, the different sowing types of *G. max*, particularly the spring-sowing accessions, independent of their geographic origin.

Key words: AFLP, China, genetic diversity, *Glycine max*, polymorphism, soyabean

Résumé

Evaluation de la diversité génétique dans un germoplasme de soja chinois par AFLP

La technique d'AFLP (Amplified Fragment Length Polymorphism) est utilisée pour évaluer la diversité génétique dans 61 accessions de soja chinois, y compris 41 sojas cultivés (*Glycine max*) et 20 sojas de type sauvage (*Glycine soja*), provenant de quatre régions de culture principales. Huit paires d'amorces AFLP permettent de détecter un total de 569 fragments, soit une moyenne de 71 fragments par amorce et par accession. Un grand nombre de fragments (273), représentant 48 % du total des fragments, sont polymorphes, 172 chez *G. max* et 249 chez *G. soia*. Il existe une très grande différence dans l'aptitude des combinaisons d'amorces à détecter du polymorphisme. La gamme des bandes polymorphes détectées par des combinaisons d'amorces particulières va de 16 à 51. La paire d'amorces E-aag/M-ctc apporte notamment beaucoup d'informations, mettant en évidence 51 fragments polymorphes, qui permettent de différencier les 61 accessions. Comme on pouvait s'y attendre, la diversité génétique observée dans les accessions sauvages est plus grande que celle présente dans les accessions cultivées, ce qui confirme que le soja de type sauvage conserve une plus grande variation génétique. Une analyse de groupes (cluster analysis), basée sur la méthode des groupes de paires non pondérées avec moyenne arithmétique (Unweighted Pair Group Method with Arithmetic Average, UPGMA), distingue nettement les deux espèces à l'instar de la taxonomie actuelle et, dans une certaine mesure, les différents types de cultivars de *G. max*, en particulier les accessions de semis de printemps, quelle que soit leur origine géographique.

Resumen

Evaluación de la diversidad genética en el germoplasma de la soja china mediante el AFLP

El AFLP (Polimorfismo de la longitud de los fragmentos amplificados) se ha utilizado para evaluar la diversidad genética de 61 accesiones de soja china, 41 cultivadas (*Glycine max*) y 20 silvestres (*Glycine soja*), de cuatro regiones principales de cultivo. Ocho pares iniciadores de AFLP individuaron un total de 569 fragmentos con una media de 71 fragmentos por iniciador en cada una de las accesiones. El 48% de los fragmentos (273), eran polimorfos, 172 en *G. max* y 249 en *G. soja*. Las combinaciones de iniciadores se diferenciaban mucho debido a la capacidad de individuación del poliformismo. La serie de bandas poliformias individuadas por las combinaciones de iniciadores de individuos se extendía de 16 a 51. La pareja de iniciadores E-aag/M-ctc es particularmente interesante ya que ha individuado 51 fragmentos poliformios que representan la característica de diferenciación de las 61 accesiones. Como era de esperar, la diversidad genética presente en las accesiones silvestres es más alta que la de las accesiones cultivadas confirmando, de esta manera, que la soja sivestre presenta una variación genética mayor. Algunos análisis conjuntos, basados en un método de grupos de pares sin peso en que se utiliza una media aritmética (UPGMA), separaban claramente las dos especies como sucede con la taxonomía actual y, hasta cierto punto, los diferentes tipos de siembra de *G. max*, especialmente las accesiones sembradas en primavera, independientemente de su origen geográfico.

Introduction

China is recognized as being the native country of soyabean (*Glycine max*), one of the most important oil and protein crops, which has been cultivated in China for more than 5000 years. *Glycine soja* is believed to be the ancestor of soyabean and is cross-compatible with it; therefore, it is considered to be a natural source of genetic diversity for soyabean improvement (Kiang *et al.* 1992). Although during the past decade, nearly 30000 accessions have been collected and documented, few investigations have been carried out to evaluate the genetic diversity of Chinese germplasm, in spite of the good characterization carried out on other genepools of the species, such as germplasm found in the USA. It is worth emphasizing that assessing genetic diversity is one of the key tools of crop improvement and germplasm conservation.

Table 1. Soyabean accessions used for AFLP analysis

Code	Accession	Name	Species	Origin[§]	Sowing type
1	D4429	Tai xing hei duo	*G. max*	III	SPR
2	D6027	Wuqingduo	*G. max*	III	SPR
3	D5979	Wu yue bai	*G. max*	III	SPR
4	D21	Heinong 26	*G. max*[†]	I	SPR
5	D13839	Zhe chun No 1	*G. max*[†]	III	SPR
6	D13840	Zhe chun No 2	*G. max*[†]	III	SPR
7	D14510	Xiangchun No 10	*G. max*[†]	III	SPR
8	D13842	Qingbai duo	*G. max*	IV	SPR
9	D13832	Liuyue duo	*G. max*	IV	SPR
10	D13836	Chengdongmeiduo	*G. max*	IV	SPR
11	D261	Hailunduludue	*G. max*	I	SPR
12	D4918	Wanbei 4918	*G. max*[†]	II	SUM
13	D13882	Tiananduo	*G. max*	IV	SUM
14	D2643	Yanhuang No 1	*G. max*[†]	II	SUM
15	X531	Nannong87c38	*G. max*[†]	III	SUM
16	X534	Luhei duo No 2	*G. max*[†]	II	SUM
17	D2573	Qihuang 2	*G. max*[†]	II	SUM
18	D6332	Mao pen qing	*G. max*	IV	AUT
19	D6066	Cu huang duo	*G. max*	III	SUM
20	A6007	Xiang qiu No 1	*G. max*[†]	IV	AUT
21	A6009	Te da duo No 1	*G. max*[†]	II	SUM
22	D5942	Kan shanbai	*G. max*	III	SPR
23	A6020	Zaoshuhuangpi	*G. max*	IV	SUM
24	D6217	Qingpiqingren	*G. max*	III	SUM
25	A6019	Zaoshuhuangpi	*G. max*	IV	SUM
26	A6029	Yi hong bao	*G. max*	IV	AUT
27	D14207	Changtinqingpi	*G. max*	IV	AUT
28	A6035	Yanqing	*G. max*	IV	AUT
29	D498	Heitiejia	*G. max*	I	SPR
30	D537	Heiqi	*G. max*	I	SPR
31	D622	Hongmaoduo	*G. max*	I	SPR
32	D745	Jinduo 33	*G. max*	I	SPR
33	D781	Huangqitiejia	*G. max*	I	SPR
34	D823	Huangke	*G. max*	I	SPR
35	D1832	Fupingdahuangduo	*G. max*	II	SUM
36	D2013	Baiheiduo	*G. max*	II	SPR
37	D2024	Dahuangduo	*G. max*	II	SPR
38	D3146	Shenqiuniumaohuang	*G. max*	II	SUM
41	D403		*G. max*[‡]	III	
42	D404		*G. max*[‡]	III	
43	D405		*G. max*[‡]	III	
44	W9622		*G. soja*	IV	
45	W9624		*G. soja*	IV	
46	W9627		*G. soja*	IV	
47	W9628		*G. soja*	IV	
48	W9637		*G. soja*	III	
49	W9642		*G. soja*	III	
50	W9619		*G. soja*	III	
51	W9623		*G. soja*	IV	
52	W9632		*G. soja*	III	
53	W9633		*G. soja*	III	
54	W9638		*G. soja*	III	
55	W9640		*G. soja*	IV	
56	W9607		*G. soja*	II	
57	W9608		*G. soja*	II	
58	W9606		*G. soja*	II	
59	W9609		*G. soja*	II	
60	W9611		*G. soja*	II	
61	W9612		*G. soja*	II	
62	W9618		*G. soja*	I	
63	W9614		*G. soja*	I	

[†]Improved lines.
[‡]Formerly classified as *G. gracilis*.
[§]Places of origin: I, Northeast Region; II, stream of Yellow River; III, downstream of Yangtze River; IV, Southeast Region.

DNA markers have been used to evaluate genetic diversity in different crop species (Cooke 1995). The most-used markers have been Restriction Fragment Length Polymorphism (RFLP) (Cox *et al.* 1986), a highly reproducible assay which, however, detects only a few loci for each experiment and Random Amplified Polymorphic DNA (RAPD) (Williams *et al.* 1990; Vierling and Nguyen 1992; Kazan *et al.* 1993), which detects a larger number of markers but is not always as reproducible as RFLP (Cooke 1995). The RAPD system has the advantage of being a fast assay, requiring only nanograms of genomic DNA and minimum equipment. Therefore, it has been used to analyze the genetic relatedness in several crop species (Chalmers *et al.* 1992; Koller *et al.* 1993; Zhu *et al.* 1997). However, both RFLP and RAPD revealed a low level of DNA polymorphism in self-pollinating species such as soyabean, (Apuya *et al.* 1988; Williams *et al.* 1990; Keim *et al.* 1992). AFLP (Vos *et al.* 1995) and microsatellites proved to be the most powerful and reliable markers and revealed a much higher level of polymorphism in soyabean than RAPD and RFLP (Becker *et al.* 1995; Maughan *et al.* 1995; Maughan *et al.* 1996; Meksem *et al.* 1995). As AFLP is simpler and quicker it is particularly useful for germplasm studies, which in view of the large number of samples are time-consuming; it can also detect a higher level of polymorphism than RAPD and RFLP.

The purpose of the present study was to investigate, through the use of AFLP markers, the genetic diversity of Chinese soyabean germplasm from different places of origin.

Material and methods

The samples represented 41 Chinese soyabean accessions (*G. max)* from four main cultivation regions, including three accessions formerly classified as *G. gracilis* and 20 *G. soja* accessions (Table 1). Most of the accessions were kindly provided by Chang Ruzheng from the Crop Germplasm Institute of Beijing, China.

DNA was isolated from a pool of five plants for each accession according to Dellaporta *et al.* (1983) and its concentration estimated through agarose gel electrophoresis by comparison with known amounts of lambda DNA. AFLP experiments were carried out as described by Vos and associates (1995) with a few modifications: (1) 250 ng of DNA were initially digested with two enzymes,

(2) digestion reaction was incubated for two hours to obtain a complete digestion, (3) ligation was carried out for two hours, and (4) the pre-amplification products were diluted 40-fold instead of 10-fold. Samples were randomly selected to be run in duplicate experiments aimed at checking the reproducibility of the amplification patterns.

A total of eight primer combinations (Gibco) were used for AFLP assay in all accessions. After excluding monomorphic bands, polymorphic fragments were scored as 1 for presence and 0 for absence of the band. Measurement of genetic similarity for pair-wise accessions was based on the simple matching coefficients (Gower 1985). A similarity matrix was constructed and subjected to cluster analysis following the UPGMA method by computer program NTSYS-pc version 1.8 (Rohlf 1993).

Results and discussion

AFLP polymorphism in all samples

AFLP markers proved to be very informative in monitoring, with very high reproducibility, the genetic diversity present in a sample of 61 cultivated and wild soyabean accessions. The fragment sizes, determined by comparing the amplicons with standard DNA, ranged from about 40 to 600 bp. An example of a typical AFLP pattern is shown in Figure 1. Analysis of the 61 accessions with eight primer combinations revealed a total of 569 bands, with an average of 71 bands per primer combination among which 273 fragments (47.9%) were polymorphic across the whole set of samples (Table 2). The total number of fragments detected by individual primer pair ranged from 40 to 109 while the number of polymorphic fragments varied from 14 to 44 (Table 2). Different primer combinations detected different levels of polymorphism ranging from 37%, detected with the primer combination E-aag / M-ctt to 53% for the combination E-aac/M-ctc. This combination proved to be particularly informative as it detected 51 polymorphic fragments able to differentiate all 61 accessions. This result shows the ability of AFLP to discriminate among genotypes and suggests their application for cultivar identification. In comparison, the majority of soyabean RFLP probes detected only one polymorphic locus (Keim *et al.* 1989, 1992), while RAPD assay generally revealed about 0.5 polymorphic loci per primer (Williams *et al.* 1990).

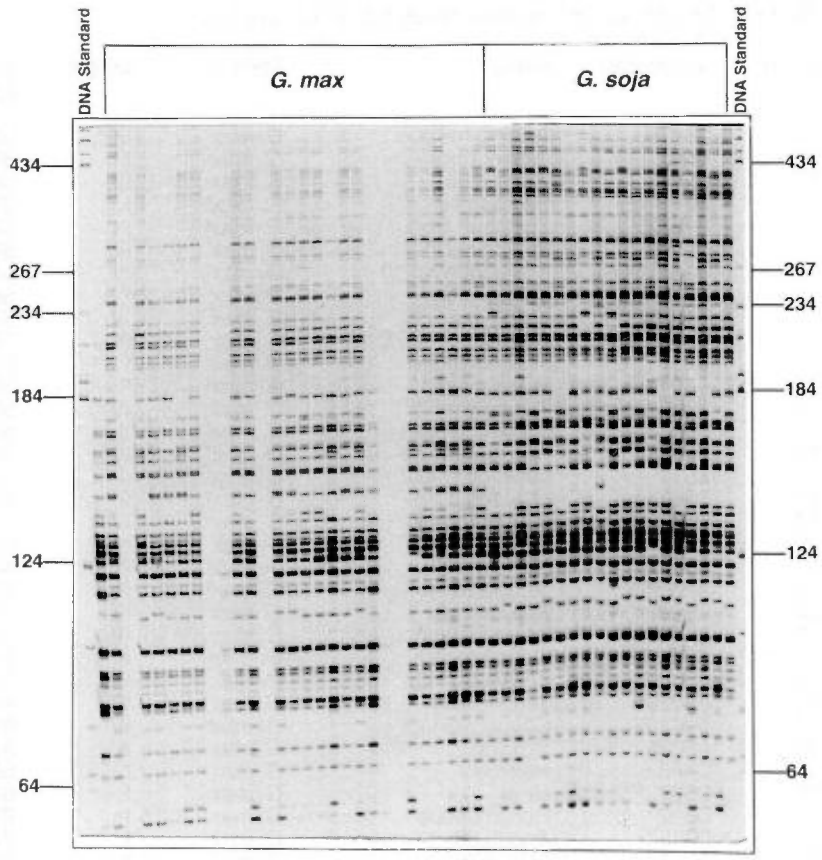

Fig. 1. AFLP patterns obtained with primer pair E-aag/M-cta of *G. max* and *G. soja* accessions. Amplified fragments were resolved on a standard DNA sequencing gel and detected by autoradiography.

Comparison of AFLP polymorphism between G. max *and* G. soja

Among the 274 polymorphic fragments, 146 (53.3%) were shared by both *G. max* and *G. soja* accessions. The frequency of polymorphism was 33.7% for *G. max* and 44.9% for *G. soja*; 31 fragments (11.4%) were uniquely associated with a single accession while 25 (9.1%) were dispersed among 14 wild soyabean accessions. Interestingly, 24 polymorphic fragments were observed only in *G. max* and among these 14 were present only in cultivated soyabean, while 103 were polymorphic only in *G. soja* and of these 65 were present

Table 2. Number of AFLP fragments detected at eight AFLP loci in soyabean

Primer combination	No. of polymorphic fragments				Total no. of fragments		% of polymorphism
	G. max	*G. soja*	shared	Total	*G. max*	*G. soja*	
E-aac/M-caa	23	34	19	39	68	77	48.8
E-aac/M-cac	14	14	12	16	40	41	39.0
E-aac/M-cat	22	34	21	35	62	70	50.0
E-aac/M-ctc	14	23	13	24	40	45	53.3
E-aac/M-ctg	16	22	14	24	51	56	42.9
E-aag/M-cta	28	41	25	44	79	88	50.0
E-aag/M-ctc	30	44	23	51	67	81	62.9
E-aag/M-ctt	23	37	19	41	97	109	37.6
Total	170	249	146	273	504	555	
Average	21.3	31.1	18.3	34.5	63	69.4	
% polymorphism	33.7	44.9		47.9			47.9

only in wild soyabean. Therefore, 79 more polymorphic bands were amplified in *G. soja* than in *G. max*, possibly due to the genetic erosion of cultivated accessions during the process of evolution. However, the reduced number of bands present in cultivated samples could also be due to breeding selection, which is known to result in the narrowing of the genetic base as a result of the loss of specific DNA sequences. The utilization of wild species in breeding programmes can be used to restore part of the diversity lost. It would be interesting to analyze in more detail the DNA sequences presently missing in cultivated species in order to have a better understanding of genetic erosion.

Genetic similarity

Based on the simple matching coefficients, a genetic similarity matrix was constructed using the AFLP data to assess the genetic relatedness among the 61 accessions. The similarity coefficients ranged from 0.606-0.945 for all accessions; the minimum genetic similarity was between a cultivated and a wild accession (accessions 22 and 57) and the greatest similarity between two wild accessions, 49 and 55. As expected, the genetic similarity present in most wild accessions is lower than that in cultivated accessions, confirming that wild soyabean preserves a larger genetic variation with which to increase genetic diversity in cultivated lines.

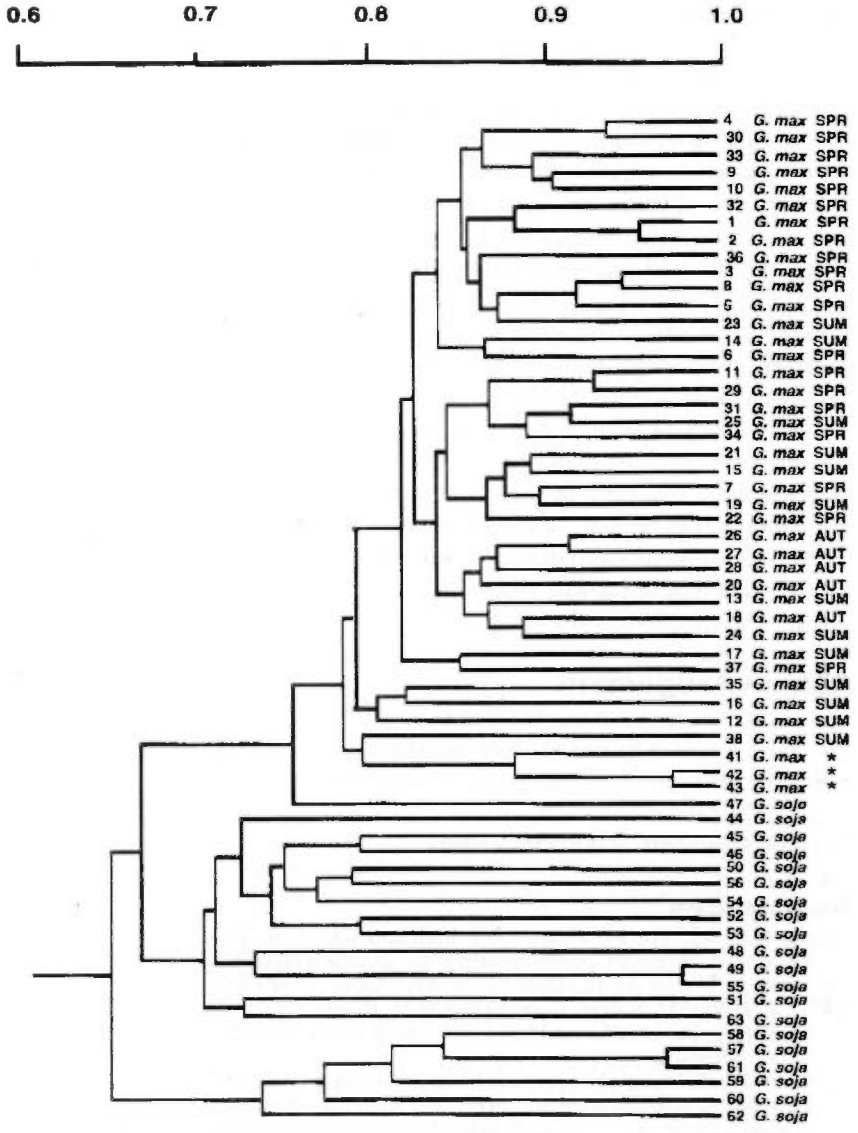

Fig. 2. Dendrogram generated by UPGMA cluster analysis based on simple matching coefficients using 273 AFLPs among 61 cultivated and wild soyabean accessions. The scale indicates genetic similarity.*Formerly classified as *Glycine gracilis*.

Cluster analysis

The similarity coefficients were used as input data for cluster analysis performed by the NTSYS-pc program. The resulting dendrogram is shown in Figure 2. *Glycine max* and *G. soja* clustered separately with the three semi-wild accessions (41, 42 and 43 classified as *G. gracilis* before the species merged with *G. max*) grouped with *G. max* at the lowest level of similarity present in the group, showing the complete concordance with Hermann's revision (Hermann 1962). Accession 47 was isolated singly from both groups in the dendrogram. This accession, however, has several agronomic characteristics of semi-wild soyabean: it develops main stem, and has ovate leaf shape and bigger seeds than wild soyabean. It possibly represents an intermediate evolutionary stage between *G. soja* and *G. max*.

To some extent, cultivated accessions were sub-clustered according to their sowing type, e.g. most of the spring-type accessions were clustered together, while four autumn accessions formed an independent group in spite of their different origin. These two types of soyabean were well separated in the dendrogram, showing a relative large genetic distance between them. Soyabean spring-types have the highest genetic similarity compared to the other types, with little genetic distance within accessions, while summer-type accessions were clustered in several groups and four (12, 16, 35 and 38) were isolated. This type of soyabean seemed to have larger intra-type genetic variations than the other two. In contrast, accessions with close geographic or pedigree relationships such as accessions 23 and 25, which are from the same village (they have even the same name), and accessions 6 and 14, where the latter is the male parent of the former, clustered together. The differences in genetic diversity among ecotypes provide valuable information for genebank curators on collecting and constructing core collections.

Conclusion

The repeatibility of AFLP banding patterns was very high. Compared to that obtained with conventional DNA markers, these markers displayed a high rate of polymorphism. The observed level of polymorphism is basically the result of two major components: primer combinations and diversity of accessions under study. In fact, Maughan and associates (1996) reported, in a sample of 23 accessions of cultivated and wild soyabean, a number of polymorphic loci detected per AFLP primer pair ranging from 9 to 27. This study revealed, in a broad range of 61 soyabean accessions analyzed with eight primer combinations, a much higher number of polymorphic loci ranging from 14 to 44. The primer combinations used in the two studies differ completely although the data suggest a larger genetic diversity in the Chinese genepool than in the American genepool. It would be very interesting to carry out a global comparison of the genetic diversity among different soyabean genepools in order to obtain further information on the species.

Acknowledgement

The authors wish to thank IPGRI for partially sponsoring this study and also gratefully acknowledge the contribution of the Institute for Vegetables and Flour Breeding, CNR, Portici. Contribution No. 201 of the University of Naples, Dept. of Agronomy and Plant Genetics.

References

Apuya, N.R., B.L. Frazier, P. Keim, E.J. Roth and K.G. Lark. 1988. Restriction fragment polymorphisms as genetic markers in soyabean, *Glycine max* (L.) Merr. Theor. Appl. Genet. 75:889-901.

Becker, J., P. Vos, M. Kuiper, F. Salamini and M. Heum. 1995. Combined mapping of AFLP and RFLP markers in barley. Mol. Gen. Genet. 249:65-73.

Chalmers, K.J., R. Waugh, J.I. Sprent, A.J. Simons and W. Powell. 1992. Detection of genetic variation between and within populations of *Gliricidia sepium* and *G. maculata* using RAPD markers. Genetics 69:465-475.

Cooke R.J. 1995. Varietal identification of crop plants. Pp.33-63 *in* New Diagnostics in Crop Science, (J.H. Skerrit and R. Appels, eds.). Biotechnology in Agriculture No. 13. CAB International, Wallingford, UK.

Cox, T.S., J.P. Murphy and M.D. Rodgers. 1986. Changes in genetic diversity in the red winter wheat region of the United States. Proc. Natl. Acad. Sci. USA 83:5583-5586.

Dellaporta, S.L., J. Wood and J.B. Hicks. 1983. A plant DNA minipreparation, Version II. Plant Mol. Biol. Rep. I:19-21.

Gower, J.C. 1985. Measure of similarity, dissimilarity, and distance. *In* Encyclopedia of Statistical Sciences, 5:397-405 (S. Kotz and N.L. Johnson, eds.). Wiley, New York.

Hermann, F.J. 1962. Revision of the genus glycine and its immediate allies. USDA Tech. Bull. 1268:1-82.

Kazan, K., J.M. Manners and D.F. Cameron. 1993. Genetic relationships and variation in the *Stylosanthes guianensis* species complex assessed by random amplified polymorphic DNA. Genome 36:43-49.

Keim, P., R.C. Shoemaker and R.G. Palmer. 1989. Restriction fragment polymorphism diversity in soyabean. Theor. Appl. Genet. 77:786-792.

Keim, P., W. Beavis, J. Schupp and R. Freestone. 1992. Evolution of soyabean RFLP marker diversity in adapted germplasm. Theor. Appl. Genet. 85:205-512.

Kiang, Y.T., Y.C. Chiang and N. Kaizuma. 1992. Genetic diversity in natural populations of wild soyabean in Iwate prefecture. Jpn. J. Hered 83:325-329.

Koller, B., A. Lehmann, J.M. McDermott and C. Gessler. 1993. Identification of apple cultivars using RAPD markers. Theor. Appl. Genet. 85:901-904.

Meksem, K., D. Leister, J. Peleman, M. Zabeau, F. Salamini and C. Gebhardt. 1995. A higher resolution map of the vicinity of the R1 locus on chromosome V of potato based on RFLP and AFLP markers. Mol. Gen. Genet. 249:74-81.

Maughan, P.J., M.A. Saghai Maroof and G.R. Buss. 1995. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soyabean. Genome 38:715-723.

Maughan, P.J., M.A. Saghai Maroof, G.R. Buss and G.M. Huetis. 1996. Amplified fragment length polymorphism (AFLP) in soyabean: Species diversity, inheritance and near-isogenic line analysis. Theor. Appl. Genet. 93:392-401.

Rohlf, F.J. 1993. NTSYS-pc Numerical taxonomy and multivariate system, version 1.8. Applied Biostatistics Inc., New York.

Vierling, R.A. and H.T. Nguyen. 1992. Use of RAPD markers to determine the genetic diversity of diploid wheat genotype. Theor. Appl. Genet. 84:835-838.

Vos, P., R. Hogers, M. Bleeker, M. Reijan, T. Van de Lee, M. Hornes, A. Freijters, J. Pot, J. Peleman, M. Kiuper and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acid. Res. 23:4407-4414.

Williams, J.G.K., A.R. Kubelik, K.J. Lavak, J.A. Rafaski and S.Y. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. 18:6531-6535.

Zhu, S., P. Manfredi, L.M. Monti and R. Rao. 1997. RAPDs markers useful for the identification of the 'Annurca' apple variety and its sport 'Rossa del Sud'. Adv. Hort. Sci. 11:120-122.

ARTICLE Plant Genetic Resources Newsletter, 1999, No. 119 Supp.:15-25

Sample size and variance effective population size for genetic resources conservation

J. Crossa[1]✉ and R. Vencovsky[2]

[1] Biometrics and Statistics Unit, International Maize and Wheat Improvement Center (CIMMYT), Lisboa 27, Apdo. Postal 6-641, Mexico D.F., Mexico. Tel: +1 804 2004; Fax: +1 804 7558; Email: j.crossa@cgiar.org
[2] Department of Genetics, ESALQ-USP, Cx.P. 83, 13.400-970, Piracicaba, SP, Brazil

Summary

Sample size and variance effective population size for genetic resources conservation

Despite the great genetic complexity of most plant populations it is possible to derive general guidelines for appropriate sampling strategies and sample sizes for maintaining genetic resources. Sample size is affected more by low allele frequency than by the number of alleles or number of loci. To maintain alleles at 3% frequency with 0.9999 probability, assuming unknown association between genes within individuals, the required sample sizes are between 177 and 564 individuals for two to 20 alleles per locus and for five to 1000 loci. For retaining alleles at 5% frequencies 105 to 335 individuals are required. For accession regeneration and for any level of accession loss, or in cases where all plants are used as male and female parents but only a portion of them are harvested, results indicate that full-sibs with female gametic control is a better system than half-sibs with female gametic control or random pollination and an unequal number of seeds taken from each pollinated parent. Accession losses increase drift due to sampling parents but controlling female and male gametes should decrease drift due to sampling of gametes and, therefore, help to control overall drift. The advantage of the half-sib system is that it does not require hand-pollination because pollen from all plants in the field can be used.

Key words: Effective population size, female gametic control, male gametic control, sample size

Résumé

Taille de l'échantillon et taille de la population générant la variance nécessaire à la conservation des ressources génétiques

Malgré la grande complexité génétique de la plupart des populations végétales, il est possible d'établir des règles générales destinées à élaborer des stratégies d'échantillonnage et des tailles d'échantillon afin de conserver les ressources génétiques. La taille de l'échantillon est davantage affectée par la faible fréquence allélique que par le nombre d'allèles ou le nombre de loci. Pour maintenir une fréquence allélique de 3 % avec une probabilité de 0,9999, en admettant une association inconnue entre les gènes chez les individus, les tailles d'échantillon requises sont comprises entre 177 et 564 individus pour 2 à 20 allèles par locus et pour 5 à 1000 loci. Pour maintenir une fréquence allélique de 5 %, il faut 105 à 335 individus. En ce qui concerne la régénération des accessions et tout niveau de perte d'accessions, ou dans les cas où toutes les plantes sont utilisées comme parents mâles et femelles, mais seule une partie d'entre elles sont récoltées, deux systèmes sont utilisables. Les résultats indiquent que des descendances biparentales avec contrôle gamétique femelle donnent de meilleurs résultats que les descendances uniparentales avec contrôle gamétique femelle ou pollinisation aléatoire et un nombre inégal de graines prélevées sur chaque parent pollinisé. Les pertes d'accessions augmentent la dérive due à l'échantillonnage des parents, mais le contrôle des gamètes mâles et femelles devrait réduire la dérive due à l'échantillonnage des gamètes et, donc, contribuer à la maîtrise de la dérive globale. L'avantage du système à descendances uniparentales est qu'il ne nécessite pas une pollinisation manuelle parce que le pollen de toutes les plantes dans le champ peut être utilisé.

Resumen

Tamaños de las muestras y dimensión de la variación efectiva de la población para la conservación de recursos genéticos

No obstante la gran complejidad genética de la mayoría de las poblaciones vegetales, se pueden obtener guías generales sobre adecuadas estrategias de muestreo y tamaños de muestras para la conservación de los recursos genéticos. El tamaño de las muestras depende más de la baja frecuencia de los alelos que de su número o de la cantidad de sitios que se toma en consideración. Se necesitan muestras de entre 177 y 564 individuos para una cantidad de alelos que va de dos a 20 por sitio y de cinco a 1000 sitios para mantener a los alelos a una frecuencia del 3% con una probabilidad del 0,9999, suponiendo una relación desconocida de genes con los individuos. Se necesitan de 105 a 335 individuos para conservar alelos a una frecuencia del 5%. Para la regeneración de la accesión y para cada nivel de accesión perdido, o en el caso en que todas las plantas se usen como progenitores masculinos y femeninos pero sólo una parte de ellas se coseche, los resultados indican que los hermanos completos con control gamético femenino representan un sistema mejor respecto al de medios hermanos con control gamético femenino o a la polinización aleatoria y presentan un número desigual de semillas escogidas de cada progenitor polinizado. Las pérdidas de las accesiones aumentan la deriva genética a causa del muestreo de los progenitores, pero controlando los gametos masculinos y femeninos debería disminuir dicha deriva a causa del muestreo de los gametos y, por consiguiente, contribuir en el control de la deriva genética en su conjunto. La ventaja del sistema de medios hermanos es que éste no requiere polinización manual porque se puede utilizar el polen de cualquier planta del campo.

Introduction

Genetic resources conservation programmes and genebanks around the world have received and stored hundreds of thousands accessions of different cultivated species, and their wild and weedy relatives, collected in their centres of origin or elsewhere. Accessions stored in genebanks must represent a wide spectrum of population diversity. Furthermore, regeneration protocols must consider the species (including allogamous, partially allogamous, autogamous and the type of material being regenerated) to ensure sample integrity and representativeness of the population. The great genetic complexity of most plant populations, and the many possible ways that genetic resources may be used in the future, makes it difficult to provide simple and efficient sampling schemes and optimal sample sizes for the mainte-

nance of all species (Namkoong 1986). Nevertheless, it is possible, based on probability theory and population genetic models, to derive general guidelines for appropriate sampling strategies and sample sizes despite the complex factors that influence sampling designs.

Stochastic changes in allele frequency caused by sampling error in small populations (random genetic drift) lead to a continuous fixation and loss of alleles, and reduce the proportion of heterozygous individuals in the population (Wright 1931; Crow and Kimura 1970). These random changes in allele frequency, that occur in finite populations subject to sampling error, affect the genetic representativeness of the population. In a large mating population of N individuals, the reduced number of progenitors whose offspring will constitute the next generation causes random genetic drift, which is quantified and predicted using the parameter called "effective population size" (N_e).The effective population size is the size of an ideal population that has the same amount of drift in allele frequency or the same rate of decrease in heterozygosity as the actual population. Historically, there have been two main approaches to quantify N_e. The first is related to the inbreeding that occurs in a breeding population and is called "inbreeding effective population size" [$N_{e\,(I)}$]. The second is concerned with the sampling variance of allelic frequency in a breeding population and is called "variance effective population size" [$N_{e\,(v)}$]. Several factors affect the effective size of a population, including the number of parents per generation, the number of corresponding offspring, the number of male and female gametes contributed per individual in the parental population, the mating system of the species, and so on. The effective population size, taken as a measure of the genetic representativeness of a seed sample, can be adapted to specific aspects of plant breeding and genetic resources conservation such as seed regeneration and collection.

During the process of collecting, storing, regenerating and restocking germplasm, genetic drift occurs and affects the genetic integrity of the population in a number of ways. First, when collecting germplasm in the field, an appropriate sampling strategy should be used to avoid, as much as possible, dramatic reduction in the size of the population (bottleneck). Second, when the accession is stored, different survival rates of genotypes and the accumulation of mutations effect the genetic integrity of the accessions. Third, when seed of an accession is drawn for regeneration, the sample needs to be optimum to avoid genetic drift due to sampling or due to differential survival (germination) or fecundity, leading to changes in the genetic constitution of the accession. Fourth, when plants of accessions are in the field for regeneration, insects, diseases or other environmental factors can affect plant stand and, therefore, limit their gametic (offspring) contribution to the next generation. Since we do not know which alleles will prove to be useful in the future, it is essential that sampling be done efficiently and that populations be of sufficient size to maintain as much genetic diversity as is practical. Large samples are expensive and difficult to manage, but if samples are too small, genetic diversity may be affected by the loss of valuable alleles through random changes in allele frequency. Maintaining allelic diversity during regeneration depends on factors such as sampling procedures, seed viability and mating systems. These factors cause random genetic drift.

Probabilistic treatment of the problem of sampling alleles originates from the theory of genetic drift in finite populations. In general, most of the emphasis on sampling an accession for regeneration is on cross-fertilizer species. However, several of the problems discussed earlier also affect mixed self- and random-mating species and self-fertilizer species, especially if gene diversity of accessions is high.

One of several reviews of theoretical and practical aspects of the sample size required for genetic conservation of cross-fertilizer species may be found in Crossa (1989) and for regenerating maize genebank accessions in Crossa *et al.* (1994). Several authors have recommended that an appropriate strategy for seed regeneration should be to equalize the number of progeny per family. This would increase the effective population size and, therefore, prevent the loss of genetic variation. A recent simulation study on evolution has shown that the practice of equalizing the number of progeny per family during 25-50 regeneration cycles allows the number of mildly deleterious mutations to build up in the germplasm accession more rapidly than when the seed is bulked (Schoen *et al.* 1998). This potentially affects the viability and fitness of the germplasm held in genebanks. However, the authors used a genetic model that did not allow for natural selection and different mutation effects.

Statistical genetic models used by Crossa *et al.* (1993) indicate that for maintaining germplasm accessions for seed regeneration, sample sizes of 160-250 plants are required for capturing alleles at frequencies of 0.05 or higher in each of 150 loci, with 90-95% probability. These formulas considered the sample size for conserving at least one allele per locus, but they do not quantify the probability of conserving two, three or more alleles per locus. Gregoruis (1980) developed a general formula for the probability that not all the alleles present in a population are represented in a sample of diploid individuals. The genetic diversity of a population depends on the number and frequency of alleles at a locus and across loci. Sample-size determination depends on whether estimates of allele frequencies are available (Hernandez and Crossa 1993). For more than one independent locus, the probability of including at least one of each allele at each locus in a sample of n gametes is the product of the probabilities for each locus.

Crossa *et al.* (1994) recommended specific practical procedures for regenerating and maintaining genetic resources in maize to minimize the loss of alleles and maintain genetic diversity. Vencovsky (1978) proposed the variance of the number of female and male contributed gametes for random-mating species based on a two-stage sampling model. Crossa and Vencovsky (1994; 1997) developed manageable computation procedures for deriving the variance of the number of contributed gametes [$V_{(k)}$] and thus computing the variance effective population size [$N_{e\,(v)}$]. They suggested sampling procedures and mating systems that could increase the effective population size for seed regeneration.

Vencovsky and Crossa (1999a) extended the two-stage sampling model for random-mating monoecious species to estimate $V(k)$ and $N_{e(v)}$ for species with mixed self- and random-mating systems of reproduction. The authors considered different values of the natural rate of self-fertilization (s) and natural level of inbreeding (f) and illustrated results for the case of regeneration and collection of genetic resources of species.

Lindgren *et al.* (1996, 1997) have defined the status number (N_s) as a suitable measure for describing the diversity of a pedigree population. The status number is half the inverse of the average co-ancestry of the group of individuals to which N_s refers. The average co-ancestry is the probability that two random genes sampled from the group are identical by descent. The status number is related to the variance effective number given by Cockerham (1969) as a function of the group co-ancestry. While Lindgren's N_s admits an arbitrary co-ancestry structure within a group of individuals, Cockerham's definition assumes a random model such that individuals of a group have all the same pedigree history. When compared with the $N_{e(v)}$ theory and formulas given by Vencovsky and Crossa (1999a), Lindgren's and Cockerham's measures correspond to a situation in which the reference population is of infinite size as may be assumed in germplasm collection. For this specific situation, however, Lindgren's N_s is more general because it does not require a regular pedigree structure and thus is particularly suited to quantify the degree of representativeness of groups of individuals in breeding programmes. Under a random model and sampling from infinite populations all three approaches are identical. The $N_{e(v)}$ theories given by Crossa and Vencovsky (1997) and Vencovsky and Crossa (1999a), on the other hand, also accommodate a situation in which sampling is from a finite reference population, as may be assumed in accession regeneration.

In this paper we use the theory and equations developed earlier and demonstrate their use for answering practical questions that a genetic resources conservation manager may pose when regenerating species. Practical examples are shown for: (1) computing the required sample sizes for maintaining at least one copy of each allele at independent loci present in the accession to be regenerated with certain probability, (2) assessing practical procedures involving crossing systems such as full-sibs (FS) and half-sibs (HS) that will tend to decrease the variance of the number of contributed gametes [$V(k)$] and thus increase the variance effective population size [$N_{e(v)}$] of random-mating populations, and (3) calculating $V(k)$ and $N_{e(v)}$ for species with mixed self- and random-mating systems of reproduction with different levels of natural rate of self-fertilization (s).

Probability models for determining sample size for regeneration

We are interested in finding a sample size n such that the probability of detecting at least one allele of each allele class is greater than a quantity 1-α or $P(n)$>1-α (for small α). In probability theory this is referred to as the coupon problem. Crossa (1989) gave the probability that each of the k allele classes at a locus will be represented at least once in the sample and Crossa *et al.* (1993) extended the expression to m independent loci. Hernandez and Crossa (1993) developed a computer program to evaluate mathematical expressions leading to estimates of the required sample size (n) for values of α, number of alleles classes (k) and number of loci (m). Since diploid individuals are sampled (not gametes) the following assumptions underlying the model should be considered: (i) seeds are sampled regardless of the genotype of the parents, (ii) linkage equilibrium (no associations among genes from different loci), (iii) if there are no associations between genes within individuals at any locus, the required sample size is half the sample size reported by the program [(½)n], (iv) if there is a perfect association between genes within individuals at any locus, the required sample size equals the sample size reported by the program (n), (v) if the association between genes within individuals at any locus is unknown, the required sample size is between (½)n and n.

Except for the case of synthetics derived from inbred lines, in real situations allele frequencies are unknown, therefore, simplification of the original equations is required. Making the assumption that k-1 alleles occur at an identical low frequency of p_o and that the k^{th} allele occurs at a frequency 1-[(k-1)p_o], Crossa *et al.* (1993) showed that the sample size (n) required to retain, with probability P=1-α, at least one copy of each of the k allelic classes in each of the m loci should be larger than

$$n > \frac{\log[1-(P)^{1/m}] - \log(k-1)}{\log(1-p_o)}$$

For loci with two, three or four alleles, each with p_o=0.05, at least 89 to 110 individuals are required if at least one allele at each of 10 loci is to be retained with a P=1-α=90% prob-

Table 1. Sample size required to retain at least one copy of each allele with k-1 alleles at frequencies p_o=0.05, 0.03, and 0.01 for several alleles (k) per locus with 0.9999 probability [P=(1-α)=0.9999]

Number of alleles per locus	Number of loci			
	5	10	100	1000
	p_o=0.05			
2	211	224	269	277
4	232	246	291	299
10	254	267	312	320
20	268	282	327	335
	p_o=0.03			
2	355	378	454	467
4	391	414	490	503
10	427	450	526	539
20	452	475	550	564
	p_o=0.01			
2	1077	1146	1375	1415
4	1186	1255	1484	1524
10	1295	1364	1593	1634
20	1370	1438	1668	1708

ability. If 100 loci are involved, at least 134 to 155 individuals are required. For two, three or four alleles at p_e=0.003 and 10 loci, at least 150 to 180 individuals are required; for 100 loci, 225 to 255 individuals are needed. Sample sizes of at least 160 to 210 individuals are required to retain alleles at frequencies of p_e =0.05 in each of 150 loci, with a 90-95% probability.

However, when the required probability (1-α) increases or the frequency of the alleles drops to 1% (p_e=0.01), larger sample sizes than those specified above are required. For example, Table 1 shows sample sizes required for achieving P=1-α= 0.9999 probability of retaining at least one allele for different number of allele classes at different frequencies and for various numbers of loci. For unknown association between genes within individuals and p_e=0.001, sample sizes between 533 and 1708 individuals will retain, with probability of 0.9999, at least one allele from each class considering a wide range of allele classes (k=2 to 20) and number of loci (m=5 to 1000). Table 1 clearly indicates that sample size is affected more by low allele frequency than by the number of alleles or number of loci. To maintain alleles at 3% frequency (p_e=0.03) with 0.9999 probability, assuming unknown association between genes within individuals, the required sample sizes are between 177 and 564 individuals for 2 to 20 alleles per locus and for 5 to 1000 loci. For retaining alleles at 5% frequencies 105 to 335 individuals are required (Table 1).

For the case of 6 alleles per locus with frequencies between 0.03 and 0.10 and for 1 to 20 000 loci, sample size between 84 to 750 individuals will preserve at least one copy of each allele class in each locus with a probability of 0.9999999 (data not shown). Assuming two alleles at each of the 20 000 loci and one of them at a 0.05 frequency, 186 individuals will preserve this allele at each locus with a 95% probability. This is similar to the 172 individuals suggested by Lawrence *et al.* (1995). However, our results indicate that sample sizes of 172 individuals will only retain an allele at 5% frequency assuming no association between genes within individuals at any locus; however, it does not seem to be sufficient if associations between genes within individuals exist at some loci and/or if maintaining alleles between 3% and 1% frequencies are required. General recommendations and/or guidelines concerning sample sizes should be given within ranges (or intervals) depending on model assumptions and biological considerations such as number of alleles, loci, genetic linkage, etc.

Note that assuming that k-1 alleles occur at an identical low frequency for all loci makes the required sample size shown in Table 1 conservative, because it is likely that some alleles will have higher frequencies at least in some loci. The assumption of independence of loci is unrealistic and genetic drift may cause non-random associations between linked loci. If the population is under random mating (cross-pollinated species), linkage equilibrium can be assumed for all possible pairs of loci so that it is expected, for each pair of loci, that similar number of coupling and repulsion combinations occur. Evidently this does not hold for self-pollinated species or species with mixed self- and random-mating systems of reproduction. Between 20% to 30% gametic disequilibrium is obtained for effective populations sizes of 50 to 100 (Crossa 1989); 20% to 30% of the gametes will contain linked loci and for smaller effective sizes the linkage disequilibrium approaches the maximum value. However, if loci are in linkage equilibrium larger population sizes are required to break linkage groups such that gametic disequilibrium approaches 0 (gametic equilibrium).

It should be pointed out that computations for determining the sample size for retaining at least one allele from each allele class at each locus with certain probability offer approximate guidelines on range of sample sizes, which, as previously mentioned, are greatly dependent on the frequency of the rare alleles. Furthermore, as stated in the assumptions of the model, sampling n=2N gametes is equivalent to sampling ½ n=N individuals (diploid zygotes) only for panmictic populations (i.e., idealized random-mating population of infinite size with no association between any two genes within individuals); in this case the number of diploid individuals needed to be sampled is exactly half the number of gametes (n).

The range of sample sizes discussed above are based solely on probability models and do not consider the genetic structure of the population or specify how well a particular sample represents the reference population in terms of genetic parameters such as variance of allele frequency, inbreeding, random genetic drift due to sampling error, genetic linkage, etc. Furthermore, by themselves probability models do not specify appropriate matings and reproduction systems such as panmixia, mixed self- and random-mating, self-fertilization, and so forth that could, under specific circumstances, maximize genetic representativeness of the sample. To illustrate the point, consider the probability (P) of including at least one allele of class a_2 in a random sample of size N_0 individuals taken from a large panmictic population as

$$P = 1 - (p_1^2)^{N_o} = 1 - p_1^{2N_o}$$

(a sample of N_0 individuals corresponds to a sample of $2N_0$ genes). If the base population is in inbreeding equilibrium, with a certain amount of natural self-fertilization, the probability is

$$P = 1 - [p_1^2 + fp_1(1 - p_1)]^{Nf}$$

where f is the inbreeding coefficient for the group of individuals and N_f is the new sample size that can be written as $N_f = b_f N_0$. Assuming the same probability value (P) in both cases and solving for b_f it is obtained that $b_f = 2/(1+C_f)$ where $C_f = \{\ln[p_1 + f(1-p_1)]\} / \ln(p_1)$ for any probability value P and $0 < p_1 < 1$. Quantity b_f represents how much larger N_f should be than N_0 for attaining the same probability P as attained with a sample of size N_0. Values of b_f range between 1 and 2 depending on the level of f; for f=0, b_f=1 and for f=1, b_f=2. To maintain a given value of P, N_f increases with f, relative to N_0.

Now introduce the variance effective population size for the sampling situation described above such that

$$N_{e(v)} = \frac{n}{1+f}$$

(Vencovsky and Crossa 1999a) or $n=(1+f)N_{e(v)}$ where n is the required sample size. Using the previous notation $N^*_f = (1+f)N_{e(v)}$ and for f=0 $N_{e(v)}=n=N_0$. Thus, $N^*_f=(1+f)N_0=b^*_fN_0$ where $b^*_f = (1+f)$ ranges between 1 and 2 as the previous b_f for f=0 and f=1, respectively but for $0 < f < 1$, $b^*_f > b_f$. For example, for f=0.2, 0.4, 0.6 and 0.8 and p_1=0.05, b_f=1.11, 1.26, 1.44, 1.67 and b^*_f =1.20, 1.40, 1.60, 1.80, respectively. Factor b^*_f is more conservative (it provides larger sample sizes) than b_f (except for f=0 and f=1) and has the advantages that it provides larger probability of retaining allele a_2 than that obtained using factor b_f and also assures fewer changes in allele frequencies due to sampling (less genetic drift). The same holds also for other values of p_1 such as p_1=0.03 or 0.01. These considerations may be especially useful when defining appropriate sample size (n) for collecting germplasm.

The effect of inbreeding in regenerating small accessions is also important. Frankel and Soulé (1981) have suggested that for seed regeneration, the rate of inbreeding (f) should be kept below 1%. For two alleles per locus, the rate of inbreeding can be approximated by $f=1/(2N_e)$ (Wright 1931). In general, the level of inbreeding when changing from x alleles to x-1 alleles is $f=x(x-1)/(4N_e)$ (Kimura 1955) and thus, $N_e = x(x-1)/(4f)$. Therefore, if $N_e \geq N$ is required, N=150 individuals will restrict inbreeding to less than 1%. However, in non-ideal populations (e.g., accession regeneration), it is very likely that $N_e < N$. For studying factors that will make $N_e \geq N$, it is necessary to incorporate aspects of population genetics theory into probability models that will facilitate estimation of parameters such as the mean and variance of male and female contributed gametes and the mating systems that would facilitate control of male and female gamete contributions.

Variance effective population size for regenerating genetic resources

Concepts and a basic model

The concept of effective population size N_e is useful when studying the breeding structure and genetic representativeness of an actual population as related to an ideal population in which individuals mate at random with no variation in fertility; the number of progeny per parent has a Binomial distribution (that approaches Poisson for large sample size) and its size is kept constant through time (Wright 1931). The effective population size determines the amount of sampling error between generations that causes random fluctuations of allele frequency. However, as opposed to the ideal population, natural and breeding populations fluctuate in size from one generation to another and some of the individuals may not produce gametes and may have different natural levels of inbreeding and natural rates of self-fertilization than others. Thus, measures of effective size often deviate from the size of the actual population.

Crow and Kimura (1970) derived models for studying the genetics of finite populations and assessed the relationship between the actual size of a population, N, and the distribution of progeny among parents by estimating its mean and variance. They derived an expression $N_{e(v)}$ when a random sample of individuals is taken from an original population of size N assuming that all the individuals potentially contribute with male and female gametes. The equation is

$$N_{e(v)} = \frac{2n}{\frac{s^2_k}{\bar{k}}(1+f)+(1-f)} \quad \text{Eq. 1}$$

where f measures the departure from Hardy-Weinberg equilibrium in generation t-1; n is the number of offspring in generation t (or total number of seeds collected from the parents); k is the number of gametes contributed by a particular parent that can be considered a Binomial random variable in the case of random sampling of gametes and no variation in fertility;

$$\bar{k} = \frac{2n}{N} \quad \text{and} \quad s^2_k = \left(\frac{N}{N-1}\right) V(k)$$

are the mean and variance of the gametes contributed by the parents, respectively. For large N, $s^2_k = V(k) = \bar{k}$ (Poisson).

Two-stage sampling model for random-mating species

As previously mentioned, Crow and Kimura (1970) assumed that all N individuals in the parental generation contribute potentially with gametes to the next generation. However, this is not a realistic assumption in many practical situations that arise in plant breeding, regeneration and collection of plant genetic resources [where some plants may fail to produce gametes (offspring) due to external factors such as poor seed viability or seed germination, insects and diseases in crossing blocks, etc.]. Therefore, a more realistic model should consider that the overall effect of sampling on allele frequency drift is the summation of drift occurring at two stages: (i) when sampling parents from the original population (first stage), and (ii) when sampling gametes from these parents (second stage). The effect of the first stage on drift is reduced by increasing the number of randomly selected parents, whereas the influence of the second stage on drift is reduced by controlling male gametes through hand-pollination and by controlling female gametes by taking equal numbers of seeds per pollinated parent.

Vencovsky (1978) was the first to outline the theory, procedures and models for computing $V(k)$ and $N_{e(v)}$ for the two-stage sampling model for monoecious species that considers that some plants may not contribute gametes to the next generation. The model was adapted to specific aspects of selection and germplasm collection and regeneration. Crossa and Vencovsky (1994) provided approximate equations for computing $V(k)$ and $N_{e(v)}$ for germplasm collection and regeneration with and without control of contributed male and female gametes. Crossa and Vencovsky (1997) provided the theoretical development of the two-stage model, gave an alternative derivation based solely on the theory of random sampling within finite population, and demonstrated practical procedures involving full-sibs and half-sibs crossing systems.

A brief description of the two-stage sampling model follows. Consider an initial set of N diploid monoecious plants (reference population) and that P plants contributing *male* and *female* gametes are randomly selected ($0 < P \leq N$).

From the remaining N-P plants, R are additionally sampled at random contributing only *male* gametes [$0 \leq R \leq (N\text{-}P)$]. Therefore, $M=P+R$ plants contribute *male* gametes, P plants contribute *female* gametes, and N-M plants do not contribute any gametes at all. Thus, the proportion of functional seed parents is $u=P/N$ (or $P=uN$) (where $0 < u \leq 1$) and the proportion of functional pollen parents is $v=(P+R)/N=M/N$ (or $M=vN$) ($0 < v \leq 1$). With these proportions it is possible to simulate several important practical situation that arise in breeding selection, regeneration and the collection of genetic resources. For example, the condition given by Crow and Kimura (1970) when developing Eq. 1 that all N individuals potentially produce male and female gametes, is achieved by considering $P=M=N$ or $u=v=1$, that is, a perfect stand of plants in the field. For seed regeneration this means that all plants in the field are used as male and female parents.

On the other hand, if $P<N$, $M<N$ such that $u<1$, $v<1$ some plants fail to produce gametes due to poor germination, poor seed viability, or due to a poor stand of plants in the field because of insects, diseases or other environmental factors. If it can be assumed (for seed collection of random-mating species or mixed self- and random-mating species) that the number of pollen plants (M) is very large, then potentially $M \cong N$ and $v \cong 1$. This is an oversimplification but is considered for calculating an upper limit for $N_{e(v)}$. Also, the number of seed parents can be considered much smaller than N, $P<<N$, and $u \cong 0$. For complete self-fertilizer species $P=M$ and $u=v$, necessarily.

Vencovsky (1978) and Crossa and Vencovsky (1994, 1997) adapted the variance of the number of contributed gametes $V(k)$ of Eq. 1 for the two-stage sampling model for random-mating species (natural rate of self-fertilization, $s = 0$) for regeneration and collection of genetic resources. $V(k)$ was partitioned in such a way that it reflects the concept of total drift being the summation of drift when sampling the parents (first stage), plus drift when sampling gametes from the parents (second stage).

For germplasm regeneration of N (finite) plants in the field, assuming that population size is kept constant, that is, $n=N$, equations can be adapted for different cases and mating designs for monoecious species. Two general cases are considered here.

Case 1. The integrity of accession is affected ($u=P/N<1$), that is, there is a loss of seeds of the original accession, but pollen and seeds are contributed by all remaining plants ($P=M$). Then $P= uN$, $M= uN$ ($P =M$; $u = v$). This case resembles loss of seed accessions due to environmental causes during storage in the genebank or while accession plants are growing in the field.

Case 2. The integrity of accession is not affected, that is, pollen come from all plants ($M=N$), but seed is collected from only $P= uN$ plants. Then, $M=N$, $v=1$, but $P= uN$. This case simulates the situation when only a portion of the total number of pollinated plants in the field are harvested.

For each case three mating designs based on different gametic control are considered:

a) female and male gametic control which represents a full-sibs mating system,

b) female gametic control only which represents a half-sibs mating system, and

c) without female and male gametic control, that is, open pollination is at random and an unequal number of seeds are randomly taken from the set of P parent.

For Case 1, male and female gametic control involves full-sib crosses and taking an equal number of seeds from $P=M=uN$ plants, whereas for Case 2, male gametic control includes full-sibs among the $M=P=N$ plants, but seeds are collected from only $P= uN$ plants with female gametic control. Furthermore, for Case 1, random pollination involves half-sibs among $P=M= uN$ plants and female gametic control implies taking an equal number of seeds from $P= uN$ plants, whereas for Case 2, random pollination includes half-sibs among all plants ($M=P=N$), but female gametic control implies taking seeds from only $P= uN$ pollinated

Table 2. $N_{e(v)}$ for germplasm regeneration of monoecious species for Cases 1 and 2 with and without gametic (male and female) control, constant population size ($n=N=1000$) and different proportion of parents contributing with male and female gametes to the next generation (u)

	Case 1 Seed $P=uN$ Pollen $M=uN$			**Case 2** Seed $P=uN$ Pollen $M=N$		
u	$N[2u/(2-u)]$†	$N[4u/(4-u)]$‡	Nu§	$N[4u/(1+u)]$†	$N[4u/(1+2u)]$‡	$N[4u/(1+3u)]$§
0.1	105	103	100	364	333	308
0.2	222	211	200	667	571	500
0.3	353	324	300	923	750	632
0.4	500	444	400	1143	889	727
0.5	667	571	500	1333	1000	800
0.6	857	706	600	1500	1091	857
0.7	1077	848	700	1647	1167	903
0.8	1333	1000	800	1778	1231	941
0.9	1636	1161	900	1895	1286	973
1.0	2000	1333	1000	2000	1333	1000

†Female and male gametic control.
‡Female gametic control and random pollination.
§No gametic control.

plants. For the purpose of comparing results, it should be pointed out that Case 1 with female and male gametic control, and Case 1 with random pollination and female gametic control correspond to the full-sibs (FS) and the half-sibs with only P parents intercrossed (HS') of Crossa and Vencovsky (1997), respectively. Case 2 with random pollination and female gametic control corresponds to the half-sibs system (HS) of Crossa and Vencovsky (1997). Case 2 with female and male gametic control was not considered by Crossa and Vencovsky (1997).

Equations for $N_{e(v)}$ expressed in term of the proportion of functional parents (u) for Case 1 and 2 and for the three mating systems are given in Crossa and Vencovsky (1994, Table 1). Here, in Table 2 (for n=N=1000), we showed the evaluation of these six expressions considering a range of accession losses from 90% (u=0.1) to 0% (u=1.0, full stand of plants in the field). For Case 1 and 2, with a perfect stand of plants (u=1.0), taking an equal number of seeds per pollinated plant (female gamete control), and using hand-pollination (full-sibs) (male gametic control) produces $N_{e(v)}$ =2N; with only female gametic control and random pollination (half-sibs), $N_{e(v)}$ =(4/3)N, and without male and female gametic control $N_{e(v)}$ =N.

Values of $N_{e(v)}$ for Case 2 (last three columns of Table 2) are always larger than their corresponding columns for Case 1 (first three columns of Table 2) due to the larger pollen pool existing in the case where M=N as compared to M= uN. A question of interest is how male and female gametic control can help overcome the loss of seed or plant accessions. In other words, how can a genebank manager regenerate an accession from a monoecious species that has lost, for example, 20% or more of its seeds or plants and still be able to attain $N_{e(v)}$ =N? For Case 1, a loss of 30% (u=0.7) of the plant accessions can be recovered by male and female gametic control ($N_{e(v)}$ =1077) (Table 2), whereas a loss of 20% (u=0.8) can be recovered by simply taking an equal number of seeds from the 800 randomly pollinated plants that remained in the field ($N_{e(v)}$ =1000) (Table 2). It is interesting to note that for Case 2, where initially all 1000 plants are used to form full-sibs, an equal number of seeds taken from only 400 of them (u=0.4 or 60% of accession loss), produces values of $N_{e(v)}$ slightly larger than N [$N_{e(v)}$ =1143]. Furthermore, if all 1000 plants are left to pollinate randomly (half-sibs), but an equal number of seeds are taken from only 500 (u=0.5) of them, $N_{e(v)}$ =N (1000). This indicates the importance of controlling female gametes as a mechanism for recovering the magnitude of $N_{e(v)}$ due to the possible loss of seed and/or plant accessions before or during the growing cycle (Case 1) [or the possible problems encountered by harvesting only a portion of the total number of pollinated plants in the field (Case 2)].

It should be pointed out that $N_{e(v)}$ is larger for female gametic control with the half-sibs of Case 2 (second column of Table 2) than for female gametic control with the full-sibs of Case 1 (first columns of Table 2) up to an accession loss equal to or greater than 30% ($u \leq 0.7$) (1167 vs 1077 for u=0.7). When accession loss is smaller than 30% (u>0.7), female gametic control with full-sib families from P=M= uN plants (Case 1) is superior to that of half-sib families from M=N plants, but female gametic control on only P= uN families (1333, 1636 and 2000 vs 1231, 1286, and 1333 for u=0.8, 0.9, and 1.0, respectively). Results of Table 2 are presented graphically in Figure 2 of Crossa and Vencovsky (1997) where it is seen that half-sib is the best option for recovering a large loss of accessions (smaller values of u) through increasing $N_{e(v)}$ in contrast to the full-sibs alternative that becomes important only for smaller loss of accessions ($u \geq 0.8$). On the other hand, the female gametic control with the full-sib system of Case 2 (first column of Table 2) is always superior to the female gametic control with the half-sib system of Case 1 (second column of Table 2) for all values of accession loss as well as for the situation of no gametic control (third columns of Table 2). This is due to the larger pollen pool of Case 2 as compared to Case 1.

For any level of accession loss (Case 1), or in the case where all N plants are used as male and female parents but only a portion are harvested (Case 2), results indicated that full-sibs with female gametic control is a better system than half-sibs with female gametic control or random pollination and unequal number of seeds taken from each pollinated parent. An exception was found when comparing {HS with female gametic control of Case 2} vs {FS with gametic control of Case 1}; the former is preferred when deterioration of the accession is intermediate to severe ($0 < u \leq 0.7$). Accession losses of Case 1 increase drift due to sampling parents (first stage of sampling) but controlling female and male gametes should decrease drift due to sampling of gametes (second stage of sampling) and, therefore, helps to control the overall drift. The advantage of the half-sib system is that it does not require hand-pollination because pollen from all plants in the field can be used. It is a valuable alternative for decreasing the contribution to drift due to sampling parents (first stage).

Two-stage sampling model for mixed self- and random-mating species

Classifying plant species strictly as self-fertilizer or cross-fertilizer is a very superficial. Many self-compatible species may have a high natural rate of self-fertilization (s) so that they can be considered to have a mixed self- and random-mating reproduction system. Evidently the genetic structure of a mixed self- and random-mating species is complex because not all plants have the same level of natural inbreeding (f). Furthermore, progenies of a population with a mixed self- and random-mating species will have a mixture of selfed seed (S_1) with proportion s and half-sibs or outcrossed seeds with proportion 1-s. The proportion s can be artificially manipulated by hand-pollination so that s=1 when selfing all plants or s=0 when crossing all plants. However, in natural populations s varies from 0 to 1 ($0 \leq s \leq 1$) and they are assumed to be in inbreeding equilibrium when they reproduce naturally. In that case the relationship between s and f is given by

$$f = \frac{s}{2 - s} \quad (\text{or } s = \frac{2f}{1+f})$$

Inbreeding equilibrium can be also assumed in germplasm collection and regeneration in which the natural mating system of the species is not disturbed by hand-pollination or other factors.

Vencovsky and Crossa (1999a) extended the two-stage sampling models for random-mating species to the case of mixed self- and random-mating species ($0 \leq s \leq 1$) and developed estimates and direct expressions for computing $V(k)$ and $N_{e(v)}$ with applications to specific aspects of germplasm regeneration and collection. The authors considered an initial set of N diploid plants where P plants, contributing with male and female gametes, are randomly sampled from N and subsequently R plants, contributing with only male gametes, are randomly sampled from N-P. Proportions u and v are defined as in the two-stage sampling for random-mating monoecious species. The model allows correlation between number of female and male gametes contributed per individual within the set of P plants (assumed to be zero for the case of random-mating species). Therefore, the natural or artificial rate of self-fertilization (assumed constant over parents) is s, the proportion of crossing is 1-s, and n is the total number of seeds collected from the set of P parents. Then, $2ns$ is the expected total number of female and male gametes contributed from selfing and $2n(1-s)$ is the expected total number of male and female gametes contributed from crossing. The overall total of contributed gametes is $2ns+2n(1-s)=2n$.

The basic mathematical expressions derived by Vencovsky and Crossa (1999a) for estimating $V(k)$ and $N_{e(v)}$ are general and cover specific cases previously developed and studied by several other authors. However, these expressions are restricted to the case of binomial variance when the number of seeds can be considered a random variable. For example, with specific restrictions, it led to the same $N_{e(v)}$ that was: (1) obtained by Cockerham (1969) using intraclass correlations, (2) used by Gale and Lawrence (1984) when discussing effective population size and rate of self-fertilization, and (3) developed by Lindgren *et al.* (1996, 1997) called the status number (N_s). The basic expressions developed by Vencovsky and Crossa (1999a) considered different mating situations and inbreeding status. They considered random sampling of seeds (RS), female gametic control (GC), unrestrictive inbreeding (UI) and inbreeding equilibrium (IE). For the cases of RS-IE, and GC-IE the general equations are as follows:

$$[\text{RS-IE}]\quad N_{e(v)} = \frac{n\,(2 - s)}{\frac{(1+s)^2[n(1-u)-1]}{2P} + \frac{(1-s)[n(1-v)(3+s)-1]}{2M}+2} \qquad \text{Eq. 2}$$

and

$$[\text{GC-IE}]\quad N_{e(v)}=\frac{2n\,(2 - s)}{\frac{(1+s)^2[n(1-u)]}{P} + \frac{(1-s)[n(1-v)(3+s)-1]}{M}+(3+s)(1-s)} \qquad \text{Eq. 3}$$

It can be observed that $N_{e(v)}$ increases linearly with the size of the sample (n) if the accession is regenerated prior to any loss of plants in the field and all original plants contribute with male and female gametes to the offspring. The equations are unrestricted with respect to values of s and can be adapted to cases where gametes are manipulated through hand-pollination.

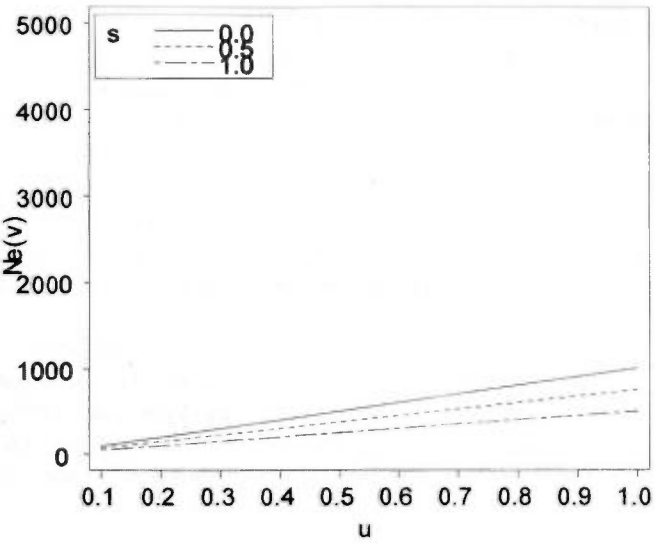

Fig. 1. Plot of the $N_{e(v)}$ against the proportion of functional parents (u) for three values of the natural rate of self-fertilization (s=0.0, 0.5 and 1.0) under random sampling (RS) and inbreeding equilibrium for Case 1 (seed parents: P=Nu and pollen parents: M=Nu) for n=N=1000.

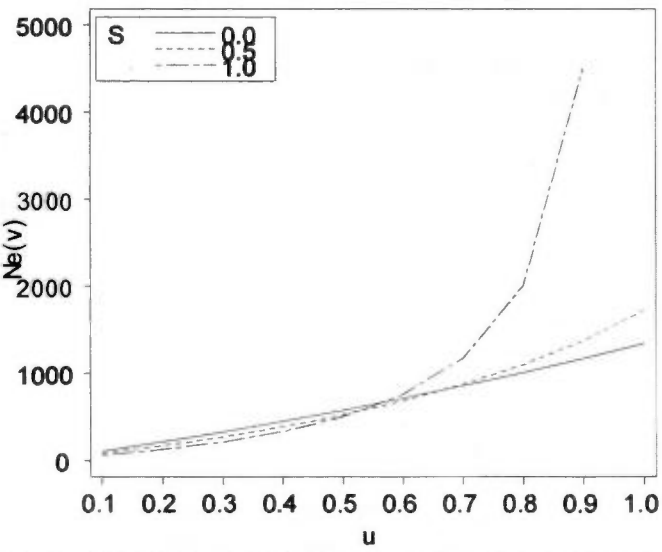

Fig. 2. Plot of the $N_{e(v)}$ against the proportion of functional parents (u) for three values of the natural rate of self-fertilization (s=0.0, 0.5 and 1.0) under female gametic control (GC) and inbreeding equilibrium for Case 1 (seed parents: P=Nu and pollen parents: M=Nu) for n=N=1000.

Eqs. 2 and 3 can be adapted to germplasm regeneration (and collecting) considering the effect of: (1) an arbitrary level of natural rate of self-fertilization (s), and (2) random sampling of seeds or female gametic control. We will examine the same two cases we studied when considering the two-stage model for random-mating species, that is, **Case 1**: loss of accessions given by u<1, with seed parents

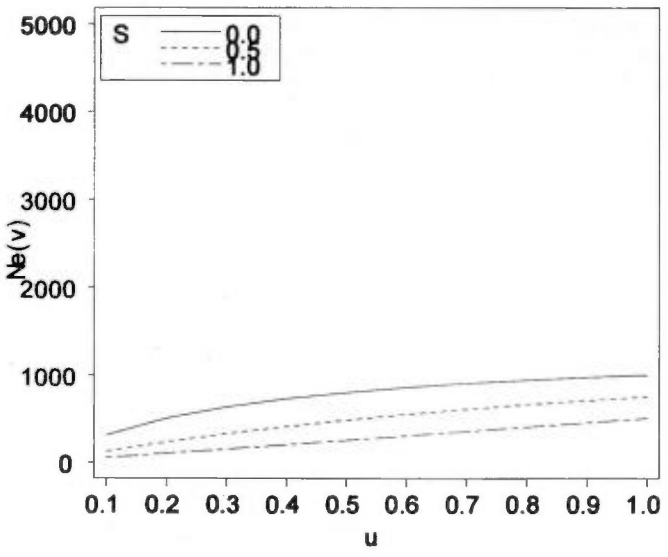

Fig. 3. Plot of the $N_{e(v)}$ against the proportion of functional parents (u) for three values of the natural rate of self-fertilization (s=0.0, 0.5 and 1.0) under random sampling (RS) and inbreeding equilibrium for Case 2 (seed parents: $P=Nu$ and pollen parents: $M=N$) for $n=N$=1000.

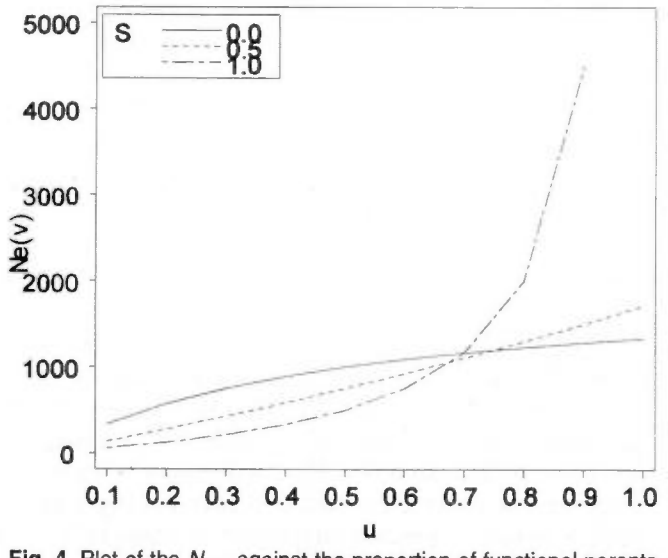

Fig. 4. Plot of the $N_{e(v)}$ against the proportion of functional parents (u) for three values of the natural rate of self-fertilization (s=0.0, 0.5 and 1.0) under female gametic control (GC) and inbreeding equilibrium for Case 2 (seed parents: $P=Nu$ and pollen parents: $M=N$) for $n=N$=1000.

$P=uN$, and pollen parents: $M=uN$, and **Case 2**: no loss of accessions but seeds taken from uN pollinated plants, that is, seed parents: $P=uN$ and pollen parents: $M=N$. We will also consider an additional effect given by the fact that the number of seed parents (number of collected seeds) that can be taken is equal to N (constant $n=N$) or arbitrary ($n\neq N$).

Figures 1 to 4 show results after adapting Eq. 2 (RS-IE) and Eq. 3 (GC-IE) to Case 1 and Case 2 for $n=N$=1000 (constant n) and different levels of natural rate of self-fertilization (s) and accession loss (u). In these Figures, $N_{e(v)}$ is plotted against u for three values of s (0.0, 0.5, and 1.0). In general, it is observed that GC gave higher values of $N_{e(v)}$ as compared to RS. Female gametic control combined with adequate levels of u (u>0.7) becomes an effective combination for maintaining values of $N_{e(v)}$ equal or higher than N (Figs. 2, 4), whereas with RS, values of $N_{e(v)}$ are always smaller than N (Figs. 1, 3).

For Case 1 and Case 2, RS of cross-fertilizer species (s=0.0) is less affected by accession deterioration than more self-fertilized species (s=0.5 and s=1.0) (Figs. 1 and 3) at any level of accession deterioration (u). However, for Case 1 and Case 2, GC of cross-pollinated species gave slightly higher values of $N_{e(v)}$ than more self-fertilized species only for severe values of accession deterioration; the reverse is true for mild accession loss (Figs. 2 and 4). Difficulties for maintaining $N_{e(v)}=N$ for RS are clearly shown in Figures 1 and 3 where it is seen that $N_{e(v)}=N$ is only achieved for high values of u and when the species approaches panmixia (s=0.0).

Case 1 with GC (Fig. 2) produces a rapid increase in $N_{e(v)}$ for autogamic species (s=1) when u>0.6; for Case 2 (Fig. 4) this happens at u>0.7. Note that $N_{e(v)}$ values of Case 2 tend to always be higher that those for Case 1 due to the larger pollen pool existing in Case 2 where $M=N$ as compared to $M=uN$ of Case 1. For Case 1 and Case 2, when seeds are taken randomly, (Figs. 1 and 3) accession regeneration is more efficient for panmictic species than for autogamic ones, under any level of accession deterioration. This is also true with female gametic control and severe deterioration loss (Figs. 2 and 4). Autogamic (and polymorphic) species with GC, on the other hand, lead to much higher $N_{e(v)}$'s in regeneration than allogamic materials, when deterioration of accession is mild (Figs. 2 and 4). Here, it is interesting to mention the findings of Vencovsky and Crossa (1999a) for germplasm collecting as opposed to germplasm regeneration. They found that, for attaining adequate $N_{e(v)}$'s, assuming inbreeding equilibrium and even with GC, it is much more difficult to collect seed from an autogamic (and polymorphic) species than to regenerate accessions of this type of species when deterioration of accessions is avoided (u>0.8). The reverse is true for panmictic species.

In summary, the positive effect of female gametic control for achieving large values of $N_{e(v)}$ is clearly observed in Figures 1-4. Under GC, autogamic species largely benefit from keeping values of u as high as possible.

Eqs. 2 and 3 can also be adapted for clarifying questions about the required sample size (n) for maintaining $N_{e(v)}=N$ when seed germination drops below a certain level or when studying possible advantages of GC vs RS systems under Cases 1 and 2. For this situation we consider any arbitrary value of n ($n\neq N$). Values of $N_{e(v)}$ for regenerating a hypothetical population of N=500 plants in the field for Cases 1 and 2 under RS and GC for various values of accessions loss (u), natural rate of self-fertilization (s), and sample size (n) are shown in Table 3.

In general, results agreed with those previously described, that is, values of $N_{e(v)}$ for Case 2 are always equal to,

Table 3. $N_{e(v)}$ when regenerating N=500 plants and harvesting an arbitrary number of seeds (n) for different values of accession loss or proportion of functional parents (u) and natural rate of self-fertilization (s). Cases 1 and 2 with random sampling of seeds (RS), female gametic control (GC) and inbreeding equilibrium (IE)

			Case 1 (Seed P=uN; Pollen M=uN)		**Case 2** (Seed P=uN; Pollen M=N)	
u	s	n	RS-IE	GC-IE	RS-IE	GC-IE
0.5	0.0	500	250	286	400	500
0.5	0.0	1000	334	364	667	800
0.5	0.0	3000	429	445	1201	1334
0.8	0.0	500	400	500	471	616
0.8	0.0	1000	667	800	890	1144
0.8	0.0	3000	1201	1334	2184	2668
0.9	0.0	500	450	581	487	643
0.9	0.0	1000	819	1029	948	1242
0.9	0.0	3000	1801	2118	2574	3275
0.5	0.5	500	188	261	240	375
0.5	0.5	1000	250	308	353	481
0.5	0.5	3000	322	350	515	590
0.8	0.5	500	300	546	329	649
0.8	0.5	1000	501	800	586	1044
0.8	0.5	3000	901	1161	1221	1756
0.9	0.5	500	338	684	353	750
0.9	0.5	1000	614	1137	668	1334
0.9	0.5	3000	1351	2038	1638	2770
0.5	1.0	500	125	250	125	250
0.5	1.0	1000	167	250	167	250
0.5	1.0	3000	214	250	214	250
0.8	1.0	500	200	1000	200	1000
0.8	1.0	1000	334	1000	334	1000
0.8	1.0	3000	601	1000	601	1000
0.9	1.0	500	225	2250	225	2250
0.9	1.0	1000	410	2250	410	2250
0.9	1.0	3000	901	2250	901	2250

or higher than, those obtained for Case 1. For small accession loss (large u), it is easier to regenerate more autogamic species (larger $N_{e(v)}$ are attained), whereas for large accession loss (small u), it is easier to regenerate more panmictic species. For an autogamic species (s=1.0) values of $N_{e(v)}$ for Case 1 and Case 2 are equal.

It is interesting to note that collecting n=N=500 seeds from 250 plants (u=0.5) by taking two seeds per plant (GC) of a panmictic species (s=0), but using pollen from all M=N=500 plants (Case 2), the required condition $N_{e(v)}$=N =500 is achieved. On the other hand, collecting n=500 seeds from 250 plants (u=0.5) by taking 2 seeds per plant (GC) of an autogamic species (s=1) (Case 1=Case 2) the required condition is never achieved because $N_{e(v)}$ is only (½)N=250. As expected, this result indicates that severe accession deterioration has more impact on autogamic species than on panmictic species. If pollen from all 500 plants from a panmictic species (s=0) is used and 12 seeds (GC) are taken from only 250 plants (u=0.5) (n=3000) (Case 2) $N_{e(v)}$=2.67N=1335 and $N_{e(v)}$=2.4N=1200 with RS (Table 3). For Case 1, where the pollen pool is smaller than for Case 2, with GC the condition $N_{e(v)}\geq N$ is only achieved for smaller accessions loss, u=0.8, than that allowed for Case 2, u=0.5.

For Case 1, a 10% accession loss (u=0.9) and 3000 seeds collected under GC from a panmictic species (s=0), $N_{e(v)}$=2118>4N. For Case 2, a 20% accessions loss (u=0.8) of a panmictic species (s=0) and collecting 3000 seeds using GC will produce $N_{e(v)}$=2668>5N. The combination of female gametic control with the possibility of having a larger pollen pool, as in Case 2, produces important increases in $N_{e(v)}$ for u=0.8 and 0.9.

References

Cockerham, C.C. 1969. Variance of gene frequency. Evolution 23:72-84.

Crossa, J. 1989. Methodologies for estimating the sample size required for genetic conservation of outbreeding crops. Theor. and Appl. Genet. 77:153-161.

Crossa, J., C.M. Hernandez, P. Bretting, S.A. Eberhart and S. Taba. 1993. Statistical genetic considerations for maintaining germplasm collections. Theor. and Appl. Genet. 86:673-678.

Crossa, J., S. Taba, S.A. Eberhart, P. Bretting and R. Vencovsky. 1994. Practical considerations for maintaining germplasm in maize. Theor. and Appl. Genet. 89:89-95.

Crossa, J. and R. Vencovsky. 1994. Implication of the variance effective population size on the genetic conservation of monoecious species. Theor. and Appl. Genet. 89:936-942.

Crossa, J. and R. Vencovsky. 1997. Variance effective population size for two-stage sampling of monoecious species. Crop Sci. 37:14-26.

Crow, J.F. and M. Kimura. 1970. An Introduction to Population Genetics Theory. Burgess Publishing, Minnesota.

Frankel, O.H. and M.E. Soulé. 1981. Conservation and evolution. Cambridge University Press, Cambridge.
Gale, J.S. and M.J. Lawrence. 1984. The decay of variability. *In* Crop Genetic Resources: Conservation and Evolution (J.H.W. Holden and J.T. Williams, eds.). Allen and Unwin, London.
Gregorius, H-R. 1980. The probability of losing an allele when diploid genotypes are sampled. Biometrics 36:643-652.
Hernandez, C.M. and J. Crossa. 1993. A program for estimating the optimum sample size for germplasm conservation. J. Heredity 84:1.
Lawrence, M.J., D.F. Marshall and P. Davies. 1995. Genetics of genetics conservation. I. Sample size when collecting germplasm. Euphytica 84:89-99.
Kimura, M. 1955. Random genetic drift in a multi-allelic locus. Evolution 9:429-435.
Lindgren, D., L.D. Gea and P.A. Jefferson. 1996. Loss of genetic diversity monitored by status number. Silvae Genet. 45(1):52-59.
Lindgren, D., L.D. Gea and P.A. Jefferson. 1997. Status number for measuring genetic diversity. Forest Genet. 4(2):69-72.
Marshall, D.R. and A.H.D. Brown. 1975. Optimum sampling strategies in genetic conservation. *In* Crop Genetic Resources for Today and Tomorrow (O.H. Frankel and J.G. Hawkes, eds.). Cambridge University Press, Cambridge.
Namkoong, G. 1986. Sampling for germplasm collections. Hort. Sci. 23(1) 79-81.
Schoen, D.J., J.L. David and T.M. Bataillon. 1998. Deleterious mutation accumulation and the regeneration of genetic resources. Proc. Nat. Acad. Sci. 95:394-399.
Vencovsky, R. 1978. Effective size of monoecious populations submitted to artificial selection. Brazil J. Genet. I(3):181-191.
Vencovsky, R. and J. Crossa. 1999a. Variance effective population size under mixed self and random mating with applications to genetic conservation of species. Crop Sci. (in press).
Vencovsky, R and J. Crossa. 1999b. Medidas de representatividade. Workshop O Melhoramento de Plantas na Virada do Milenio. Universidad Federal de Vicosa, MG, Brasil.
Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97-159.

Use of molecular markers to increase understanding of plant domestication and to improve the management of genetic resources; some examples with tropical species

C. Lanaud
Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), BIOTROP, BP 5035, 34032 Montpellier, France. Tel: +33 467615831; Fax +33 467615605; Email: lanaud@cirad.fr

Summary

Use of molecular markers to increase understanding of plant domestication and to improve the management of genetic resources; some examples with tropical species

Large germplasm collections are often available for use by plant breeders to improve most cultivated species. However, a very small percentage of genotypes are actually used. There are many reasons for the under-exploitation of genetic resources, one of which is a lack of characterization of plant material in the collection. Molecular markers could aid in the management of genetic resources and could lead to greater benefits being obtained from them. Even if molecular markers are not directly involved in the selection for traits of interest, they increase knowledge of domestication, the structure of variability within and between species, the relationships between populations, and the geneflow between wild and cultivated species. They can also provide a tool for early screening of genetic resources. These aspects are illustrated by examples taken from the use of molecular markers for tropical crops. However, there is a need to simplify the use of molecular markers in order to facilitate their transfer to tropical countries. These countries possess the greatest range of genetic resources and the availability of molecular markers near to the places where collections are maintained would be an important tool in optimizing the conservation and use of genetic resources.

Key words: Banana, cocoa, domestication, genetic resources, molecular markers, sugar cane

Résumé

Utilisation des marqueurs moléculaires pour mieux comprendre la domestication des plantes et gerer les ressources génétiques. Quelques exemples pris chez les plantes tropicales

Les sélectionneurs disposent souvent de larges collections pour améliorer la plupart des espèces cultivées. Malgré cela, un très faible pourcentage de plantes issues de ces collections sont réellement utilisées. Une des raisons qui expliquent cette sous exploitation des ressources génétiques est une mauvaise caractérisation du matériel végétal en collection. Les marqueurs moléculaires peuvent représenter un outil de choix pour aider à gérer et à mieux exploiter les ressouces génétiques: s'ils n'ont pas un interêt direct en sélection, ils permettent, en revanche, une meilleure compréhension des processus de domestication, de la structuration de la variabilité au sein de ou des espèces, des apparentements entre populations ou des flux de gènes entre espèces sauvages et cultivées. Ces aspects sont illustrés par des exemples pris chez les plantes tropicales. La simplification et le transfert possible des méthodes de marquage génétique en régions tropicales sont discutées. En effet, ces régions détiennent la majorité des ressources génétiques et la mise à disposition de l'outil moléculaire pour les sélectionneurs de tous les pays représente un enjeu important pour la conservation et l'utilisation optimum des ressources génétiques.

Resumen

Uso de marcadores moleculares para una mayor comprensión de la domesticación de las plantas y una mejora del manejo de los recursos genéticos; algunos ejemplos con especies tropicales

Muy a menudo se dispone de grandes colecciones de germoplasma que los fitomejoradores utilizan para mejorar las especies más cultivadas. Sin embargo, se usa realmente un porcentaje muy limitado de genotipos. Una de las principales causas de la escasa explotación de los recursos genéticos es la falta de caracterización del material vegetal en la colección. Los marcadores moleculares pueden ser útiles para el manejo de los recursos genéticos aportando mayores beneficios. Si bien los marcadores moleculares no están comprometidos directamente con la selección de los rasgos de mayor interés, ellos aumentan los conocimientos sobre la domesticación, la estructura de la variabilidad en y entre las especies, las relaciones entre las poblaciones y el flujo de genes entre especies silvestres y cultivadas. Asimismo pueden representar un instrumento fundamental para el control precoz de los recursos genéticos. Ejemplos del uso de los marcadores moleculares para cultivaciones tropicales ilustran tales aspectos. Sin embargo, hay que simplificar el uso de dichos marcadores para facilitar su traslado a los países tropicales. Éstos poseen una gama más amplia de recursos genéticos y el hecho de tener marcadores moleculares cerca de los lugares donde se mantienen las colecciones sería un instrumento importante para optimizar la conservación y el uso de los recursos genéticos.

Introduction

Over several centuries of plant cultivation and domestication, farmers have played an important role in maintaining the genetic diversity of cultivated plants. Although farmers used mass selection they often maintained genetic mixing with neighbouring wild species and so a certain diversity among cultivated varieties was perpetuated and voluntarily maintained (Pernes *et al.* 1993).

The 20th century has been marked by the development of high-yielding monocultures, which together with increasing populations and reduced amounts of arable land and forest, has contributed greatly to the loss of biodiversity. Thus, a 1996 FAO report stated that in the United States, many of the crop varieties described between 1804 and 1904 are no longer in existence; for example, only 14% of apple varieties still exist, 5% of cab-

bage varieties, 9% of maize varieties, 6% of pea varieties and 19% of tomato varieties. In order to prevent further genetic erosion and preserve biodiversity, the international community has organized plant collections that contain increasing numbers of economically important plants. FAO (1996) reported a total of 5554 505 accessions held in national and international collections; of these 593 000 accessions are held in the CGIAR Centres. The National Genebank of China alone houses 358 000 accessions (Li *et al.* 1998). Plant breeders now have available large collections of most cultivated species and these are constantly being enriched with new specimens. In spite of this, the genetic diversity actually used by breeders is in general quite restricted. In China, only 3 to 5% of conserved accessions are used in breeding programmes (FAO 1996).

There are many possible reasons for this under-exploitation of genetic resources. One is probably the poor characterization of the plant material in the collections and a lack of understanding of the relations existing between plant populations. Moreover, some genotypes or related wild species are difficult to use although they could constitute an important reservoir of genes. Using a few tropical plants as examples, we will see how molecular markers can supply basic information that can help to better manage and exploit genetic resources.

Strategies employed

Modern varieties are sometimes the result of hundreds or thousands of years of domestication of plants that either no longer exist or have evolved so much that they can no longer be recognized by using only morphological analyses. Sometimes ancestors are even more difficult to identify because presently cultivated varieties can be the result of often poorly understood inter-specific crossings, as is the case for banana.

Banana is an important economic food crop throughout tropical regions. The different varieties of banana have many uses and can be eaten raw or cooked or used to make textiles and beer, as in East Africa. Most of the plantations that grow bananas for export cultivate only one variety. At the turn of the century, the variety 'Gros Michel' was used, then with the appearance of Panama disease in around 1950, due to *Fusarium oxysporum*, farmers changed to Cavendish, a closely related but resistant variety. Cavendish clones are at present coming under attack from a new strain of Panama disease and plant improvement programmes have been set up to deal with this and other pests such as Cercosporiose and nematodes.

Banana cultivars originated in Southeast Asia and are mainly triploid varieties. Some of them have a multi-specific origin; this involved principally two species of the genus *Musa* - *M. acuminata* (AA genome) and *M. balbisiana* (BB genome) (Simmonds and Sheferd). *Musa acuminata* is itself constituted of subspecies differentiated by their chromosomal organization (Sheferd 1987). Most wild bananas are fertile, produce gametes and are diploid. Bananas would have been domesticated initially by hybridization between several of these subspecies and by the acquisition of both the parthenocarpy and sterility favoured by the structural heterozygosity of these diploid cultivars (Fig. 1). The appearance and domestication of more vigorous triploid forms arose from the subsequent pollination of the diploid cultivars, which often produce non-reduced gametes. Pollen could have come from *M. acuminata* itself or from *M. balbisiana* with which it co-exists, particularly in India and Burma. Triploid cultivars were then dispersed around the world, particularly into Africa where some arrived in Madagascar and the Comoros.

The triploid nature of current cultivars makes them difficult to improve. One strategy for widening their narrow genetic base would be to create new triploid varieties starting from wild diploids or diploid *M. acuminata* cultivars (Bakry *et al.* 1990). The genetic origins of the current cultivars have been investigated using molecular markers to obtain a better idea of which genetic resources to use while retaining the same commercial characters in new varieties.

For this, nuclear or cytoplasmic RFLP probes were used to analyze the diversity of wild and cultivated AA, BB or AB diploids and the various AAA, AAB or ABB triploid cultivars. Banana has the particularity of possessing chloroplastic DNA inherited from the female parent and mitochondrial DNA inherited from the male parent (Fauré *et al.* 1993). This enables more precise information to be obtained on both the maternal and paternal contribution and history.

The nuclear diversity of the wild *M. acuminata* (Aaw) can be arranged into four groups within which the diploid cultivars can be replaced (Fig. 2) (Carreel *et al.* 1994). These are distributed in continuous fashion between the four groups, which shows that their formation involved

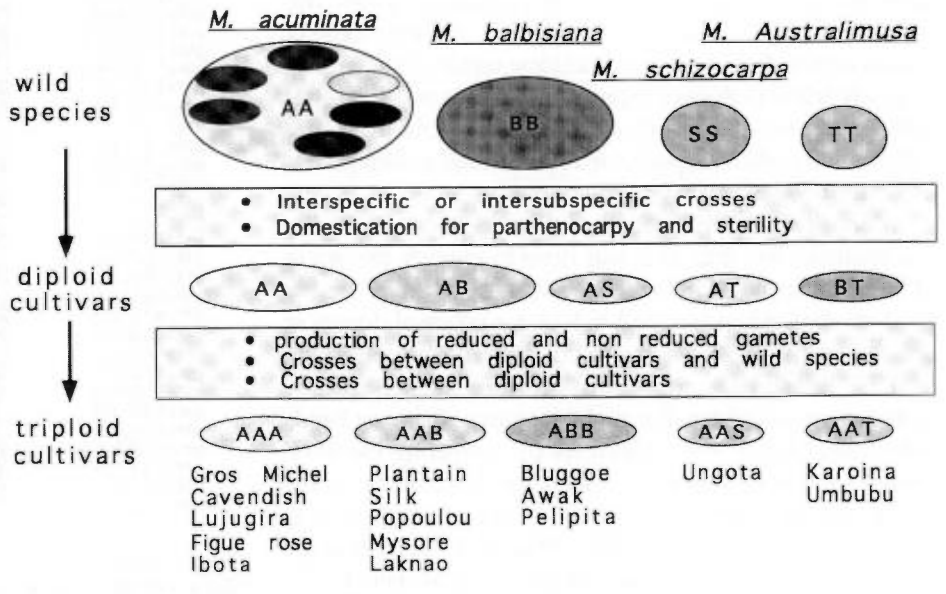

Fig. 1. Schematic representation of banana domestication.

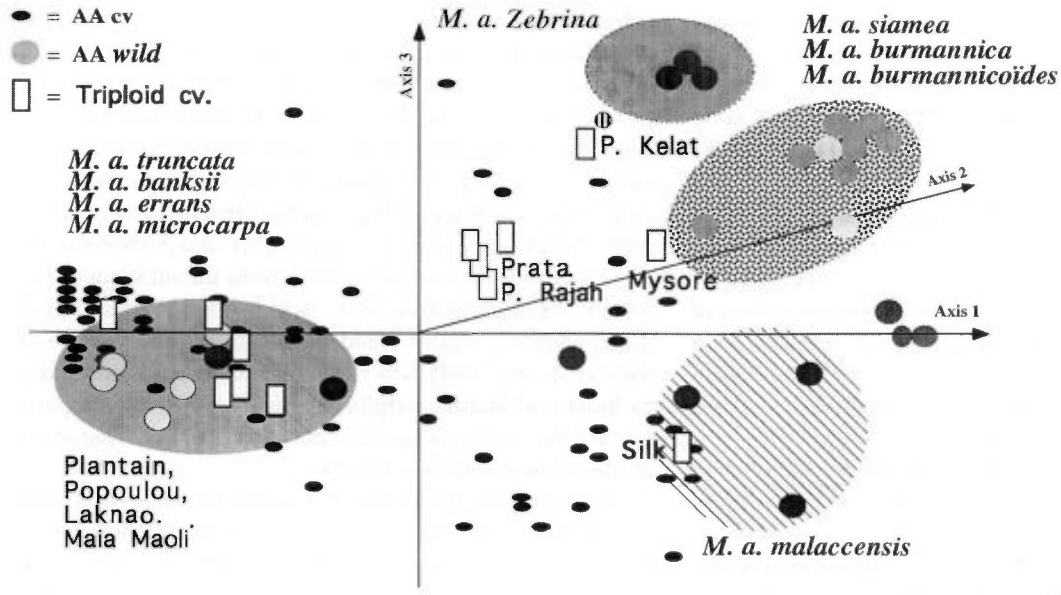

Fig. 2. Nuclear structuration of wild *M. acuminata* subspecies (AAw), diploid (AAcv) and triploid cultivars. Plan 1-3 of an AFTD (Factorial Analysis on Matrix Distances) realized on Jacquard distances. The pool '*M. a. siamea, M. a. burmannica, M. a. burmannicoides*', isolated on the axis 2, has been also reported on this figure (according to Carreel 1994).

hybridization between different subspecies of *M. acuminata*. A great number of them are found around the focus of *banksii/microcarpa/errans*, which emphasizes the importance of these subspecies in the first stages of domestication.

In the same way, the triploid cultivars can be replaced on this diagram to investigate their relation with the diploids. We were thus able to see that cultivars of the plantain group AAB, which produce the cooking bananas widely consumed in Africa, are closely related to the banksii and errans groups. This suggests that their A genome could come from these subspecies. In general, the cultivars close to the banksii focus have more starchy fruits and those close to the malaccensis focus have sweet fruits, which might suggest that these two subspecies groups were involved in the domestication of these characters (Carreel 1994).

Further information on the maternal and paternal origins of the cultivars can be obtained by comparing the highly variable cytoplasmic profiles of wild subspecies with those of diploid and triploid cultivars (Carreel 1994). Thus, the cytoplasm of plantains is associated with that of wild bananas of the banksii subspecies. This confirms the involvement of this subspecies, and with those of around twenty diploid cultivars, including certain AB's probably originally linked to these cultivars. In general, the two subspecies banksii and errans have cytoplasm associated with the majority of cultivars and could have played an important role in their genesis and in their domestication for parthenocarpy.

More precise identification of the ancestors of cultivars is sometimes possible. This was done to identify the origin of the AAA triploid cultivar 'Gros Michel', which is used to produce dessert bananas, and to look for potential donors of non-reduced and reduced gametes (Raboin *et al.* unpublished).

First, the diploid ancestor that provided the non-reduced gamete was looked for. The diploids sharing the most bands with 'Gros Michel' were looked for by comparing the RFLP profiles of 36 nuclear probes mapped and distributed over the whole genome. Three clones were thus identified which came from Indonesia. These have similar RFLP profiles and have the totality of their genome in common with 'Gros Michel' (Fig. 3). Moreover, for one of these clones, these results were confirmed by enzymatic data and by variations of the rDNA IGS, showing profiles totally incorporated in those of the triploid cultivar.

Next, the diploid ancestor that provided the reduced gamete was investigated. This clone should bring additional bands to the triploid cultivar. Two clones from Thailand fulfilling this criterion and which had the same genotype were identified. Therefore, 'Gros Michel' can be entirely explained by the fusion of a non-reduced

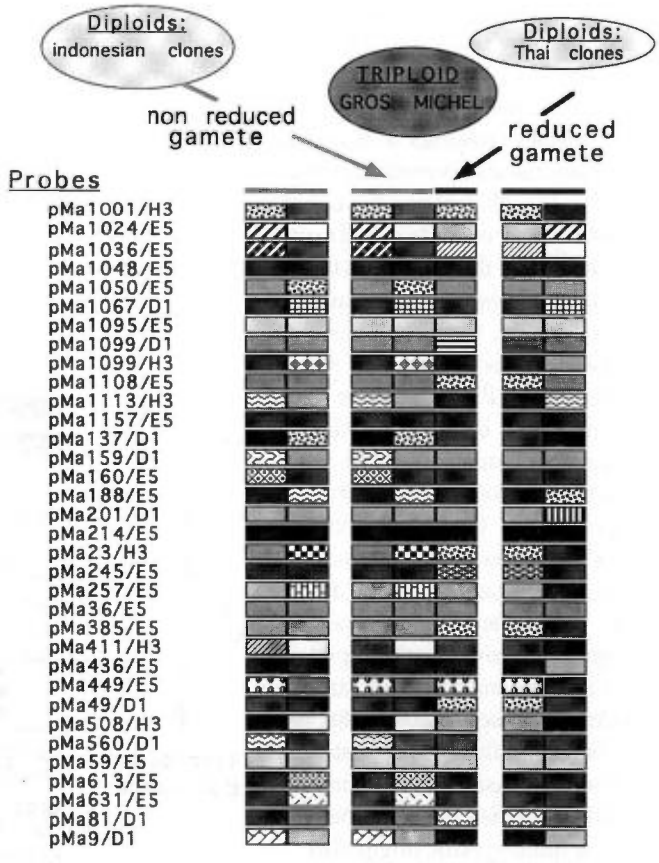

Fig. 3. Genotype representation of diploid ancestors of the triploid "Gros Michel" banana cultivar. The rectangles represent the several alleles present in the diploid and triploid clones.

gamete from an Indonesian diploid clone, and a reduced gamete from a diploid Thai clone. This exercise was carried out on other important cultivars and enabled possible ancestors to be identified, or at least those in the collections that were closest.

Taken together, all this information increases understanding of how diploid then triploid bananas became domesticated. According to Carreel (1994), the primary centre of domestication would have been located in Papua New Guinea and/or the Philippines; then, the first domesticated genotypes would have been transported to neighbouring regions where they met other *M. acuminata* subspecies or other species similar to *M. balbisiana*. Hybrid forms appeared, corresponding to the diploid and triploid cultivars that were later to become widespread in Africa and in all the world (Fig. 4).

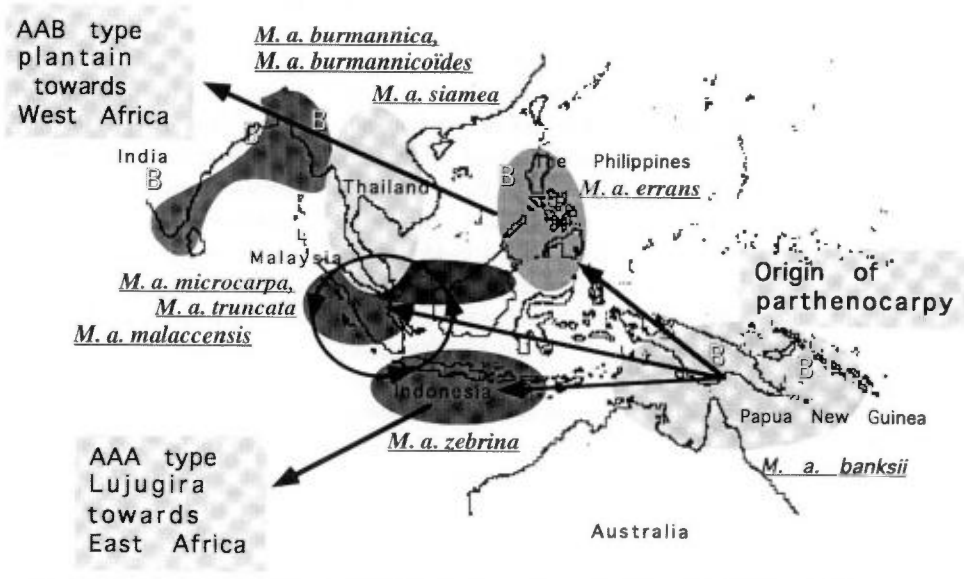

Fig. 4. Hypothesis of domestication of banana cultivars proposed by Carreel (1994).

This information allows us to better target the subspecies or diploid cultivars for use in breeding programmes to create new improved triploid varieties from an enlarged genetic base. The information obtained jointly from the genetic diversity study and the genome analysis can also be used to generate tools for making better use of genetic resources for targeted characters. With the development of genetic maps, which has progressed rapidly in the last ten years, it has become possible to identify the regions of the genome (QTLs) involved in the characters of interest and to evaluate their importance, stability and diversity (Paterson *et al.* 1988; Stuber *et al.* 1992; Tanksley *et al.* 1996; Young 1996). The markers that closely delimit these QTLs enable the presence of these characters to be checked, either during the selection stages or directly in certain populations if the genetic mixing has remained insufficient to break the genetic links between the markers and the characters (Jannoo *et al.*, in press).

This favourable situation could occur for the selection of quality characters in the cocoa tree, whose fruits contain the cocoa beans, which after fermentation, drying and roasting are used to make chocolate. Indeed, the Maya in Central America began the domestication of the cocoa tree over 2000 years ago. It was used not to make chocolate but in a mix with corn and pepper to make a high-energy food (Paradis 1979). A very particular type of cocoa tree – the Criollo – has long remained the only variety cultivated throughout Central America. Criollo varieties are currently used to make the finest chocolate. Other wild populations originate from the Amazon, Orinoco and Guyana regions. These South American populations (Forastero group), possess agronomic characters that are often more interesting than those of Criollo, in particular for disease resistance, although they generally produce a lower quality chocolate. Over the last few decades Criollo has been replaced in many plantations by Trinitario – a hybrid between Criollo and Forastero – which is more vigorous and productive. Considerable genetic mixing has taken place between these two groups.

A diversity study, using RFLP and RAPD molecular markers, was carried out on 300 clones representing most of the populations present in the international collections (Laurent *et al.* 1994; N'Goran *et al.* 1994). The Criollo group particularly is represented by cultivated varieties or those selected in the last few decades. Diversity was clearly structured into Forastero as one group and Criollo as the other, with considerable genetic diversity within the two groups (Fig. 5). More recently, a new study of Criollo diversity has been developed (Motamayor *et al.* 1997) and a very narrow genetic base at the origin of actual cultivars can be identified. The current, recently selected varieties of Criollo, of which most are partially heterozygous, are variable and probably resulted from more or less important introgressions of Forastero genes in an ancestral Criollo type. In this case, the quality characters, widespread in the Criollo and Trinitario groups, would be of a nearly unique origin.

These results have important consequences for the selection of Criollo quality characters. A saturated genetic map of *T. cacao* L. has been produced (Lanaud *et al.* 1995) and analysis of quality QTLs is in progress. Given the small number of generations leading to the hybrid forms, the markers very close to the QTLs should enable us to identify the hybrid forms that carry the quality characters in homozygous or heterozygous form. This could greatly facilitate the use of genetic resources from the Trinitario and Criollo groups for the selection of these characters.

Sometimes the complexity of plant genomes is such that only the use of markers can lead to significant progress in the understanding and exploitation of genetic resources for plant improvement. This is the case for sugar cane. The first interspecific hybrids, similar to those presently cultivated, were created at the beginning of the century. This was achieved by successive hybridization and backcrossing between

two species - *Saccharum officinarum* ($2n = 80$), the noble sweet species and *S. spontaneum* ($2n = 40$ to 128), a wild species.

The genome of these cultivars is highly complex; it contains over 100 chromosomes, of which a minority, less than twenty, come from *S. spontaneum* (Bremer 1961). Chromosomes of *S. spontaneum* are particularly important because they confer vigour, adaptation and resistance to disease in these modern, interspecific varieties. In the last decades, new cultivars have continued to be produced through crossings between existing varieties, but a yield plateau seems to have been reached. Current improvement programmes are again using wild species of the *Saccharum* genus or related genera such as *Erianthus* to enlarge the genetic basis. However, as in classical analysis, genetic analysis of the characters is very difficult. In order to gain better control of the construction of new varieties, where only a small number of chromosomes from the wild species need be conserved, current strategies are based on several different aspects of genetic marking: *in situ* hybridization and the search for specific repeated sequences, genetic diversity studies, genome mapping and exploitation of the synteny. *In situ* hybridization enables the chromosomes of each species present in the hybrids to be visualized. It also enables analysis of the genome structure of the varieties by identifying both the chromosomes of each species and the recombinations between them (D'Hont *et al.* 1996). By associating the hybridization of probes specific to certain chromosomes, as it is the case for the rDNA probes (18S or 5S) which hybridize only on a single base chromosome, it becomes possible to decipher the structure of these complex genomes by counting very precisely the number of examples of each of these chromosomes provided by the two species (D'Hont *et al.* 1998).

In addition, genetic marking of these chromosomes, as well as location of the genome regions involved in the characters of interest (QTLs), has begun. Genetic diversity studies carried out on species of the genus *Saccharum*

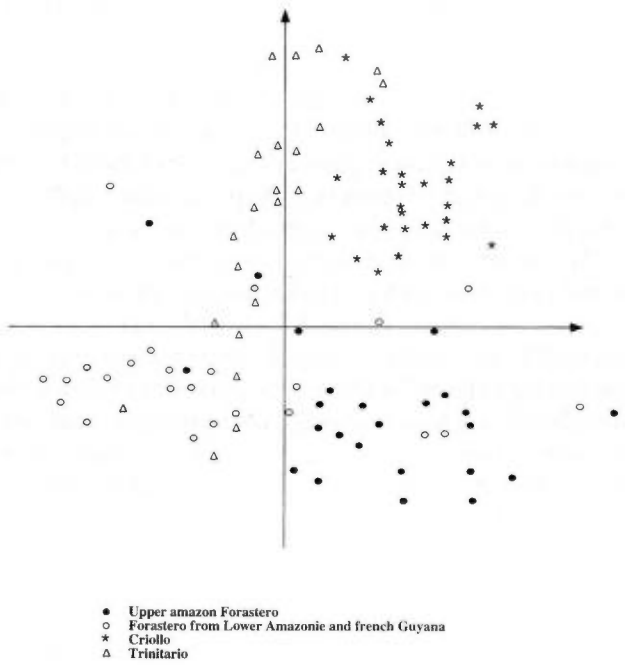

Fig. 5. Structuration of diversity in *Theobroma cacao* L. Plan 1-2 of a FAC (Factorial Analysis of Correspondance) based on RAPD data provided by 49 polymorphic bands (N'Goran *et al.* 1994).

reveal a clear differentiation of the different species (Lu *et al.* 1994) and have shown the possibility of identifying specific markers for these species. These specific markers were used to map the chromosomes present in the cultivars. Thus, by studying the self-fertilization of one variety, 99 linkage groups were obtained. The homologous chromosomes could be identified using RFLP probes located simultaneously on several groups and a composite map has been constructed (Grivet *et al.* 1996). In most cases, the use of markers specific to each of these species enabled the identification of the S. *spontaneum* or S. *officinarum* origin of each cosegregation group, as well as the recombinations possible between the two species. Research is currently being carried out to find the QTLs for the characters of principal interest for plant improvement. Because of the complexity of the genome and the redundancy of the alleles in *S. officinarum*, the QTLs will be more easily detected on the part of the genome corresponding to *S. spontaneum*, where each chromosome has a small number of copies and can be located by specific markers.

Marking of characters of interest in sugar cane is also based on information obtained by analysis of the simplest genomes making use of the phenomenon of synteny (Grivet *et al.* 1994; Dufour *et al.* 1997). Sugar cane belongs to the family *Poaceae*, tribe *Andropogonae*, to which maize and sorghum also belong. Other species such as rice or wheat are more distant but still belong to the same family. The genomes of these species are of variable size, although less complex, and possess a certain degree of conservation between themselves. This will enable the information that they provide to be used separately and to be transposed to related species. More than 80% of maize RFLP probes hybridize well on sugar cane and sorghum, thus it is possible to compare their genome structures and to locate the homologous zones. In particular, a very good homology can be observed between the genomes of sugar cane and sorghum (Dufour *et al.* 1997). This phenomenon of synteny has been used directly, for example, to saturate a region of the sugar cane genome where a major gene controlling resistance to rust had already been identified (Daugrois *et al.* 1996). Indeed, localized markers in homologous regions of other species have been used in sugar cane to delimit this gene.

In this complex species, where only molecular markers enable us to truly tackle the analysis of character genetics, all this information taken together can be used to exploit existing genetic resources more effectively and to monitor the construction of new varieties by introgressions made from wild species in a controlled manner.

Discussion and conclusion

Molecular markers have provided powerful tools to analyze the genetics of cultivated species and to decipher the genome structure even for plants with a complex genome. Therefore, they can help to solve problems for which classic selection was inefficient. They have also allowed the domestication processes to be traced and have focused the use of genetic resources on the improvement of new varieties.

Tropical regions constitute an important genetic reservoir for a very large number of plants cultivated around the world, including

those cultivated in temperate regions. Despite the local and international collections that have been built up, it is in these regions that most of the genetic resources useful to the whole planet are held. It is, therefore, necessary for these countries to have effective tools available on site to manage and exploit genetic resources, and to complement morphological observations. Up to the present, molecular markers have been essentially developed and used in northern countries. Their possible transfer to tropical countries, near to the places where collections are maintained, has now became an important issue. This transfer implies the development of simple and reproducible markers.

Despite the fact that these techniques have already been hugely simplified, notably with the development of PCR, progress remains to be made in providing plant breeders with easy-to-use tools. Consensus targeted PCR markers could provide such markers. To date, such consensus primers have been developed in most cases to amplify chloroplastic and mitochondrial DNA (Demesure *et al.* 1995) and only a few, such as the primers developed to amplify rDNA 5S (D'Hont *et al.* 1995), have been defined in the nucleus. Such universal markers have to be developed in order that they become available for all plant studies. The growing number of sequences presented in databases, notably for model species such as *Arabidopsis* and rice, should enable location of the conserved regions where such universal markers could be defined for a great number of these genes.

With the new objectives of agriculture: creating varieties that are more diversified, more resistant to disease, requiring fewer inputs and compatible with the local ecology, a facilitated access to genetic resources becomes of utmost importance. The progress that can be made with molecular markers for the characterization and exploitation of genetic resources justifies the efforts that must still be made so that they can be truly integrated into the toolbox of the plant-improvement scientist.

Acknowledgements

I thank A. D'Hont, P. Lagoda and J.C. Glaszmann for their helpful comments.

References

Bakry, F., J.P. Horry, C. Teisson, H. Tezenas du Montcel and J. Ganry. 1990. L'amélioration génétique des bananiers à l'IRFA/CIRAD. Fruits, n° special banane, 25-40

Bremer, G. 1961. Problems in breeding and cytology of sugar cane. Euphytica 10:121-258.

Carreel, F. 1994. Etude de la diversité génétique des bananiers (genre *Musa*) à l'aide des marqueurs RFLP. Thèse I.N.A. Paris-Grignon.

Carreel, F., S. Fauré, D. Gonzalez-De-Leon, P.J.L. Lagoda, X. Perrier, F. Bakry, H. Tezenas-Du-Montcel, C. Lanaud and J.P. Horry. 1994. Evaluation de la diversité génétique chez les bananiers diploïdes (*Musa* sp.). Genet. Sel. Evol. 26 (suppl. 1) 125-136

Daugrois, J.H., L. Grivet, D. Roques, J.Y. Hoarau, H. Lombard, J.C. Glaszmann and A. D'Hont. 1996. A putative major gene for rust resistance linked with an RFLP marker in sugar cane cultivar R570. Theor. Appl. Genet. 92:1059-1064.

Demesure, B., N. Sodzi and R.J. Petit. 1995. A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants. Mol. Ecol. 4:129-131.

D'Hont, A., P.S. Rao, P. Feldmann, L. Grivet, N. Islam Faridi, P. Taylor and J.C. Glaszmann. 1995. Identification and characterisation of sugarcane intergeneric hybrids, *Saccharum officinarum* x *Erianthus arundinaceus*, with molecular markers and DNA *in situ* hybridisation. Theor. Appl. Gent. 91:320-326

D'Hont, A., L. Grivet, P. Feldmann, P. Rao, N. Berding and J.C. Glaszmann. 1996. Characterisation of the double genome structure of modern sugarcane cultivars (*Saccharum* spp.) by molecular cytogenetics. Mol. Gen. Genet. 250:405-413.

D'Hont, A., D. Ison, K. Alix, C. Roux and J.C. Glaszmann. 1998. Determination of basic chromosome numbers in the genus *Saccharum* by physical mapping of ribosomal RNA genes. Genome 41:221-225.

Dufour, P., M. Deu, L. Grivet, A. D'Hont, F. Paulet, A. Bouet, C. Lanaud, J.C. Glaszmann and P. Hamon. 1997. Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. Theor. Appl. Genet. 94:409-418.

FAO. 1996. Rapport sur l'état des ressources phytogénétiques dans le monde, FAO,Rome.

Fauré, S., J.L. Noyer, F. Carreel, J.P. Horry, F. Bakry and C. Lanaud. 1993. Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). Current Genet. 25:265-269.

Grivet, L., A. D'Hont, P. Dufour, P. Hamon, D. Roques and J.C. Glaszmann. 1994. Comparative genome mapping of sugar cane with other species within the Andropogoneae tribe. Heredity 73:500-508.

Grivet, L., A. D'Hont, D. Roques, P. Feldmann, C. Lanaud and J.C. Glaszmann. 1996. RFLP mapping in cultivated sugarcane (*Saccharum* spp.): Genome organization in a highly polyploid and aneuploid interspecific hybrid. Genetics 142:987-1000.

Jannoo, N., L. Grivet, A. Dookun, A. D'Hont and J.C. Glaszmann. 1999. Linkage disequilibrium among modern sugarcane cultivars. Theor. Appl. Genet., in press.

Lanaud, C., A.M. Risterucci, J.A.K. N'Goran, D. Clement, M.H. Flament, V. Laurent and M. Falque. 1995. A genetic linkage map of *Theobroma cacao* L. Theor. Appl. Genet. 91:987-993

Laurent, V., A.M. Risterucci and C. Lanaud. 1994. Genetic diversity in cacao revealed by cDNA probes. Theor. Appl. Genet. 88:193-198.

Li Yu, Jie Wang, Yongsheng Cao, Weidong Gao, Jiahe Fang and Xizhi Lou. 1998. The use of genetic resources in crop improvement: Lessons from China. Genet. Res. and Crop Evol. 45:181-186.

Lu, Y.H., A. D'Hont, F. Paulet, L. Grivet, M. Arnaud and J.C. Glaszmann. 1994. Molecular diversity and genome structure in modern sugarcane varieties. Euphytica 78:217-226.

Motamayor, J.C., A.M. Risterucci, V. Laurent, A. Moreno and C. Lanaud. 1997. The genetic diversity of criollo cacao and the consequences for quality breeding. 1e Congreso Venezolano del Cacao y su industria, del 17 al 21 de Noviembre, Maracay Estado Aragua, Venezuela.

N'Goran, J.A.K., V. Laurent, A.M. Risterucci and C. Lanaud. 1994. Comparative genetic diversity of *Theobroma cacao* L. using RFLP and RAPD markers. Heredity 73:589-597.

Paradis, L. 1979. Le cacao précolombien: Monnais d'échange et breuvage des dieux. Journ. d'Agric. Trad. et de Bota. Appl. XXVI, 3-4.

Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln and S.D. Tanksley. 1988. Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Pernes, J., D. Combes and J.M. Leblanc. 1993. Le mil. Pp. 159-197 *in* Gestion des ressources génétiques des plantes, Tome 1: monographies.

Sheferds, K. 1987. Translocations in *Musa* acuminata. *In* Genetic improvement of Bananas and Plantains, and the INIBAP Project on International *Musa* Germplasm Exchanges. INIBAP 16-18 March 1987. Bogota, Colombia.

Simmonds, N.W. and K. Sheferds. 1955. The taxonomy and origins of the cultivated bananas. J. Linn. Soc. (Bot.) 55:302-312.

Stuber, C.W., S.E. Lincoln, D.W. Wolff, T. Helentjaris and E.S. Lander. 1992. Identification of genetic factors contributing to heterosisin a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823-839.

Tanksley, S.D., S. Grandillo, T.M. Fulton, D. Zamir, Y. Eshed, V. Petiard, J. Lopez and T. Beck-Bunn. 1996. Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor. Appl. Genet. 92:213-224.

Young, N.D. 1996. QTL mapping and quantitative disease resistance in plants. Ann. Rev. Phytopathol. 34:479-501.

Organellar molecular markers applied for determination of genetic diversity of selected crop species

H. Augustyniak[1]✉ and S. Bartkowiak[2]

[1] Institute of Molecular Biology and Biotechnology, Department of Plant Molecular Biology, A. Mickiewicz University, 5 Miedychodzka St., 60-371 Poznan, Poland. Tel: +48 61 8615511; Fax: +48 61 8615596; Email: haaugust@main.amu.edu.pl
[2] Institute of Plant Genetics, Polish Academy of Sciences, 34 Strzeszynska St., 60-479 Poznan, Poland

Summary

Organellar molecular markers applied for determination of genetic diversity of selected crop species

The usefulness of organellar markers in determination of genetic diversity in different species, lines and hybrids of *Lupinus*, as well as in lines and hybrids of *Brassica napus*, was analyzed. The 2D-PAGE pattern of proteins isolated from mitochondria and chloroplasts of *Lupinus* showed the presence of species specific polypeptides. In mitochondria we found a different number of specific protein isoforms with the same molecular weight of about 35 kDa. In chloroplasts, the main differences in the distribution of polypeptides are observed in the 26-42 kDa region in two pI: one about 5.5-6 and the other eight. PCR and gel blot analysis of mitochondrial and total DNA isolated from *Lupinus* and *Brassica* revealed the usefulness of tRNA-like sequences for detection of mitochondrial DNA polymorphism. The presence and copy number of *Lupinus albus* mitochondrial plasmid-like DNAs were tested in different genotypes and hybrids. In general, the absent DNA molecules were observed to be those found in mitochondria in a smaller number of copies.

Key words: *Brassica napus*, DNA analysis, *Lupinus*, 2D-PAGE proteins

Résumé

Marqueurs moléculaires d'organites utilisés pour la détermination de la diversité génétique de certaines espèces cultivées

L'utilité des marqueurs spécifiques d'organites cellulaires dans la détermination de la diversité génétique dans différentes espèces, lignées et hybrides de *Lupinus*, ainsi que dans des lignées et des hybrides de *Brassica napus*, a été analysée. Le profil en électrophorèse bidimensionnelle (2D-PAGE) des protéines isolées de mitochondries et de chloroplastes de *Lupinus* indique la présence de polypeptides spécifiques à l'espèce. Dans les mitochondries, on trouve un nombre différent d'isoformes de protéines spécifiques avec le même poids moléculaire d'environ 35 kDa. Dans les chloroplastes, les principales différences dans la distribution des polypeptides sont observées dans la région de 26-42 kDa pour deux pI : l'un d'approximativement 5,5-6 et l'autre de 8. L'étude par PCR et analyse de Southern de l'ADN mitochondial et de l'ADN total isolés à partir de *Lupinus* et de *Brassica* révèle l'utilité des séquences de type ARNt pour la détection du polymorphisme de l'ADN mitochondrial. La présence et le nombre de copies d'ADN mitochondriaux de type plasmidique de *Lupinus albus* ont été évalués dans différents génotypes et hybrides. En général, on note que les molécules d'ADN absentes sont celles que l'on trouve en un plus petit nombre de copies dans les mitochondries.

Resumen

Marcadores moleculares de organelos aplicados en la determinación de la diversidad genética de los cultivos seleccionados

Se ha analizado la utilidad de marcadores de organelo en la determinación de la diversidad genética de distintas especies, líneas e híbridos del *Lupinus*, así como también en las líneas e híbridos de la *Brassica napus*. El modelo 2D-PAGE de proteínas aisladas de mitocondrios y cloroplastios del *Lupinus* ha evidenciado la presencia de polipéptidos específicos en las especies. En los mitocondrios se han encontrado cantidades distintas de isoformas protéicas específicas con el mismo peso molecular, 35 kDa. En los cloroplastios, las diferencias principales en la distribución de los polipéptidos se encuentran en un ámbito de 26 a 42 kDa en dos pI: uno cerca de 5.5-6 y otro igual a ocho. Los análisis mediante reacción en cadena de la polimerasa (RCP) y los análisis del gel trasferido a membranas realizados sobre el ADN mitocondrial y total aislado del *Lupinus* y de la *Brassica*, revelaron la utilidad de las secuencias parecidas al tRNA para la individuación del polimorfismo del ADN mitocondrial. Se han analizado la presencia y la cantidad de copias de los ADN mitocondriales como el plasmidio del *Lupinus albus* en varios genotipos e híbridos. En general, se ha observado que las moléculas de ADN ausentes eran las que se individuaron en los mitocondrios en un número inferior de copias.

Introduction

The identification and conservation of genetically diverse germplasm (nuclear as well as cytoplasmic) are considered essential for breeding programmes. In the assessment of genetic variation by plant breeders the evaluation of cytoplasmic variation has been largely neglected. A growing number of studies have demonstrated the usefulness of organellar markers in determining plant genetic diversity, even though the number of organellar markers known to date is relatively small. They include mostly the sequences of genes, intergenic sequences and repeated sequences.

We are interested in the elaboration of organellar molecular markers to enable the monitoring of particular cytoplasm in germplasm material and in hybrids. In addition, we expect to be able to distinguish fragments coming from mitochondrial or chloroplast DNA, and to determine to what extent and at what developmental stage they may interfere with data obtained by very common methods such as RAPD or RFLP (Lorenz *et al.* 1994; Aagaard *et al.* 1995; Dumolin *et al.* 1995; Jack *et al.* 1995)

This study characterized the lupin and rape cytoplasms with regard to the presence and expression of important organellar genes. Its objectives were to examine mitochondrial and chloroplast polypeptide variability in *Lupinus* species. The usefulness of other organellar sequences, such as mitochondrial repeats (tRNA-like) in characterization of various genotypes of *Lupinus* as well as *Brassica napus*, was also checked. *Lupinus* genotypes and hybrids were also tested for the presence and copy number of mitochondrial plasmid-like minicircular DNAs - K1 and K2 (Gorączniak and Augustyniak 1991).

Materials and methods

Plant material including different species, lines and hybrids of *Lupinus* as well as lines and hybrids of *Brassica napus* were used. *Lupinus: L. albus, L. luteus, L. angustifolius L. termis, L. vavilovii* and *L. mutabilis* represented by genotypes XM-5, LM-34, LM-13 and mutant line Mut 45 and its interspecific hybrids with species of the section *Albus* were obtained from the collection of E. Sawicka-Sienkiewicz at the Department of Plant Breeding and Seeds Production, University of Agriculture, Wroclaw, Poland. *Brassica napus*: cv. POLO, cv. MAR, cv. Samuraj (CMS line) and hybrids 9232 (CMS Ogura x POLO) 9234 (CMS Ogura x MAR) were from the Oil Crop Department, Plant Breeding and Acclimatization Institute, Poznan Poland.

Organellar proteins were extracted as described by Hurkman and Tanaka (1986). Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) was performed according to the procedure of Hochstrasser *et al.* (1988). Proteins were silver-stained as described by Heukeshoven and Dernick (1985). Gels, after staining, were read on Image 2-D Elite LKB-Program.

Total DNA for PCR analyses was extracted from 3-day-old seedlings by Floraclean Kit (Bio 101, Inc.) for lupin plants, and from 20-day-old plant leaves by Nucleon Phytopure Kit (Amersham) for *Brassica napus*. *Brassica napus* plants were grown on Murashige-Skoog medium supplemented with micro and macroelements, vitamins and proper organic compounds.

The PCR analyses with matrix DNA concentration ranging from 25 to 75 ng were carried out in a Biometra thermal cycler. EtBr and silver staining quantified DNA. Digestion with restrictases and labelling by the Feinberg method as well as hybridization analysis were performed according to established procedures (Sambrook *et al.* 1989)

Purified mitochondria and chloroplasts were isolated from lupin plants essentially as described by Karpińska and Augustyniak (1989) and Bookjans *et al.* (1984).

Results and discussion

There are many reasons for the rather limited use of genetic resources, one of which is the lack of proper characterization of plant material. Characterization should concern the genetic information stored in the nucleus and the cytoplasm. The latter has a share in all kinds of genetic diversity. Molecular markers could facilitate the characterization of genetic resources. In the case of developing organellar markers it should be realized that the contribution of organellar genomes to total cellular DNA varies among species. In the present study we characterized representatives of two families: Cruciferae – *Brasica napus* and Fabacae – *Lupinus*.

This investigation is also intended to be a preliminary study towards a better understanding of barriers to the crossing of lupin species. To this end, protein and DNA polymorphism of chloroplasts and mitochondria were studied, thereby obtaining data for two different sets of molecular markers. The 2-D PAGE pattern of proteins isolated from

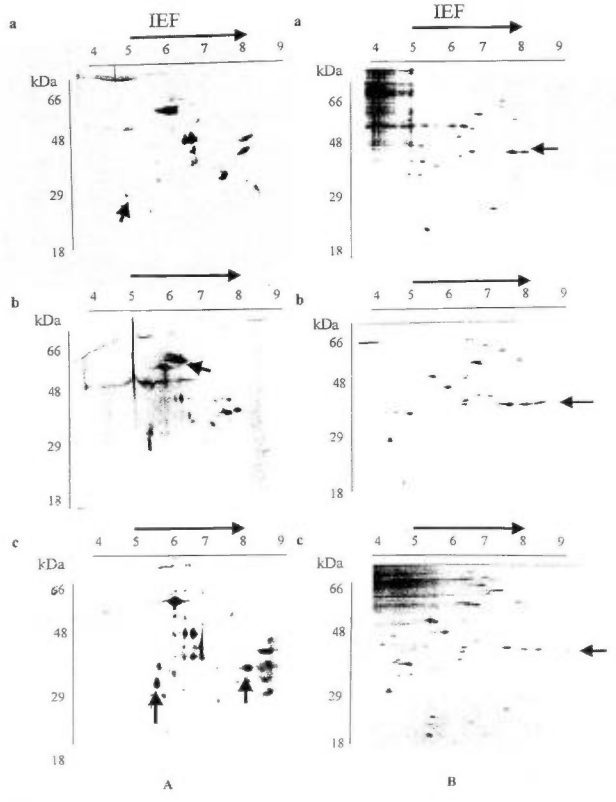

Fig. 1. 2-D PAGE pattern of (A) chloroplast and (B) mitochondrial polypeptides. The black arrows indicate marker polypeptides. a) *Lupinus albus*, b) *L. Luteus*, c) *L. angustifolius*. Chloroplast polypeptide useful for distinguishing species is indicated with a light arrow

mitochondria and chloroplasts shows the presence of species-specific polypeptides. In mitochondria (Fig. 1) it can be seen that the polypeptide patterns of *L. angustifolius* and *L. luteus* are more similar to each other than to those of *L. albus*. The presence of "marker proteins", characteristic of the particular lupin species with pI > 7.0 and molecular weight of about ~35 kDa, is also observed. *Lupinus albus* shows the presence of two isoform polypeptides: *L. luteus* - 3 and *L. angustifolius* - 4.

For chloroplasts (Fig. 1), the main differences in the distribution of polypeptides are observed in the 26-52 kDa region in two pI: one about 5.5-6 and the other about eight. Marker polypeptide of *L. angustifolius* has pI 5.5-6.0 and ~30 kDa; *L. albus* about 29 kDa (pI 5.0), and *L. luteus* - of about 56 kDa polypeptide (pI 6.5).

The usefulness of the mitochondrial tRNA-like sequence that represents direct repeats surrounding mainly the 18S rRNA genes of lupin species (Karpińska *et al.* 1994) in detection of polymorphism was also checked in *Brassica napus* plants. Figure 2 shows the result of this investigation in *Brassica* and other *Lupinus* genotypes. The presence of additional bands beside the expected 138 bp indicates that the sequence in those regions of the genome is different. To investigate whether the tRNA-like sequence undergoes rearrangements in crosses, we analyzed some interspecific hybrids between *L. mutabilis* and lupins from the section *Albus* (Table 1). The crosses were also analyzed for the

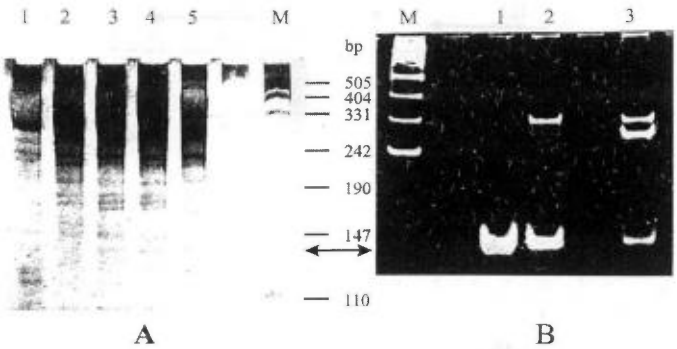

Fig. 2. PCR analysis of the mitochondrial tRNA–like repeated sequences in some *Brasica napus* (A) and *L. mutabilis* genotypes (B). Arrows indicate the proper length amplified fragment.
(A) 1 – cv MAR; 2 – cv POLO; 3 – cv SAMURAJ (CMS line); 4 – CMS Ogura x Polo; 5 – CMS Ogura x MAR.
(B) 1 – L. vavilovii; 2–*L. mutabilis* Mut 45; 3 – *L. albus* x *L. mutabilis* LM34; 4 – *L. vavilovii* x *L. mutabilis* Mut 45.

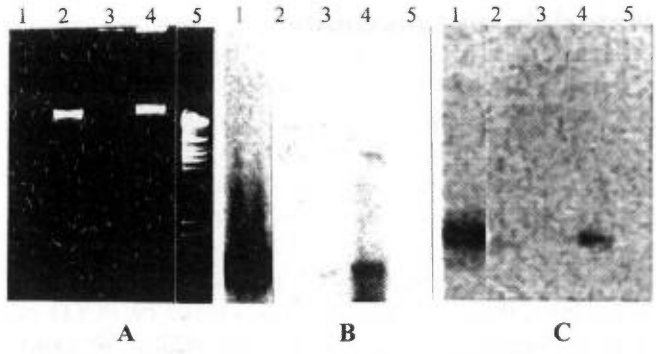

Fig. 3. Hybridization of the *L. albus* K_1 and K_2 plasmid like-DNA with Xho I digested DNAs from different lupin genotypes 1 - control DNA; 2 - *L. vavilovii* x *L. mutabilis* XM-5; 3 - *L. vavilovii*; 4. – *L. albus*; 5 – marker DNA. **A** – EtBr stained gel of the digested DNAs. **B** – autoradiogram of Southern blot of digested K_1 DNA. **C** - autoradiogram of Southern blot of digested K_2 DNA.

Table 1. Presence (+) or absence (-) of the plasmid like DNAs (K1 and K2) and the tRNA-like repeated sequences in DNA of *L. mutabilis* Sweet and its F1 interspecific hybrids – PCR analysis

Plant material	Molecular markers		
	K_1	K_2	tRNA-like
Female parent			
L. albus 520	+	+	+ s
L. termis 518	+	+	+ s
L. vavilovii 519	+	+	+ s
Male parent of *L. mutabilis*			
Mut 45	+ s	+ s	+ s
LM 34	+ s	+ s	+ s
XM 5	+	+ s	+ s
LM 13	+	+	+ s
Interspecific hybrids			
L. termis 518 x *L. mutabilis* LM 13	+	-	+ s
L. vavilovii 519 x *L. mutabilis* XM5	-	+	+ s
L. vavilovii 519 x *L. mutabilis* Mut45	+	-	+ s
L. albus 520 x *L. mutabilis* LM 34	+	+	+ s

s = detection of bands by silver staining.

presence of the K1 and K2 plasmid-like DNA molecules in our search for PCR fragments of 343 and 374 bp, respectively (Table 1).

Figure 3 shows the variation in the presence and copy number of two mitochondrial plasmid-like DNAs. This could indicate that the differences in the sorting of mitochondria (which influences the copy number) or mutations in the sequence may be responsible for the observed variability of plasmid-like DNAs.

Conclusions

1. The mitochondrial proteins contain specific marker proteins in lupin species.
2. 2-D PAGE of chloroplast proteins can be useful in determining the cytoplasmic variability of different plant species.
3. The mitochondrial repeated tRNA-like sequence provides a new possibility of detecting DNA polymorphism.
4. Molecular analyses of mitochondrial plasmid-like DNA enabled the determination of the mitochondrial variability in interspecific hybrids.

References

Aagaard, J.E., S.S. Vollmer, F.C. Sorensen and S.H. Strauss. 1995. Mitochondrial DNA products among RAPD profiles are frequent and strongly differentiated between races of Douglas fir. Mol. Ecol. 4:441-447.

Bookjans, G., B.M. Stummann and K.W. Henningsen. 1984. Preparation of chloroplast DNA from pea plastids isolated in a medium of high ionic strength. Anal. Biochem. 141:244-247.

Dumolin, S., B. Demesure, R.J. Petit. 1995. Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theor. Appl. Genet. 91: 1253-1256.

Gorączniak, R.M. and H. Augustyniak. 1991. Properties of lupin mitochondrial plasmid-like DNAs and nucleotide sequence of a new minicircular DNA. Plant Sci. 79:173-179.

Heukeshoven, J. and R. Dernic. 1985. Simplified method for silver staining of proteins in polyacrylamide gels and the mechanism of silver staining. Electrophoresis 6:103-112.

Hochstrasser, D.F., M.G. Harrington, A.C. Hochstrasser, M.J. Miller and C.R. Merril. 1988. Methods for increasing the resolution of two-dimensional protein electrophoresis. Anal. Biochem. 173:424-435.

Hurkman, W.J. and C.K. Tanaka. 1986. Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. Plant Physiol. 81:802-806.

Jack, P.L., T.A.F. Dimitrijevic and S. Mayes. 1995. Assessment of nuclear, mitochondrial and chloroplast RFLP markers in oil palm (*Elaeis guineensis* Jacq.). Theor. Appl. Genet. 90:643-649.

Karpińska, B. and H. Augustyniak. 1989. Lupin chloroplast and mitochondrial tRNA populations and their genes. Acta Biochim. Polon. 36:245-256.

Karpińska, B., K. Leśniewicz, G. Pietkiewicz and H. Augustyniak. 1994. Organization of the 18S, 5S, 4S rRNA genes and the tRNA-like repeat in the mitochondrial genomes of three lupin species. Acta Biochim. Polon. 41:1-8.

Lorenz, M., A. Weihe and T. Börner. 1994. DNA fragments of organellar origin in random amplified polymorphic DNA (RAPD) patterns of sugar beet (*Beta vulgaris L.*). Theor. Appl. Genet. 88:775-779.

Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Molecular cloning. A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Assessment of genetic diversity, stratification of germplasm accessions in diversity groups and sampling strategies for establishing a core collection of Indian sesame (*Sesamum indicum* L.)

I.S. Bisht[1]✉, R.K. Mahajan[1], T.R. Loknathan[2], P.L. Gautam[1], P.N. Mathur[3] and T. Hodgkin[4]

[1] National Bureau of Plant Genetic Resources, Pusa Campus, New Delhi-110 012, India. Email: techcell@nbpgr.delhi.nic.in
[2] National Bureau of Plant Genetic Resources, Regional Station, PKV Campus, Akola (MS)-444 104, India
[3] IPGRI Office for South Asia, c/o NBPGR, Pusa Campus, New Delhi - 110012, India
[4] International Plant Genetic Resources Institute (IPGRI), Via delle Sette Chiese, 00145 Rome, Italy

Summary

Assessment of genetic diversity, stratification of germplasm accessions in diversity groups and sampling strategies for establishing a core collection of Indian sesame (*Sesamum indicum* L.)

Diversity in the Indian sesame collection (3129 accessions), which represents all ecogeographic regions, was studied for a range of morphological and agronomic characters. A detailed multivariate analysis was performed on a set of 100 selected accessions representing different agro-ecological zones. The accessions were classified into seven discrete clusters and the entire collection was allocated to different clusters. Representation of various zones in seven clusters gave the opportunity of forming distinct diversity groups, making combined use of passport and characterization data. Twenty diversity groups were finally formed and used for sampling accessions to build up a core collection. Various stratified random sampling strategies were compared for developing a core collection of indigenous sesame accessions. The Shannon Diversity Index pooled over 19 descriptors was used as the measure of diversity. Simple random sampling and six stratified random sampling strategies, were compared for varying sample fractions ranging from 5 to 30%. Stratified random sampling was always superior to simple random sampling. Of the six stratified sampling strategies, a genetic diversity dependent strategy with sampled accessions from various strata directly proportional to the product of the diversity and logarithm of the size of the strata (GL strategy) appeared to be the best strategy. Invariably a 10% sample fraction was determined as the suitable sample fraction for all sampling strategies and served the objectives of both germplasm management and use.

Key words: Core collection, genetic diversity, germplasm, Sesame, *Sesamum indicum* L., Shannon Diversity Index, stratified sampling

Résumé

Evaluation de la diversité génétique, stratification des accessions de germoplasmes dans des groupes de diversité et stratégies d'échantillonnage pour l'établissement d'une collection représentative du sésame indien (*Sesamum indicum* L.)

La diversité au sein de la collection de sésame indien (3129 accessions), qui représente toutes les régions écogéographiques, a été étudiée par rapport à une gamme de caractères morphologiques et agronomiques. Une analyse multivariée approfondie a été réalisée sur un ensemble de 100 accessions choisies, représentant différentes zones agro-écologiques. Les accessions ont été réparties en sept groupes distincts et l'appartenance de l'ensemble des membres de la collection aux différents groupes a été déterminée. La répartition des diverses zones en sept groupes a permis de constituer des groupes de diversité distincts, en utilisant à la fois des passeports et des données de caractérisation. Vingt groupes de diversité ont finalement été formés et utilisés pour échantillonner les accessions afin de constituer une collection de référence. Différentes stratégies d'échantillonnage aléatoire stratifié ont été comparées afin de constituer une collection de référence d'accessions de sésame indigène. L'index de diversité de Shannon établi à partir de 19 descripteurs a été utilisé pour mesurer la diversité. Un simple échantillonnage aléatoire et six stratégies d'échantillonnage aléatoire ont été comparés pour différentes fractions de l'échantillon allant de 5 à 30 %. Un échantillonnage aléatoire stratifié donne toujours de meilleurs résultats qu'un simple échantillonnage aléatoire. Sur les six stratégies d'échantillonnage stratifié, la meilleure est une stratégie dépendant de la diversité génétique avec des accessions provenant de différentes strates, qui est directement proportionnelle au produit de la diversité et du logarithme de la taille des strates (stratégie GL). Invariablement, une fraction de 10 % de l'échantillon est déterminée comme la fraction d'échantillon appropriée pour toutes les stratégies d'échantillonnage et répond à la fois aux besoins de gestion des germoplasmes et à leur utilisation.

Resumen

Valoración de la diversidad genética, estratificación del germoplasma de las accesiones en los grupos de diversidad y estrategias de muestreo para establecer una colección núcleo del sésamo de la India (*Sesamum indicum* L.)

Sobre la base de características agrícolas y morfológicas se ha estudiado la diversidad en las colecciones de sésamo de la India (3129 accesiones), que representan todas las regiones ecogeográficas. Se ha realizado un análisis variado y detallado sobre un conjunto de 100 accesiones seleccionadas que representaban diferentes zonas agro-ecológicas. Las accesiones se clasificaron en siete grupos distintos y la colección entera se asignó a grupos separados. La represen-tación de las zonas en siete grupos dio la posibilidad de formar grupos con distinta diversidad, utilizando conjuntamente los datos de caracterización y pasaporte. Al final se formaron 20 grupos de diversidad y se usaron para el muestreo de las accesiones con el fin de crear una colección núcleo. Se compararon varias estrategias de muestreo aleatorio estratifi-cadas para desarrollar una colección núcleo de accesiones de sésamo autóctono. Como medida de la diversidad, se ha utilizado el Shannon Diversity Index que acomuna más de 19 descriptores. Se compararon muestreos aleatorios simples y seis estrategias de muestreo aleatorio estratificadas para la variación de fracciones simples que van de un cinco a un 30%. El muestreo aleatorio estratificado se demos-traba siempre superior al simple. De las seis estrategias de muestereo aleatorio estrati-ficadas, la estrategia que se basa en la diveridad genética con muestreos de accesiones de varios estratos directamente proporcional al producto de la diversidad y al logaritmo del tamaño de los estratos (estrategia GL) demostró ser la mejor estrategia a seguir. Una fracción de muestra del 10% resultaba (sin variar) adecuada para todas las estrategias de muestreo y cumplía con los objetivos de manejo y uso del germoplasma.

Introduction

Sesame (*Sesamum indicum* L. syn. *S. orientale* L.) is considered to be the oldest of the oilseed plants and has been cultivated in Asia for over 5000 years. In India, the antiquity of sesame is known from the use of its seeds in religious ceremonies and its mention in old Hindu literature. It is commonly known as *til* (Hindi), *tal* (Gujarati), *tili* (Punjabi), *nuvvulu* (Telugu), *ellu* (Tamil) and *rasi* (Oriya), and also as *gingelly* in different parts of the country. Approximately 36

species (Kobayashi 1981) are said to be in existence and these wild species are located mainly in Africa, with only a few in India. Ethiopia is said to be the original home of this crop (Hilterbrandt 1932). However, India is rich in diversity of cultivated sesame (Joshi 1961; Nayar and Mehra 1970; Simmonds 1976).

De Candolle (1836) suggested that India might have received sesame from Malaya and Indonesia in pre-Aryan times, prior to 1500 BC. Vavilov (1926) considered sesame to be of polytypic origin and India and Abyssinia the main centres of origin together with central Asia. China was considered to be a secondary centre of origin for a special group of dwarf varieties. According to van Rheenen (1981), Africa is poor in different forms of the cultivated *S. indicum* but rich in other species of the genus *Sesamum*. The opposite holds true for India. In Africa, 18 endemic species have been found, of which only two species are reported in India and four in Sudan. Hilterbrandt (1932) observed little variation in African material while considerable diversity was observed in Indian material. However, he still considered Africa to be the centre of diversity owing to the richness of wild species. Bedigian (1981) feels that it is appropriate to consider Africa as the primary centre of origin owing to the wide genetic diversity to be found and that India may be considered a secondary centre.

India ranks first in the world in land area devoted to sesame (approximately 2.47 million ha annually – 40% of total area) and in production (0.74 million tonnes, 27% of total world production). It is the third major oilseed crop after groundnut and rapeseed mustard. Sesame is grown in different seasons in various regions of the country and covers practically all agro-ecological zones (Sharma 1994).

At the national level the National Bureau of Plant Genetic Resources (NBPGR) has the main responsibility for the sustainable management and safe conservation of diverse germplasm of crop plants and their wild relatives. The NBPGR maintains 6658 accessions of sesame including 4136 indigenous and 2522 from exotic sources. Much of this diversity was augmented during 1991-93 and the existing collections were enriched through a programme supported by the IBPGR (now IPGRI), and the receipt of the world collection from Israel and some exotic accessions/representative material from various countries grown during the Asian Workshop on Sesame Evaluation and Improvement held in India (Rana *et al.* 1994). In the past, accessions have been characterized and evaluated at two NBPGR centres: Jodhpur in a semi-arid location and Akola located in a subhumid agroclimate. Based on these past evaluations, three crop catalogues were brought out in phases (Chopra and Thomas 1982; Umesh Chandra *et al.* 1983; Loknathan *et al.* 1993).

Large collections of plant germplasm have accumulated in genebanks over time and have led to a serious concern among the scientific community as to whether the full range of genetic diversity they contain can be effectively managed and utilized. Developing procedures for reducing the size of a collection to a manageable and accessible level (a core size) is becoming one of the more important issues in the management and utilization of plant germplasm collections (Frankel 1984; Brown 1989a, 1989b; Marshall 1990).

For the efficient conservation and exploitation of genetic diversity it is important to know its nature and structure. This becomes even more important if the objective is to maximize the amount of useful genetic diversity in a germplasm collection such as a core collection. The work carried out so far clearly indicates that it is possible to develop a core collection using existing data on the accessions in a collection. The basic issues that are to be addressed for developing a core collection are (i) stratification of the collection into homogenous (diversity) groups, (ii) optimum sample size, i.e. number of accessions to be sampled from the whole collection, and (iii) sampling strategies, i.e. how to select core entries from groups.

The grouping approaches described and used in most of the core collection studies are mainly hierarchical. Taxonomic, geographical, agro-ecological, morphoagronomic, biochemical and molecular markers have often been used individually or in combination for hierarchical grouping of accessions. Data from genetic markers can also be used in a variety of ways to create a representative (core) collection (Brown and Schoen 1994; Gepts 1995; Bataillon *et al.* 1996). The issue on optimum sample fraction was addressed by Brown (1989a, 1989b) and a sample size of 10% of the whole collection was suggested based on a theoretical model put forward by Ewens (1972). However, in practice situations may differ markedly from the assumptions of the model. For sampling within groups, two techniques are mainly employed: (i) random sampling and (ii) use of principal components score strategy. Brown (1989a), Charmet and Balfourier (1995), Hintum (1995) and Yonezawa *et al.* (1995) proposed a stratified random sampling strategy. Different strategies viz. constant (C), proportional (P), logarithmic (L) by Brown (1989b); genetic diversity dependent (G) by Yonezawa *et al.* (1995); and M and H by Schoen and Brown (1995) were proposed for random selection of entries from groups. Principal components score strategy has been used in common bean (Singh *et al.* 1991), okra (Hamon and Van Sloten 1989; Mahajan *et al.* 1996), coffee (Hamon *et al.* 1995) and mungbean (Bisht *et al.* 1998).

A project was undertaken at the NBPGR to build up a core collection of Indian sesame germplasm accessions with the financial assistance of IPGRI. The objective of this study is to describe the diversity pattern in the Indian sesame collection and the relevant multivariate methods and to illustrate their use in forming groups that would facilitate the choice of accessions with maximal diversity within each group in order to develop core collections and help promote utilization of germplasm. Estimates of diversity using qualitative and quantitative descriptors were compared for various stratified random sampling allocation strategies and the optimum sample size determined for developing a core collection of Indian sesame accessions.

Material and methods

Diversity in Indian sesame accessions

A total of 4136 indigenous accessions from diverse ecogeographic regions were grown during the 1996 crop-

ping season at the NBPGR regional station Akola (Maharashtra). The material included maximum representation from Maharashtra, Madhya Pradesh, Rajasthan, Uttar Pradesh, Andhra Pradesh, Orissa, Bihar and Gujarat followed by a fair representation from Tamil Nadu, Kerala, Kranataka, Punjab, Himachal Pradesh, West Bengal and parts of northeastern India.

To classify environments in a region, data must be uniformly comparable across the region. While for some areas some detail is available this is not universally true with regard to altitude, mean temperature, rainfall, soil characteristics, expected phenology during the growing season, moisture stress and photoperiod at flowering. Therefore, the material was broadly categorized into seven main groups: Group I – northern region (including accessions from Uttar Pradesh, Himachal Pradesh, Punjab, Jammu & Kashmir), Group II – central region (accessions mainly from Madhya Pradesh and Maharashtra), Group III – southern region (comprising accessions from Andhra Pradesh, Tamil Nadu, Karnataka and Kerala), Group IV – eastern region (accessions from Bihar, West Bengal and Orissa), GroupV – northeastern region (accessions from parts of northeastern India), Group VI – western region (including parts of Maharashtra and Gujarat) and Group VII – arid region (accessions from Rajasthan and parts of Gujarat). The accessions were grown in two-row plots of 3 m row length with a row-to-row spacing of 50 cm and plant-to-plant spacing of 15 cm in an augmented block design. The material was grown in 83 blocks of 50 accessions each with two controls, replicated and randomized throughout the blocks. Final analysis was performed on a set of 3129 accessions (including 1207 accessions from the central region, 640 accessions from arid regions, 465 accessions from northern regions, 351 accessions from southern regions, 131 accession from western regions, 233 accessions from eastern regions and 102 from northeastern regions) after eliminating apparent duplicates and certain other accessions with incomplete passport data.

Data were recorded for a set of 19 distinct descriptors (Table 1) on five randomly selected plants per accession. The adjusted treatment (accession) means from the augmented design analysis were standardized to mean 0 and standard deviation 1. A range of variation for both qualitative and quantitative descriptors, and correlations between pairs of various quantitative and qualitative descriptors were studied.

A set of 362 accessions was randomly selected from the whole collection, with proportionate representation from different zones. From these 362 accessions, a set of 100 accessions with maximal diversity was finally selected us-

Table 1. Range of variation in quantitative characters and predominance of qualitative descriptor states of the whole sesame collection

	Variables	Min.	Max.	Mean	CV	Skewness
1	Days to flower	31.00	70.00	41.94	11.83	1.49**
2	Days to maturity	69.00	130.00	98.62	8.38	-0.07**
3	Plant height	22.30	120.30	82.47	19.66	-0.67
4	Capsules per plant	2.00	84.00	18.81	50.08	1.05**
5	Seeds per capsule	10.00	98.00	57.57	20.67	-0.34**
6	100-seed weight	0.07	0.47	0.32	15.62	-0.52**
7	Yield per plant	0.05	8.00	1.10	68.18	1.89**
8	Branching habit	Predominantly with moderate habit (83.72%)				
9	Stem hairiness	Predominantly sparse (96.09%)				
10	Flower colour	Predominantly white with deep violet/purple shading (66.78%) followed by white with light violet/purple shading (22.52%)				
11	Corolla hairiness	Predominantly hairy (58.69%) followed by glabrous to light hairy (31.81%)				
12	Internode length	Medium (68.90%), low (31.00%)				
13	Capsule shape	Broad oblong (58.73%), narrow oblong (41.27%)				
14	Density of capsule hair	Predominantly hairy type (83.15%)				
15	Number of capsule per leaf axil	One(68.08%), more than one (31.92%)				
16	Number of locules per capsule	Predominance of four loculed capsules (92.36%)				
17	Capsule length	Small (50.04%), medium (47.08%)				
18	Seed colour	Predominantly white (55.14%) followed by brown (17.70%), red (17.37%) and black (10.30%)				
19	Incidence of Phyllody disease	Predominance of moderate to high susceptibility				

**P (<0.01).

ing principal components score strategy (Hamon *et al.* 1995; Mahajan *et al.* 1996; Bisht *et al.* 1998). We ensured that these 100 accessions were, to a large extent, representative of the whole spectrum of diversity of the entire collection from different zones. The accessions were grown at two NBPGR locations, Delhi and Akola, and a detailed multivariate analysis was performed. Different hierarchical clustering techniques viz. single linkage, complete linkage, unweighted pair group arithmetic averaging (UPGMA) and Ward's minimum variance technique were tried to classify the accessions into discrete clusters (Romesburg 1984). Ordination of the accessions was carried out by using principal components analysis (PCA). The INDOSTAT statistical package developed at INDOSTAT Services, Hyderabad, India was used for classification and ordination analysis.

Stratification of germplasm accessions in groups of diversity

A computer programme was developed to allocate the entire collection (3129 accessions) in different clusters (obtained on the basis of 100 selected accessions). The distribution pattern of accessions from various zones in different clusters based on major character associations was studied and used for stratification (grouping) of germplasm accessions for sampling the entries for developing a core collection.

Computation of population Shannon Diversity Index

Data on 19 characters, both qualitative and quantitative, were used to estimate the SDI (Shannon and Weaver 1963). Quantitative characters were converted into qualitative descriptors by making frequency classes for computation of a diversity index.

Consider a population of N accessions spread over different descriptor states of each of the d qualitative descriptors. Let there be k_i (i=1,2,3,...,d) states for the i^{th} descriptor. Denote N_{ij} for the number of accessions associated with j^{th} descriptor state of the i^{th} descriptor such that $\sum_j N_{ij} = N \forall i$. Then $P_{ij} = N_{ij}/N$ denote the proportion of accessions belonging to the j^{th} descriptor state of the i^{th} descriptor. Thus the population Shannon Diversity Index (SDI_i) for the i^{th} descriptor is

$$SDI_i = -\sum_j P_{ij} \log_e P_{ij}$$

and the population diversity index is

$$SDI = \sum_i SDI_i$$
$$= -\sum_i \sum_j P_{ij} \log_e P_{ij}$$

Estimation of SDI

Simple Random Sampling Without Replacement (SRSWOR). A simple random sample of size n is drawn by SRSWOR from N accessions. The frequency distribution of these n units over descriptor states for each of the descriptors is computed. Let p_{ij} (i=1,2,3,...,d; j=1,2,...k_i) denote the sample proportion of accessions for j^{th} state of the i^{th} descriptor. Then an estimate of SDI is

$$sdi = -\sum_i \sum_j p_{ij} \log_e p_{ij}$$

Due to the complex nature of sdi, it is difficult to judge its unbiasedness. However, sdi is a consistent estimate of SDI. For computation of E(sdi) and V(sdi), 100 repeated independent random samples of size n by SRSWOR were drawn from the population and sdi, average sdi and variance among sdi values were computed. Then mean square error (MSE) of sdi is

$$MSE(sdi) = V(sdi) + bias^2(sdi)$$

where bias (sdi) = E(sdi) - SDI

and then 95% confidence interval (CI) can be computed.

Stratified random sampling. Let the population of N units be grouped into k strata with size $\{N_s\}$, s = 1,2,...,k for the s^{th} stratum such that $\sum_s N_s = N$. Samples of sizes $\{n_s\}$ are drawn independently by SRSWOR from the s^{th} strata such that $\sum_s n_s = n$. Assuming the accessions in a stratum constituting the entire population, SDI is computed for the s^{th} stratum. Then SDIst, the pooled SDI_s over different strata, is defined as

$$SDIst = \sum_s W_s SDI_s$$

Where W_s is the weight attached to the s^{th} stratum and SDI_s is the SDI for the s^{th} stratum. The weights W_s are to be selected such that SDIst $\cong$ SDI. The weights W_s are chosen by the researcher on the basis of past experience. In the absence of any idea about the weights, the following three set of weights were considered:

Equal weights: equal weights being attached to each of the k strata, i.e. $Ws = 1/k \; \forall \; s$.

Proportional weights: weights being proportional to the size of the strata, i.e. $W_s \propto N_s \Rightarrow W_s = N_s/N$ for the s^{th} stratum.

Proportional to stratum diversity: weights being proportional to diversity in the s^{th} stratum, i.e. $W_s \propto SDI_s \Rightarrow W_s = SDI_s / \sum_s SDI_s$.

The set of weights for which SDIst $\cong$ SDI was considered for estimation of population diversity. Let it be W'_s for s^{th} stratum. Let $(sdi)_{sts}$ be the sample diversity index for the s^{th} stratum, then the estimator of SDI is

$$(sdi)_{st} = \sum_s W'_s (sdi)_{sts}$$

The researcher may select another set of weights on the basis of his past experience about the importance to be attached to different strata. In situations where it is not possible to choose W_s, $(sdi)_{st}$ will be an unbiased estimate of SDIst but a biased estimate of SDI.

Allocation strategies

Assuming that the sampling fraction (5, 10, 15, 20, 30%) is fixed in advance, the following six allocation strategies were compared.

Strategy I (P strategy). The number of accessions sampled from different strata are directly proportional to the size of the strata i.e. $n_s \propto N_s$

Strategy II (C strategy). Equal number of accessions i.e., $n_1 = n_2 = n_3 ... = n_s = n/k$ are selected from each stratum

Strategy III (L strategy). The number of accessions sampled from different strata are directly proportional to the logarithm of the size of the strata, i.e. $n_s \propto \log_e N_s$

Strategy IV (G strategy). The number of accessions sampled from different strata are directly proportional to the diversity in the strata i.e., $n_s \propto SDI_s$

Strategy V (GP strategy). The number of accessions sampled

from different strata are directly proportional to the product of size and diversity in the strata i.e., $n_s \propto N_s\ SDI_s$

Strategy VI (GL strategy). The accessions from different strata are directly proportional to the product of diversity and logarithm of the size of strata i.e., $(SDI)_{sts} \log_e N_s$.

For obtaining the expectation, MSE of $(sdi)_{st}$ and the corresponding CI the procedure adopted was the same as in SRSWOR with bias being given by

$Bias((sdi)_{st}) = E((sdi)_{st}) - SDI_{st}$

Results

Studies on the whole collection revealed significant differences for all the quantitative variables except yield per plant. Block differences were also found to be significant for all other variables except yield. Therefore, the treatment means were adjusted for the block effects for further analysis.

The range of variation with regards to seven quantitative descriptors and predominance of descriptor states (frequencies) for 12 qualitative descriptors are presented in

Table 2. List of selected accessions with source

S. no.	Accession†	State	S. no.	Accession	State
1	NIC-16245 (1)	Andhra Pradesh	51	IC-132090 (7)	Maharashtra
2	NIC-16227 (3)	Andhra Pradesh	52	IC-132033 (1)	Maharashtra
3	NIC-08009 (4)	Andhra Pradesh	53	IC-131798 (6)	Maharashtra
4	NIC-07934 (3)	Andhra Pradesh	54	IC-131783 (5)	Maharashtra
5	NIC-07929 (3)	Andhra Pradesh	55	IC-132449 (5)	Maharashtra
6	NIC-07925 (5)	Andhra Pradesh	56	NIC-08874 (2)	Maharashtra
7	NIC-07921 (3)	Andhra Pradesh	57	IC-132541 (5)	Maharashtra
8	NIC-16153 (7)	Andhra Pradesh	58	IC-132128 (6)	Maharashtra
9	IC-096222 (1)	Andhra Pradesh	59	IC-006168 (6)	Maharashtra
10	IC-132293 (2)	Arunachal Pradesh	60	NIC-09972 (1)	Maniipur
11	NIC-14723 (2)	Assam	61	NIC-09985 (3)	Nagaland
12	NIC-14726 (1)	Assam	62	NIC-09986 (2)	Nagaland
13	NIC-14833 (2)	Assam	63	NIC-09999 (1)	Nagaland
14	NIC-14829 (2)	Assam	64	NIC-09998 (1)	Nagaland
15	NIC-14714 (1)	Assam	65	NIC-09980 (4)	Nagaland
16	NIC-14729 (2)	Assam	66	NIC-09982 (1)	Nagaland
17	NIC-10618 (6)	Bihar	67	NIC-09995 (2)	Nagaland
18	NIC-10628 (5)	Bihar	68	NIC-17386 (1)	Orissa
19	IC-096127 (5)	Bihar	69	NIC-17354 (3)	Orissa
20	IC-096128 (4)	Bihar	70	NIC-17363 (1)	Orissa
21	IC-043203 (5)	Gujarat	71	NIC-17384 (4)	Orissa
22	IC-043135 (5)	Gujarat	72	IC-056151 (5)	Orissa
23	NIC-08579 (3)	Gujarat	73	IC-096021 (5)	Orissa
24	NIC-08528 (7)	Gujarat	74	IC-131496 (7)	Punjab
25	NIC-08543 (7)	Gujarat	75	IC-014130 (6)	Rajasthan
26	NIC-17281 (5)	Himachal Pradesh	76	NIC-08388 (7)	Rajasthan
27	NIC-09840 (1)	Karnataka	77	IC-131986 (1)	Rajasthan
28	NIC-09845 (1)	Karnataka	78	NIC-08481 (4)	Rajasthan
29	NIC-09861 (3)	Kerala	79	NIC-08356 (5)	Rajasthan
30	NIC-09866 (4)	Kerala	80	IC-043230 (6)	Rajasthan
31	NIC-09874 (3)	Kerala	81	NIC-08380 (7)	Rajasthan
32	NIC-09863 (2)	Kerala	82	NIC-17846 (3)	Rajasthan
33	NIC-09868 (3)	Kerala	83	NIC-17494 (7)	Rajasthan
34	IC-052619 (4)	Madhya Pradesh	84	NIC-17886 (6)	Rajasthan
35	IC-052594 (1)	Madhya Pradesh	85	NIC-17857 (7)	Rajasthan
36	IC-063608 (1)	Madhya Pradesh	86	NIC-17880 (7)	Rajasthan
37	IC-052573 (1)	Madhya Pradesh	87	NIC-17847 (1)	Rajasthan
38	IC-021699 (1)	Madhya Pradesh	88	NIC-17472 (4)	Rajasthan
39	IC-041961 (1)	Madhya Pradesh	89	NIC-13601 (3)	Tamil Nadu
40	IC-023340-1 (5)	Madhya Pradesh	90	IC-073517 (5)	Tamil Nadu
41	IC-023314 (1)	Madhya Pradesh	91	IC-132262 (6)	Tripura
42	IC-132669 (5)	Maharashtra	92	NIC-07831 (5)	Uttar Pradesh
43	IC-132584 (6)	Maharashtra	93	NIC-16345 (5)	Uttar Pradesh
44	IC-132139 (7)	Maharashtra	94	NIC-07786 (7)	Uttar Pradesh
45	IC-019247 (5)	Maharashtra	95	NIC-16355 (1)	Uttar Pradesh
46	IC-131604 (1)	Maharashtra	96	NIC-07819 (5)	Uttar Pradesh
47	NIC-08242 (4)	Maharashtra	97	NIC-07785 (6)	Uttar Pradesh
48	NIC-08291 (7)	Maharashtra	98	NIC-09886 (3)	West Bengal
49	IC-132126(7)	Maharashtra	99	NIC-09897 (4)	West Bengal
50	IC-000498-1 (6)	Maharashtra	100	NIC-09899 (4)	West Bengal

† Numerals in parenthesis represent the cluster numbers.

Table 1. Wide variation was found for all the quantitative variables. The coefficient of variation was more than 20% for yield per plant, capsules per plant and seeds per capsule, indicating skewed distribution both positive and negative. Accessions differed in height and the extent and number of branching. In some types, the basal branches appeared very low on the stem and the branches produced numerous laterals and the plants appeared bushy. In other types, the secondary branches appeared high up on the stem and did not form any laterals, therefore, the plants looked open in habit. Some types, particularly those from arid zones with three flowers in the axil, had few or no branches. Greater height was observed in late, single flowered, many branched forms. Hairy stems in accessions from south zones and very profusely hairy stems and leaves in accessions from northeastern regions were recorded. The flowers exhibited a fair range of colour and are solitary in the majority of cases. However, two to three flowers in a leaf axil were observed in accessions from arid zones, and also from central and northwestern regions. Accessions from northeastern regions and a few from eastern and southeastern zones flower and fruit late in the season and showed photoperiodic response. They tend to produce excessive vegetative growth while reproduction is suppressed.

The capsules vary in their shape, size and number in the axil of a leaf and were hairy in the majority of cases. In a capsule, the number of locules ranged from four to eight, sometimes even more. Accessions from northeast and southern regions were invariably four loculed. A fair representation of four, six and eight-loculed capsules, usually borne singly in the leaf axil, was recorded mainly in accessions from arid, central and eastern regions. Relatively larger sized capsules were recorded in accessions from central, arid and northwestern regions. Accessions from northeast and south had narrow oblong capsules.

Seed colour varied from white to various shades of brown and grey to black or red. Seeds in accessions from the northeast were predominantly black, from eastern regions brown to black and from the south predominantly red or reddish brown. Accessions from other regions were predominantly white seeded. Seed size also varied from small to large. Accessions from the northeast were invariably small seeded. A low incidence of phyllody was observed in collections from the northeast whereas from other regions susceptibility varied from medium to high.

A significantly high and positive correlation was observed between days to flowering and maturity. Days to flowering were negatively correlated with plant height, capsules per plant, seeds per capsule, 100-seed weight and yield per plant. Plant height was positively and significantly correlated with capsules per plant, seeds per capsule, 100-seed weight and yield. Capsules per plant were significantly and positively correlated to seeds per capsule, 100-seed weight and yield and mean seed weight was also positively correlated with yield per plant. High and significant correlations between various qualitative descriptors viz. branching habit, capsule characteristics, seed colour, etc. were recorded. Accessions with a non-branching

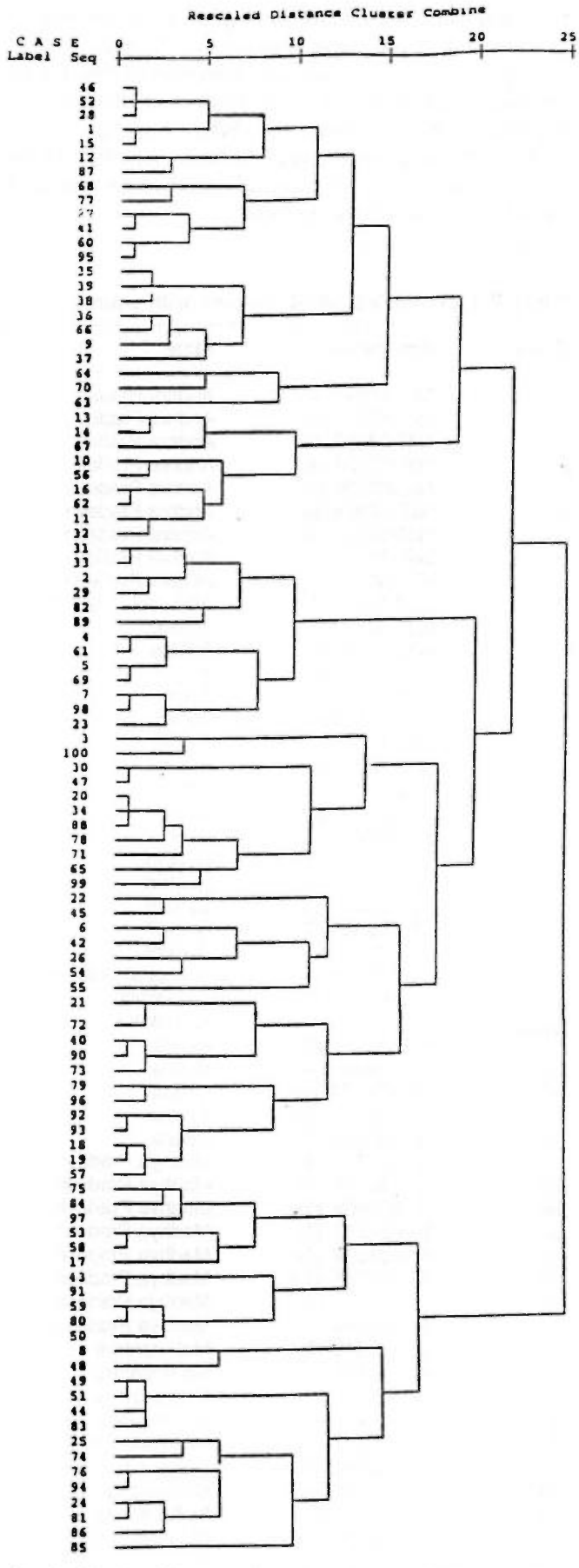

Fig. 1. Ward's minimum variance dendrogram of 100 accessions.

habit were early maturing, had a low number of capsules/plant, were broad oblong in shape, and had two to three capsules per leaf axil with predominantly white seed colour and bold seeds.

Clustering of 100 selected accessions (Table 2) revealed that Ward's minimum variance technique was found to be most suitable in classifying accessions in discrete and well-defined clusters. The dendrogram of 100 accessions is presented as Figure 1. Table 3 summarizes the number of clusters in different ranges of average Euclidean distance coefficients in order of decreasing width of the range. We had a choice of three, five and seven clusters as the wide range indicates that clusters are well separated in the attribute space. The tree was cut at the range of $15.25 < d < 16.70$ to produce classes that are maximally related to other specific variables of interest and served our purpose better. The accessions were, therefore, grouped into seven clusters.

Table 3. Number of clusters for different ranges of average euclidean distance coefficients (*d*) for the dendrogram shown in Fig 1

Number of clusters	Range	Width of the range
3	$19.25 < d < 21.75$	2.00
5	$17.00 < d < 18.75$	1.75
7	$15.25 < d < 16.70$	1.45
10	$13.70 < d < 15.00$	1.30
4	$18.75 < d < 19.75$	1.00
6	$16.70 < d < 17.00$	0.30
8	$15.00 < d < 15.25$	0.25

Critical examination of the clusters revealed that clusters were homogenous within themselves and heterogeneous between each other based on major character associations. Cluster I contained maximum accessions (23) followed by cluster V (19), VII (14), III (13), IV (11), VI (11) and II (9). Table 4 gives cluster means for various quantitative descriptors and predominant qualitative descriptor states for each cluster. The clusters were characterized as follows:

Cluster I: Cluster I comprises accessions with low plant height, late maturity, small seeds, low number of seeds per capsules, predominantly black in colour, highly pubescent corolla and stem.

Cluster II: Cluster II was mainly characterized by accessions with a high branching habit, very small seed predominantly black, highly pubescent stem and corolla, low internode length and low incidence of phyllody.

Cluster III: Accessions in cluster III were characterized with light purple flowers, relatively more capsules per plant, highly branched stems, early flowering and maturity, and red seed colour.

Cluster IV: Cluster IV was mainly characterized by accessions with a highly hairy corolla, moderate branching habit, small capsules broad oblong in shape, and a predominance of red and brown seeds.

Cluster V: Accessions with a light purple flower, narrow oblong capsules highly hairy; mainly one capsule per leaf axil, fair representation of four, six and eight locules per capsule, capsules medium to long in size, bold seeds, a large number of seeds per capsule black/brown in colour and a high incidence of phyllody.

Table 4. Means of various quantitative descriptors and predominance of qualitative descriptor states for various clusters of 100 selected sesame accessions

Variable	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI	Cluster VII
Days to flowering	49.95	48.00	42.23	42.82	40.21	38.55	40.07
Days to maturity	103.61	96.33	92.39	94.36	96.84	101.09	98.36
Plant height	61.09	70.70	83.78	78.06	84.93	101.09	98.36
Number of capsules per plant	12.44	16.22	23.23	16.18	18.95	28.82	12.93
Number of seeds per capsule	48.87	53.11	54.85	52.36	58.90	67.36	47.85
100-seed weight	0.26	0.23	0.30	0.28	0.33	0.34	0.34
Branching habit	Moderate	High	High	Moderate	Moderate	Low	Low
Stem hairiness	Sparse	Very hairy	Sparse	Sparse	Sparse	Hairy	Sparse
Corolla hairiness	Light hairy	Hairy	Light hairy	Very hairy	Hairy	Hairy	Hairy
Flower colour	White with deep purple shading	White with deep purple shading	White with light purple shading	White with deep purple shading	White with light purple shading	White with deep purple shading	White with deep purple shading
Internode length	Medium	Low	Medium	Medium	Medium	Low	Low
Capsule shape	Narrow oblong	Narrow oblong	Narrow oblong	Broad oblong	Narrow oblong	Broad oblong	Broad oblong
Density of capsule hair	High	High	Low	Sparse	High	High	Sparse
Capsules per leaf axil	One	One	One	One	One	Two to three	Two to three
Locules per capsule	Four	Four	Four	Four and more	Four and more	Four and more	Four and more
Capsule length	Small	Small	Small	Small	Medium	Long	Medium
Seed colour	Black	Black	Red	Red/brown	Brown/black	White	White
Incidence of phyllody	Low to medium	Low	Low to medium	High	High	High	High

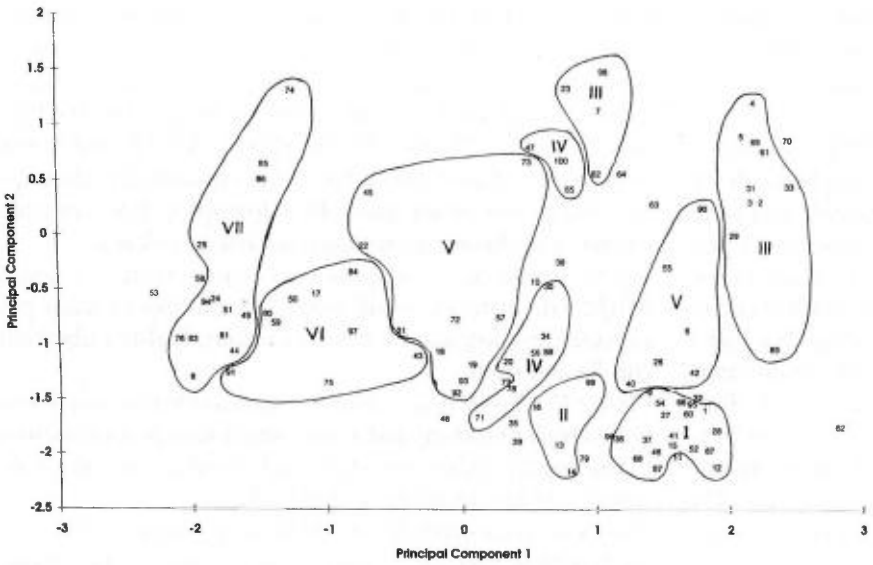

Fig. 2. Principal components analysis ordination of 100 accessions. The encircled accessions indicate the groups (clusters) obtained through cluster analysis (Fig.1).

Table 5. Representation of entire accessions from different zones in various clusters

Clusters	North	Central	South	East	Northeast	West	Arid
I	68	333	34	28	18	10	64
II	5	3	2	2	38	-	-
III	-	5	59	34	15	-	5
IV	8	6	190	85	12	5	7
V	239	478	51	49	10	55	180
VI	21	8	7	12	4	2	12
VII	124	374	8	23	5	59	372

Cluster VI: Accessions very early in flowering, predominantly tall non-branching and indeterminate, relatively more capsules per plant, two to three capsules per leaf axil and white seed colour.

Cluster VII: Accessions early flowering, fairly good representation of non-branching types, low internode length, medium to tall plant height, capsules broad oblong in shape, predominance of two to three capsules per leaf axil, capsules medium-long in size and bold seeds predominantly white in colour.

The clustering pattern on 100 selected accessions revealed that accessions from northeast, east and southern zones formed distinct groups. Cluster II contained the majority of the accessions from the northeast. Accessions from the south were grouped in clusters III and IV while accessions from the eastern region were grouped in cluster IV. Accessions from all other zones did not form distinct groups and showed overlapping mainly in clusters I, V, VI and VII.

The use of Principal Components Analysis (PAC) revealed that only the first three principal components (PC) had Eigen values of more than 1 and accounted for 26.18, 15.10 and 14.90% of total variation individually and 56.18% altogether. Fig 2 presents the scatter plot of PC scores for the first two PCs. In general the PCA result confirmed those obtained by cluster analysis, though it did not form robust groups. In some cases the pairs of accessions originating from the same location fall into tight groups within clusters whereas in other cases, although they are from the same location, they are far apart in ordination. Cluster VI and VII, and I and II fall close together in ordination. Cluster III formed a distinct group. Accessions in clusters IV and V were well dispersed in the two-dimensional space and showed overlapping.

An almost similar pattern was revealed when the whole collection was allocated to different groups (Table 5). Cluster V, VII and I comprised the maximum number of accessions i.e. 1062, 965 and 555 accessions, respectively, with the majority of the accessions from north, central, west and arid zones. Cluster III and IV had maximum representation from south and east. Cluster II had major representation from the northeast. Cluster VI was represented from almost all zones. Based on major character associations, accessions from north, central, arid and west zones were grouped mainly in clusters I, V and VII. Accessions from these zones show a high degree of genetic similarity and overlapping. Therefore, three major diversity groups and 12 groups altogether could be formed from these four zones. While adopting a subjective approach certain minor groups were readjusted/merged and finally 20

Table 6. Diversity group size and their Shannon Diversity Indices (SDI)

Diversity group	No. of accessions	SDI
1	333	11.65
2	59	12.00
3	60	11.76
4	142	12.00
5	50	13.82
6	66	14.43
7	211	9.43
8	65	13.83
9	100	12.31
10	235	9.26
11	470	10.81
12	200	11.92
13	65	10.78
14	37	10.07
15	66	13.36
16	124	10.62
17	395	11.65
18	402	11.89
19	28	11.75
20	22	8.58

Table 7. Shannon Diversity Index and its statistical parameters for simple random sampling and other allocation strategies

Sample size	Statistical parameters	Simple random sampling	Stratified sampling allocation strategies					
			I	II	III	IV	V	VI
5%	SDI	9.66	7.55	9.60	9.59	9.62	7.83	9.87
	%SDI†	68.23	51.68	65.66	65.59	65.82	53.61	67.51
	MSE	29.60	8.20	4.90	4.94	4.78	7.62	4.12
	CI	1.00-20.32	1.94-13.16	5.26-13.93	5.23-13.94	5.33-13.90	2.42-13.24	5.88-13.86
10%	SDI	11.34	9.66	10.75	10.65	10.99	9.71	11.05
	%SDI	80.15	66.14	73.58	72.82	75.24	66.47	75.58
	MSE	12.86	4.59	1.12	1.35	1.02	4.37	0.95
	CI	4.31-18.37	5.47-13.86	8.68-12.83	8.38-12.93	9.02-12.97	5.61-13.81	9.14-12.96
15%	SDI	12.59	10.35	-	11.12	11.18	10.41	11.28
	%SDI	86.12	70.80	-	76.07	76.49	71.24	77.15
	MSE	5.88	2.13	-	0.47	0.39	1.96	0.29
	CI	7.84-17.34	7.49-13.21	-	9.77-12.47	9.92-12.41	7.66-13.15	10.23-12.33
20%	SDI	12.88	10.71	-	11.38	11.37	10.79	11.58
	%SDI	88.12	73.96	-	77.90	77.83	73.81	79.21
	MSE	3.13	1.21	-	0.18	0.19	1.04	0.14
	CI	11.15-18.09	8.55-12.86	-	10.55-12.22	10.52-12.23	8.78-12.79	10.85-12.31
30%	SDI	13.70	11.16	-	11.55	11.56	11.19	11.62
	%SDI	93.72	76.36	-	79.03	79.10	76.62	79.48
	MSE	1.00	0.42	-	0.07	0.06	0.37	0.05
	CI	11.74-15.66	8.82-12.43	-	11.04-12.06	11.08-12.03	10.00-12.39	11.18-12.06

† SDI expressed as percentage of population SDI.

distinct groups were formed using information primarily on passport data (Table 6).

The population SDI was recorded to be 14.62. The SDI from stratified sampling with weights proportional to (i) size of strata, (ii) equal weights and (iii) diversity in the strata was found to be 11.36, 11.59 and 11.82, respectively. SDI with weights proportional to diversity was the highest but not very close to the population SDI. In the absence of any knowledge of W_s, the weights proportional to diversity were used to compare alternative sampling allocation strategies in stratified sampling.

SDI of different strata are presented in Table 6. From this it is evident that diversity in different strata is not uniform but ranges from 8.58 to 14.43. Diversity was at a maximum in stratum 6 (14.43) followed by stratum 8 (13.83) and 15 (13.36). It varied from 10 to 12 in the remaining strata except 7, 10 and 20 where the SDI was observed to be 9.43, 9.26 and 8.58, respectively. SDI was comparatively higher in small groups than in large groups.

Comparison of different sampling strategies is presented in Table 7. In strictly random sampling, the diversity varied from 68.23 to 93.72% for a sample size of 5 to 30%. The mean square error (MSE) varied from 1.00 to 29.60. The width of the confidence interval decreased with increase in sample size. There was considerable increase in %SDI for all sampling strategies from 5 to 10% sample fraction. Thereafter, there was no substantial increase in the %SDI with increasing sample size. A comparison of MSE clearly indicates that the genetic diversity dependent strategy, particularly IV and VI, always performed better in comparison to all other strategies. For equal allocation strategy it was not possible to go beyond 10% sample size. Strategy VI was consistently superior to other strategies with least MSE and a narrow 95% confidence interval for different sample sizes considered. Strategy I with %SDI ranging from 51.68 to 76.36% was a poor performer with highest MSE for different sampling fractions. A 10% sample fraction was, therefore, found appropriate for the development of the sesame core set as it is likely to serve both the objectives of germplasm management and utilization.

Discussion

Assessment of genetic diversity gives the breeder an opportunity to take up a hybridization programme for specific agro-ecological conditions and situations, keeping in mind the appropriate plant type for the situation. A plant form with medium height, short internodes, good branching at acute angles starting at lower height on stem, synchronous flowering, and multicapsular and multilocular habit should be arrived at (Sharma 1994). Knowledge of photo- and therm-sensitivity is also important and will permit incorporation of sesame in multiple cropping systems. Genotypes with good emergence ability, seedling vigour, resistance to seed and seedling rot diseases and ability to to combat diseases and pests late in the season are desirable. As the world production of sesame has to be maintained and even expanded, several key breeding objectives must be maintained These are yield potential and harvest index, seed retention, uniform maturation through determinate habit, and resistance to diseases. By utilizing genetic diversity, it is

possible to improve productivity of existing sesame cultivars. Earlier studies conducted by Kashi Ram (1930), Hiltebrandt (1932), Ali Mohammad & Alam (1932), Sen Gupta and Payne (1947), Sen Gupta and Sen (1947) and Ghosh (1955) give a detailed account of variability in morphological characters of sesame.

Clustering of selected 100 accessions using Ward's method placed the accessions into seven major groups. Accessions from the northeastern, southern and eastern regions of India formed distinct groups and the results indicated that the accessions had distinct morphological and agronomic traits. On the other hand, accessions from other zones did not seem to contribute a separate set and had a relatively high degree of genetic similarity or overlapping. Distinct subgroups within the main groups could, however, be identified based on microclimatic differences within zones. The distinctness of these subgroups is more prominently reflected in the PCA ordination. Cluster I formed one big tight group and the rest of the accessions revealed overlapping with clusters III and IV in the two dimensional space. Accessions in cluster II were relatively more dispersed in ordination. Cluster III formed two distinct subgroups in ordination. Cluster IV formed one major group and other accessions were well dispersed and showed overlapping with clusters I and V. Cluster V also formed two major groups far apart in ordination. Clusters VI and VII formed one major group each and were close to each other in ordination. The PCA, in general, confirmed the groupings obtained through cluster analysis. We cannot, however, expect robust groupings as the accessions are representative of the whole range of diversity of the entire collection. Consequently the variation explained by the first principal component axis are capsule shape, capsule per leaf axil and seed colour. Traits that separate accessions along the first principal component axis are capsule shape, capsule per leaf axil and seed colour. Traits that separate accessions along the second component are density of capsule hair, flower colour and corolla hairiness. Traits separating accessions along the third principal components axis are internode length, capsule length and locules per capsule.

A good core collection needs to define genetic diversity groups precisely. To build up the core collection of Indian sesame collection we used the database with passport and characterization data. Passport data are systematic and include mainly the origin, site and type of collected material. Charaterization data include mainly morphological and agronomic data. Combined use of agro-ecological and characterization data have also been reported to establish the CIAT *Phaseolus vulgaris* core collection (Tohme *et al.* 1995). Characterization data have also been used to develop a core collection of *Sorghum vulgare* (Prasada Rao and Rao 1995).

Allocation of the entire collection (3129 accessions) in seven clusters based on a clustering pattern of 100 selected accessions (representing different zones) results in grouping the accessions based on major character associations. Depending upon the size of the accessions representing various zones in different clusters, we had the opportunity to form distinct diversity groups making use of both passport and characterization data, and to select accessions from these groups to build up the core collection using suitable sampling strategies. In the present study, both agro-ecological passport data on origin of accessions and clustering technique based on phenetic analysis of individual accessions have been used to stratify accessions in diversity groups (Bisht *et al.* 1998).

To obtain a core sample the sampling strategy should maximize the diversity in the sample while attempting to reduce the redundancy of identical genotypes (Frankel and Brown 1984). Genetic diversity, however, is not randomly distributed over plant populations. It has a structure that can generally be summarized in a hierarchical model (van Hintum 1995).

A commonly used measure of diversity is the Shannon Diversity Index and this, pooled over different descriptors, has been considered as a measure of diversity. It gives larger values when more descriptor states are present, and other things being equal, it gives larger values when descriptor states are common than when some are common and others are rare. However, it is rarely affected by inclusion or exclusion of rare descriptor states in the sample (Emlen 1973). This is a desirable attribute in some contexts but in a core collection rare descriptor states are of considerable importance and need to be included with a subjective approach wherever possible.

In stratified sampling, weights are required to be attached to different strata in view of their relative importance over one another based on the degree of diversity in the different strata. All the weights are equal when the strata have equal variability and then stratification is as good as random sampling. Stratification is done when sampling units are heterogeneous with regards to characters of interest. Pooling of results over different strata is done so that the resulting SDI is approximately equal to the population SDI. In development of SDI for stratified sampling, unweighted estimates are considered. Due consideration for suitable weights is a practical necessity. Therefore, weights are considered here on the basis of closeness of pooled estimates to the population value. As with weights proportional to diversity the SDI_{st} is closer to population SDI, these weights formed the basis of comparison of different sampling strategies in stratified sampling.

In the present study, sampling through stratification based on well-identified diverse groups was more efficient than simple random sampling. To overcome the problem of giving undue higher weights to large groups, likely to contain a higher level of genetic redundancy, Growth and Roelfs (1987) and Brown (1989b) suggested that the number of accessions drawn should be in proportion to the logarithm of the frequency in each group (L strategy). The C strategy gives each group equal weighting and biases the core in favour of small groups. The P strategy considers a fixed proportion of each group so that the group is represented in the core in proportion to the frequency in the group. Clearly this strategy is biased towards large groups (Brown 1989b). The P strategy is likely to perform better when diversity is distributed according to the size of groups, which in prac-

tice rarely occurs. The C strategy is likely to perform better than the P strategy when diversity is concentrated in smaller groups. In other situations, constant allocation can be inferior to simple random sampling.

The genetic diversity dependent (G) strategy proposed by Yonezawa *et al.* (1995) takes into consideration the diversity in different groups irrespective of their sizes. It has been emphasized that this strategy is always better when the diversity in different groups is known. This directed us towards the development of some other intermediate strategies. Therefore, in the present study two more probable diversity dependent strategies, GL and GP, were also considered to take care both of group sizes and the level of diversity (Mahajan *et al.* 1999). We find that the GL strategy performed better than the G and GP strategies. The GL strategy is more appropriate when one encounters situations with different levels of diversity and sufficiently large strata sizes, while GP is likely to perform better in situations where groups are small and diversity is unevenly distributed within groups. In the last three strategies, we come across situations where the diversity in certain strata is very large and the corresponding strata size is small enough. Then n_s is likely to be greater than N_s. In such situations we take $n_s = N_s$. Apart from the above-suggested strategies, other sampling strategies could be applied to develop a core collection but their performance will depend upon the species, the composition and distribution of diversity in a collection, and the type of characters of interest (Spagnoletti Zeuli and Qualset 1993).

While comparing various sampling strategies on durum wheat, Spagnoletti Zeuli and Qualset (1993) found that the three strategies: (i) random stratified by geographic origin with P strategy, (ii) random stratified by geographic origin with L strategy, and (iii) random stratified with canonical variables, produced the desired effect of increasing frequencies from the less representative countries of origin. Diwan *et al.* (1994) proposed random sampling and classified the accessions of annual *Medicago* species-wise and accessions within species countrywide. Sampling within species was done by cluster analysis on agronomic and morphological traits. Basigalup *et al.* (1995) classified accessions of perennial *Medicago* plant introductions according to country of origin and eight sampling strategies were compared.

Two methods, combined cluster analysis based on principal components within each geographical group with random selection of entries within each cluster and direct selection of entries within each geographical group, were judged to be the best strategies for designating the core collection. Charmet and Balfourier (1995) compared SDI averaged over 200 samples generated by computer simulations using various proportions for sampling methods. The clustering method based on agronomic traits with geographic contiguity constraints was better than others. A random sample of 5% of accessions maintained 86% of diversity while a clustering method based on geostatistics gave the best results with 92% of the variation being maintained in 5% of the core sample.

Considerable increase in %SDI was observed when the sample fraction increased from 5 to 10%. No substantial increase was observed beyond a 10% sample fraction. Brown (1989a) also proposed an optimum sample size of 10%. Applying Ewen's theory, Brown (1989a) showed that the fraction of alleles retained from a population increased rapidly as the sample fraction increased to 0.1, but rather slowly after it surpassed 0.1. This is similar to the findings in the present study in all the sampling strategies tested.

On the basis of the theoretical model, Yonezawa *et al.* (1995) showed that the optimum sample fraction depends upon various genetic and resource parameters, primarily on the degree of genetic redundancy among accessions comprising the whole collection and the amount of resources available for maintenance of the core collection. The optimum sampling fraction was large, with a lower redundancy among accessions or a lower initial allelic diversity within accessions, indicating that a large sample fraction is better in species with a large selfing rate. A 20-30% sample was estimated to be appropriate in situations where accessions in the collection are neither very heavily nor very lightly redundant in terms of diversity.

The establishment of patterns and classification of genetic diversity within the germplasm surveyed through the employment of uni-, bi- and multivariate analysis has enabled clear groupings to be identified. The resulting information is useful in improving understanding of the diversity within the collection and how it is structured. Morphometric characters underlying these groups provide a useful aid in helping to target the search for new germplasm in order to augment the genetic base for breeding programmes required for future crop improvement. Also, the stratification of germplasm accessions into distinct diversity groups will enable curators to select accessions for forming a core collection. Based on the above discussion, it is evident that a suitable choice of weights and the selection of an appropriate sampling allocation strategy, together with an optimum sampling fraction, are essential in capturing the maximum diversity.

In the present study, the weights proportional to diversity in various strata, sampling allocation strategy with sample size directly proportional to the product of diversity, and logarithm of the strata size and 10% sample fraction are appropriate for the development of a sesame core set which can serve both the objectives of germplasm management and utilization.

Acknowledgements

We are thankful to IPGRI for providing financial support for sesame germplasm exploration and evaluation programmes in the past, and also for funding the present project on establishing a core collection of Indian sesame accessions at the NBPGR.

References

Ali Mohammad and Z. Alam. 1933. Types of *Sesamum indicum* D.C. in the Punjab. Indian J. of Agric. Sci. 3:897-911.

Basigalup, D.H., D.K. Barnes and R.E. Stucker. 1995 Development of core collection for perennial *Medicago* plant introductions. Crop Sci. 35:1163-68.

Bataillon, T.M., J.L. David and D.J. Schoen. 1996. Neutral genetic markers and conservation genetics. Genetica 144:409-417.

Bedigian, D. 1981. Origin, diversity, exploration and collection of sesame.

Pp. 164-169 *in* Sesame: Status and Improvement. Proceedings of Expert Consultation, Rome, Italy, 8-12 December, 1980. FAO plant production and protection paper 29.

Bisht, I.S., R.K. Mahajan and D.P. Patel. 1998. The use of characterisation data to establish the Indian mungbean core collection and assessment of genetic diversity. Genet. Resour. Crop Evol. 45:127-133.

Bisht, I.S., R.K. Mahajan, T. R. Loknathan and R.C. Agrawal. 1998. Diversity in Indian sesame collection and stratification of germplasm accessions in different diversity groups. Genet. Resour. Crop Evol. 45:325-335.

Brown, A.H.D. and D.J. Schoen. 1994. Optimum sampling strategy for core collections of plant genetic resources. Pp. 357-370 *in* Conservation Genetics (V. Loeschgke, J. Tomiuk and S.K. Jain, eds) Birkhauser Verlag, Basel, Switzerland.

Brown, A.H.D. 1989a. The case for core collections. Pp. 136-156 *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.

Brown, A.H.D. 1989b. Core collections: A practical approach to genetic resources management. Genome 31:818-824.

Charmet, G. and F. Balfourior. 1995. The use of geostatistics for sampling a core collection of perennial ryegrass populations. Genet. Resour. Crop Evol. 42:303-309.

Chopra, D.P. and T.A. Thomas. 1982. Catalogue on sesame germplasm, NBPGR Regional Station, Jodhpur.

De Candolle, A.P. 1836. Prodromus systematis naturalis regni vegetabilis sive enumeratio contracta ordinum, generum, specierumque plantarum uuc usque cognitarum, juxta methodi naturalis normas digesta. Pars quinta: Sistens Calycereas et Compositarum tribus priores. Treuttel et Würtz, Paris

Diwan, N., G.R. Bauchan and M.S. McIntosh. 1994. A core collection for the United States annual *Medicago* germplasm collection. Crop Sci. 34:279-85.

Emlen, J.M. 1973. Ecology: An evolutionary approach. Addison-Wesley, Reading, Massachusetts, USA.

Ewens, W.J. 1972. The sampling theory of selectively neutral alleles. Theor. Pop. Biol. 3:87-112.

Frankel, O.H. and A.H.D. Brown. 1984. Current plant genetic resources - a critical appraisal. Pp. 1-11 *in* Genetics: New Frontiers 4. Oxford and IBH Publishing Co., New Delhi.

Frankel, O.H. 1984. Genetic perspectiveness of germplasm conservation. Pp. 161-170 *in* Genetic Manipulation: Impact on Man and Society (W.K. Arber, K. Llimensee, W.J. Peacock and P. Starlinger, eds.). Cambridge University Press, Cambridge.

Gepts, P. 1995. Genetic markers and core collections. Pp. 127-146 *in* Core Collection of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, Th.J.L. van Hintum and E.A.V. Morales, eds.). John Wiley & Sons, Chichester,UK.

Ghosh, B.N. 1955. Photoperiodic response in til (*Sesamum indicum* Linn.). Curr. Sci. 24:170.

Growth, J.V. and A.D. Roelfs. 1987. The concept and measurement of phenotypic diversity in *Puccinia graminis* on wheat. Phytopathology 77:1395-1399.

Hamon, S and D.H. Van Sloten. 1989. Characterization and evaluation of okra. Pp. 173-196 *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.

Hamon, S., M. Noirot and F. Anthony. 1995. Developing coffee core collection using the principal components score strategy with quantitative data. Pp. 117-126 *in* Core Collection of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, Th.J.L. van Hintum and E.A.V. Morales, eds.). John Wiley & Sons, Chichester,UK.

Hilterbrandt, V.M. 1932. Sesame (*Sesamum indicum* L.). Bull. Appl. Bot. Plant Breed. 9:1-114.

Joshi, A.B. 1961. Sesame – a monograph. Indian Central Oilseeds Committee, Hyderabad, India.

Kashi Ram. 1930. Studies in Indian oilseeds (4). The types of *Sesamum indicum* D.C. Mem. Dep. Agric. India. Bot. 18:127-147.

Kobayashi, T. 1981. The wild and cultivated species in the genus *Sesamum*. Pp. 157-163 *in* Sesame: Status and Improvement. Proceedings of Expert Consultation, Rome, Italy, 8-12 December, 1980. FAO Plant Production and Protection Paper 29.

Loknathan, T.R., D.P. Patel, V.D. Verma, R.K. Mahajan, Bhag Singh, M.N. Koppar and R.S. Rana. 1993. NBPGR-IBPGR Collaborative Project: Catalogue on sesame (*Sesamum indicum* L.) germplasm. NBPGR, New Delhi.

Mahajan, R.K., I.S. Bisht and P.L. Gautam. 1999. Sampling strategies for developing Indian sesame core collection. Indian J. Plant Genet. Resour. 12:26-35.

Mahajan, R.K., I.S. Bisht, R.C. Agrawal and R.S. Rana. 1996. Studies on South Asian okra collections: Methodology for establishing a representative core set using characterisation data. Genet. Resour. Crop Evol. 43:249-255.

Marshall, D.R. 1990. Crop genetic resources: Current and emerging issues. *In* Breeding and Genetic Resources (A.H.D. Brown, M.H. Clegg, A.L. Kahlor and B.S. Weir, eds.). Sinauer, Sunderland, Massachusetts, USA.

Nayar, N.M. and K.L. Mehra. 1970. Sesame – its uses, botany, cytogenetics and origin. Econ. Bot. 24:20-31.

Peeters, J.P. and J.A. Martinelli. 1989. Hierarchical cluster analysis as a tool to manage variation in germplasm collections. Theor. Appl. Genet. 78:42-48.

Peeters, J.P. and J.T. Williams. 1984. Towards better use of genebanks with special reference to information. Plant Genet. Resour. Newsl. 60:22-32.

Prasada Rao, K.E. and V. Ramnatha Rao. 1995. The use of characterisation data in developing a core collection of sorghum. Pp. 109-115 *in* Core Collection of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, Th. J.L. van Hintum and E.A.V. Morales, eds.). John Wiley & Sons, Chichester,UK.

Rana, R.S., R.K. Arora, T.R. Loknathan and D.P. Patel. 1994. Sesame genetic resources in India: Their diversity, utilisation and conservation. *In* Sesame Biodiversity in Asia: Conservation, Evaluation and Improvement (R.K. Arora and K.W. Riley, eds.). IPGRI Office for South Asia, New Delhi, India.

Romesburg, H.C. 1984. Cluster analysis for researchers. Lifetime Learning Publications, Belmont, California.

Schoen, D.J. and A.H.D. Brown. 1995. Maximising genetic diversity in core collections of wild relatives of crop species. Pp. 55-76 *in* Core Collection of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, Th.J.L. van Hintum and E.A.V. Morales, eds.). John Wiley & Sons, Chichester,UK.

Sen Gupta, J.C. and N.K. Sen. 1947. Photoperiodic effect on *Sesamum indicum* Linn. Sci. and Culture 13:203-204.

Sen Gupta, J.C. and S.K. Payne. 1947. Leaf heteromorphism and photoperiods in *Sesamum orientale* L. Nature, Lond. 160:510.

Shannon, C.E. and W. Weaver. 1963. The mathematical theory of communication. Urbana, Illinois.

Sharma, S.B. 1994. Utilisation of national collections of sesame in India. *In* Sesame Biodiversity in Asia – Conservation, Evaluation and Improvement (R.K. Arora and K.W. Riley, eds.). IPGRI Office for South Asia, New Delhi, India.

Simmonds. 1976. (ed). Evolution of Crop Plants. Sesame, pp. 231-233. Longman, London, UK.

Singh, P.S., R. Nodari and P. Gepts. 1991. Genetic diversity in cultivated common bean: I. Allozymes. Crop Sci. 31:19-23.

Spagnoletti Zeuli, P.L. and C.O. Qualset. 1993. Evaluation of five strategies for obtaining a core subset from a large genetic resource collection of durum wheat. Theor. and Appl. Genet. 87:295-304.

Tohme, J., P. Jones, S. Beebe and M. Iwanaga. 1995. The combined use of agro-ecological and characterisation data to establish the CIAT *Phaseolus vulgaris* core collection. Pp. 95-108 *in* Core Collection of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, Th. J.L. van Hintum and E.A.V. Morales, eds.). John Wiley & Sons, Chichester, UK.

Umesh Chandra, R.L. Sapra and K.L. Mehra. 1983. Catalogue on sesame (*Sesamum indicum* L.) germplasm. NBPGR Regional Station, Shimla.

Van Hintum, Th. J. L. 1995. Hierarchical approaches to the analysis of genetic diversity in crop plants. *In* Core Collection of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, Th.J.L.van Hintum and E.A.V:Morales, eds.). John Wiley & Sons, Chichester, UK.

Van Rheenen, H.A. 1981. Genetic resources of sesame in Africa: Collection and exploration. Pp. 170-172 *in* Sesame: Status and Improvement. Proceedings of Expert Consultation,

Vavilov, N.I. 1926. Studies on the origin of cultivated plants. Bull. of Appl. Bot. 26. Institute of Applied Botany and Plant Breeding, Leningrad, USSR.

Yonezawa, K., T. Nomura and H. Morishima. 1995. Sampling strategies for use in stratified germplasm collections. Pp. 35-54 *in* Core Collection of Plant Genetic Resources (T. Hodgkin, A.H.D. Brown, Th.J.L. van Hintum and E.A.V. Morales, eds.). John Wiley & Sons, Winchester, UK.

ARTICLE Plant Genetic Resources Newsletter, 1999, No. 119 Supp.: 47-50

Establishment and development of sesame germplasm core collection in China

Zhang Xiurong[1]✉, Zhao Yingzhong[1], Feng Xiangyun[1], Cheng Yong[1], Guo Qingyuan[1], Li Yurong[2] and Wen Yongning[3]

[1] Oil Crops Research Institute of CAAS, Wuhan, Hubei Province, 430062, China
[2] Grain and Oil Crops Institute of Hebei Province of China
[3] Industrial Crops Institute of Guangxi Academy of Agricultural Sciences of China

Summary

Establishment and development of sesame germplasm core collection in China

A core collection was established from the whole collection of sesame germplasm resources in China. Sets of 50 accessions were randomly selected from the 4251 accessions. Different hierarchical clustering procedures were used with data on 14 traits and the accessions were grouped into seven ecotypes. The whole collection was then divided into 14 groups according to three levels of stratification: origin, varietal type and ecotype. In this way 884 accessions were pre-selected from the whole collection and planted in three ecological experiment sites. The results showed that 97.85% of the values obtained were the same as the original data from previous trials. Re-clustering identified 453 core accessions for the core collection with good representativeness, confirmed by zymogram characteristics. The traits selected have high heritability and distinct characteristics among the collected materials, making it possible to undertake an all-round, systematic and thorough study of sesame germplasm resources and to enhance identification and evaluation work.

Key words: China, cluster analysis, core collection, genetic diversity, sesame, *Sesamum indicum*

Résumé

Etablissement et développement d'une collection représentative de germoplasmes de sésame en Chine

Une collection représentative a été constituée à partir de l'ensemble de la collection des ressources en germoplasmes de sésame en Chine. Des ensembles de 50 accessions ont été sélectionnés au hasard à partir de 4251 accessions. Différentes méthodes de groupement hiérarchisé (hierarchical clustering) ont été utilisées avec des données concernant 14 caractères et les accessions ont été groupées en sept écotypes. L'ensemble de la collection a ensuite été divisé en 14 groupes selon trois niveaux de stratification : origine, type variétal et écotype. De cette manière, 884 accessions ont été pré-sélectionnées à partir de l'ensemble de la collection et plantées dans trois sites écologiques expérimentaux. Les résultats montrent que 97,85 % des valeurs obtenues sont les mêmes que les données originales des essais précédents. Un nouveau regroupement a permis d'identifier 453 accessions représentatives pour la collection de référence, et ce regroupement est confirmé par les caractéristiques des zymogrammes. Les caractères sélectionnés ont un taux élevé d'héritabilité et des caractéristiques distinctes parmi le matériel récolté, permettant ainsi de réaliser une étude complète, systématique et approfondie des ressources en germoplasmes de sésame et de faire progresser le travail d'identification et d'évaluation.

Resumen

Establecimiento y desarrollo de la colección núcleo del germoplasma de sésamo en China

Del total de las colecciones de los recursos de germoplasma de sésamo en China se ha establecido una colección núcleo. De las 4251 accesiones, se han seleccionado, de manera aleatoria, conjuntos de 50 accesiones. Para agrupar a las accesiones en siete ecotipos, se utilizaron distintos procedimientos jerárquicos con los datos sobre 14 características. Luego, se dividió la colección en 14 grupos según tres niveles de estratificación: origen, tipo de variedad y ecotipo. De esta manera se preseleccionaron 884 accesiones de la colección total que se plantaron en tres lugares de experimentación ecológica. Los resultados demuestran que un 97,85% de los valores resultantes eran iguales a los datos originales de pruebas precedentes. Una ulterior agrupación identificó 453 accesiones núcleo para la colección básica con un buen carácter representativo, confirmado por las características del zimograma. Los rasgos seleccionados son altamente heredables y tienen características distintas entre los materiales recolectados, todo ello permite emprender un estudio global, sistemático y completo de los recursos de germoplasma de sésamo para una mejora del trabajo de identificación y evaluación.

Introduction

Sesamum indicum L. is the only cultivated species in the genus *Sesamum*. Sesame has been grown for over 2200 years in China and is one of the four major oil crops (rapeseed, soyabean, peanut and sesame) with an oil content up to 60%. In China, the sesame production area was 800 000-900 000 ha with an average production of 568 400 t/year for the years 1994-96. China's sesame exports account for 10-15% of the world sesame trade.

Extensive collecting and studies on the conservation, evaluation and utilization of sesame germplasm genetic diversity have been conducted in China. In the 1950s, following the collecting of more than 3000 Chinese landraces, a group of good resources such as Shangcai Zihuaye 23, Nanyang Badacha and Wuchang Jiugentou were identified and used both for sesame production and for crossbreeding to produce cultivars with high yield, good quality, resistance to diseases and tolerance to waterlogging. As a result, good-quality cultivars such as 'Zhongzhi 7', 'Zhongzhi 8', 'Zhongzhi 10', 'Yuzhi 2', 'Yuzhi 4', 'Zhuzhi 1' and 'Jizhi 1' were released. During the last 10–15 years, the government has provided much support. So far about 4500 sesame accessions have been collected which originated from 29 provinces of China and more than 20 other nations. Both long-term (30-40 years) and medium-term (10-

15 years) conservation of this material has been carried out. Using this material 27 basic traits were identified including ecogeographic origin, agronomic traits, quality traits, and resistance to pests and diseases. A computer database with more than 110 000 important trait data was established and three volumes of The Catalogue of Varietal Germplasm Resources of Sesame in China were published in 1981, 1992 and 1997, respectively (Anon. 1981, 1992, 1997).

The ultimate objective of collecting, conservation and evaluation of crop germplasm resources is to use them. The establishment of a core collection in sesame is one of the effective approaches for utilization. Frankel (1984) first proposed the idea of developing core collections, i.e. collections which contain, with a minimum of repetitiveness, the maximum possible genetic diversity of a crop species and its wild relatives. Substantial research progress on core collections has been achieved in some crops such as peanut in the USA (Holbrook *et al.* 1993), perennial *Glycine* species and wheat in Australia and barley and okra in some international organizations (Hodgkin *et al.* 1995). During 1995-97, funded by IPGRI, we conducted a project on the establishment of a sesame germplasm core collection in China. The investigation included analysis of ecogeographic origins of the accessions and of the variation of important traits in the total sesame collection in China. A procedure for establishing a sesame germplasm core collection was systematically studied and a core collection established representing over 4000 accessions of sesame germplasm genetic diversity.

Research methods and results

Selection of clustering method

Sets of 50 accessions were randomly selected from the 4251 accessions which constituted the whole collection of sesame germplasm resources in China. Different hierarchical clustering procedures were used with data on 14 traits (plant type, no. of capsules/axil, no. of locules/capsule, hairiness on stem, flower colour, seed colour, capsule dehiscence, tolerance to waterlogging, resistance to *Fusarium* wilt, resistance to charcoal rot, growth period, 1000-seed weight, oil content and protein content). The dendrograms and distances of eight methods were obtained for the 50 accession sets tested and compared. Based on the comparison the Ward method was considered as the best for sesame germplasm resources as it had good evenness of distribution and expansion with little overlap between accessions from different ecological regions or with different traits.

Grouping and pre-selection of core collection

Out of the 4251 sesame accessions, 4073 originated from 29 provinces of China and 178 were introduced from 16 nations including Bulgaria, the USA and Japan. The 4073 accessions from China were grouped into seven ecotypes (Fig. 1). The whole collection was thus divided into 14 groups according to three levels of stratification: origin, varietal type and ecotype (Table 1).

The sesame germplasm materials of each of the 14 groups were clustered by the Ward method using data on the 14 traits listed above. A pre-selected core was identified using three criteria: the selection of 20% of the total collection, the maximization of diversity differences between groups, and one accession was randomly selected from each cluster. In this way 884 accessions were pre-selected from the whole collection as a pre-selected core.

In 1995 and 1996, the pre-selected core collection was planted in three ecological experiment sites: North China (Shijiazhuang), Central China (Wuhan) and South China (Nanning) and data on qualitative and quantitative traits were compared with those obtained in previous trials. Plant type, no. of capsules/axil, no. of locules/capsule, hairiness on stem, flower colour and seed colour were investigated. The result showed that 97.85% of the values obtained were the same as the

Table 1. Grouping of sesame collection in China

Origin	Varietal type	Ecotypes	Groups
China	Landrace	Ecotype 1	1
		Ecotype 2	2
		Ecotype 3	3
		Ecotype 4	4
		Ecotype 5	5
		Ecotype 6	6
		Ecotype 7	7
	Improved		8
		European zone	9
		American zone	10
	Landrace	East Asia zone	11
		Southeast and	12
Overseas		South Asia zone	13
		African zone	
	Special genetic material		14

Fig. 1. Seven ecological regions of sesame in China: 1. Northeast and northwest spring sesame region with one crop in a year; 2. Northern spring sesame region with one crop in a year; 3. Huanghuai summer sesame region with two crops in a year; 4. Jianghan summer sesame region with two crops in a year; 5. Middle and lower reaches of Yangtze River summer intercropping and interplanting sesame region with two crops in a year; 6. South-central and southern spring, summer and autumn sesame region with two or three crops in a year; 7. Southwest plateau summer with partly spring and autumn sesame region.

original data from previous trails. Three quantitative traits, i.e. total growing period, plant height and 1000-seed weight for five groups (3,4,5,6, 7) were also measured. Four correlations were analyzed: mean of two years and the original data; 1995 and the original data; 1996 and the original data; 1996 and 1995. All correlations were positive and significant in the range of 60-100%. Means, ranges and variances of trait performance data also showed that the pre-selected core represented the diversity in the whole collection.

Final selection of core collection

Re-clustering was undertaken of the 14 selected qualitative and quantitative traits. The clusters were delimited using Ward's method and a final core collection and were identified by taking one accession at random from each cluster. Finally, 453 core accessions were identified for the core collection (Table 2).

Table 2. Number of accessions in different groups and in the core collection.

Group	No. accessions	No. in pre-selected core	No. in final selected core
1	462	93	42
2	910	178	78
3	543	114	55
4	691	145	69
5	463	104	56
6	537	110	57
7	378	81	47
8	89	23	18
9 - 13	153	29	24
14	25	7	7
Total	4251	884	453

Representative analysis of sesame germplasm core collection in China

Six values including maximum (Max), minimum (Min), range, mean, standard derivation (SD) and coefficient of variation (CV) of 14 traits of the 884 accessions pre-selected by the above procedure were similar to those of the whole collection. The pre-selected core collection generally represented the whole collection in genetic variation of the 14 traits. The percentages of 35 phenotypes in 10 qualitative traits such as plant type, no. of capsules/axil, no. of locules/capsule, flower colour, etc. of the final core collection were statistically analyzed and were similar to those of the pre-selected core. Six values (max, min, range, mean, SD and CV) of four quantitative traits including oil content, protein content, 1000-seed weight and total growing period in the final core collection were also similar to those of the pre-selected core. These results indicated that both the pre-selected core and the final core collection had good representativeness (Zhang Xiurong *et al.* 1998).

Isozyme studies on the core collection

The pre-selected core and final core collection were compared using esterase isozyme diversity on a small set of 18 accessions from 3 groups (1,4,6). Both included eight zymogram types and had band polymorphisms both among and within accessions. Similar percentages were found for each of seven bands in zone B. Overall the results indicated that the zymogram characteristics of the pre-selected core represented that of the reserve collection well (Table 3).

Table 3. Percentages of zymogram types of sesame esterase isozyme

Group	Source	Access. no.	Zymogram type and percentage							
			I	II	III	IV	V	VI	VII	VIII
1	Pre-selected core	1	0	98.3	0	0	0	0	1.7	0
		2	23.3	43.4	0	20	3.3	10	0	0
		3	3.3	65	10	0	0	21.7	0	0
	Reserve collection	4	5	10	30	0	0	46.7	0	8.3
		5	0	0	58.3	0	0	36.7	1.7	3.3
		6	0	35	30	0	35	0	0	0
4	Pre-selected core	7	26.7	0	31.6	35	5	1.7	0	0
		8	6.7	25	60	0	5	0	0	3.3
		9	13.3	65	15	5	0	0	1.7	0
	Reserve collection	10	35	38.3	5	18.4	3.3	0	0	0
		11	6.7	0	85	0	0	0	0	3.3
		12	95	0	5	0	0	0	0	0
6	Pre-selected core	13	68.3	0	0	31.7	0	0	0	0
		14	31.7	3.3	0	61.7	3.3	0	0	0
		15	0	0	0	0	0	90	10	0
	Reserve collection	16	53.4	13.3	0	13.3	0	0	20	0
		17	60	3.3	0	36.7	0	0	0	0
		18	25	58.3	0	5	8.4	0	0	3.3
	Average		25.2	25.5	18.3	12.6	3.5	11.5	1.9	1.5

Conclusions and future research needs

The material in collections of many plant genetic resources is not only plentiful but also possesses substantial repetitions of genetic diversity (up to 60% duplication has been reported in some collections). In this research, the core collection was selected by cluster analysis, which groups genetically similar material into the same clusters. The traits which are selected for this analysis should have high heritability and distinct characteristics among the collected materials. In determining the number of accessions selected using a cluster analysis one needs to ensure that a certain proportion is included and that differences in the distribution of diversity between groups and within a group are considered. Cluster distances can be used for this modification and the number of samples in the group with larger genetic difference can be increased or, when distances are small, reduced.

The Chinese sesame germplasm core collection will be provided with long-term and medium-term storage and regenerated together with the reserve collection. A catalogue and the accessions will be distributed where requested. It is intended that the core collection be relatively stable, providing a working sample set for genebank managers and other users.

The major objective of establishing the sesame germplasm core collection is to use it effectively. In this research, 10.7% of the whole collection has been selected out which makes it possible to undertake an all-round, systematic and thorough study of sesame germplasm resources and to enhance identification and evaluation work. With this number of accessions the polymorphism of DNA molecular markers can be used to analyze genetic diversity at the molecular level. Combining the conventional characterization and evaluation with molecular marker technology, sesame germplasm genetic diversity can be studied to determine the extent, distribution regularities, and relations between ecogeographical differences and genetic backgrounds. These studies can provide technical guidance and a theoretical basis for further collecting, effective protection and rational use of sesame genetic resources in breeding.

Acknowledgements

This project was supported by IPGRI. Dr Toby Hodgkin at HQ and Professor Zhou Mingde at IPGRI East Asia Office have given a lot of technical guidance.

References

Anon. 1981. The Catalogue of Varietal Germplasm Resources of Sesame in China, Vol. 1. Oil Crops Research Institute of CAAS. China Agricultural Scientech Press.

Anon. 1992. The Catalogue of Varietal Germplasm Resources of Sesame in China, Vol. 2. Oil Crops Research Institute of CAAS. China Agricultural Scientech Press.

Anon. 1997. The Catalogue of Varietal Germplasm Resources of Sesame in China, Vol. 3. Oil Crops Research Institute of CAAS. China Agricultural Scientech Press.

Frankel, O.H. 1984. Genetic perspectives of germplasm conservation. Pp. 161-170 *in* Genetic Manipulation: Impact on Man and Society (W.K. Arber *et al.*, eds.). Cambridge University Press, Cambridge, UK.

Holbrook, C.C., W.F. Anderson and R.N. Pittman. 1993. Selection of a core collection from the U.S. germplasm collection of peanut. Crop Sci. 33:859-861.

Hodgkin, T., A.H.D. Brown, Th.J.L. van Hintum and E.A.V. Morales (eds.). 1995. Core Collections of Plant Genetic Resources. John Wiley & Sons.

Zhang Xiurong, Guo Qingyuan and Zhao Yingzhong. 1998. Establishment of sesame germplasm core collection in China Scientia. Agricultura Sinica 31(3):49-55.

Plant Genetic Resources Newsletter

Aims and scope: The *Plant Genetic Resources Newsletter* publishes papers in English, French or Spanish, dealing with the genetic resources of useful plants, resulting from new work, historical study, review and criticism in genetic diversity, ethnobotanical and ecogeographical surveying, herbarium studies, collecting, characterization and evaluation, documentation, conservation, and genebank practice.

Management: The *Plant Genetic Resources Newsletter* is published under the joint auspices of the International Plant Genetic Resources Institute (IPGRI) and the Plant Production and Protection Division of the Food and Agriculture Organization of the United Nations (FAO).

Availability: The *Plant Genetic Resources Newsletter* appears as one volume per year, made up of four issues, published in March, June, September and December. Plant Genetic Resources Newsletter is available free of charge to interested libraries of genebanks, university and government departments, research institutions, etc. The periodical may also be made available to individuals who can show that they have a need for a personal copy of the publication.

Types of paper

Articles: An article will publish the results of new and original work that makes a significant contribution to the knowledge of the subject area that the article deals with. Articles, which should be of a reasonable length, will be considered by the Editorial Committee for scope and suitability, then assessed by an expert referee for scientific content and validity.

Short communications: A short communication will report results, in an abbreviated form, of work of interest to the plant genetic resources community. Short communications in particular will contain accounts of germplasm acquisition missions.

Other papers: The *Plant Genetic Resources Newsletter* will publish other forms of reports such as discussion papers, critical reviews, and papers discussing current issues within plant genetic resources. Book reviews will be printed, as well as a News and Notes section. Suggestions for books to review are invited, as are contributions to News and Notes.

Submission: In the first instance papers may be submitted in typescript form or as an Email message. The final version may be submitted as an Email file or as an MS-DOS-readable file on diskette. Manuscripts submitted for publication and other communications on editorial matters should be addressed to IPGRI's Editorial and Publications Unit.

Bulletin des ressources phytogénétiques

Domaine d'intérêt: Le *Bulletin des ressources phytogénétiques* publie des articles en anglais, en espagnol et en français, sur les ressources génétiques de plantes utiles, fruit de nouvelles recherches, d'études historiques, d'examens et de critiques concernant la diversité génétique, d'études ethnobotaniques et écogéographiques, d'études d'herbiers, d'activités de collecte, de caractérisation et d'évaluation, de documentation, de conservation et les pratiques des banques de gènes.

Parrainage: Le *Bulletin des ressources phytogénétiques* est publié sous les auspices de l'Institut international des ressources phytogénétiques (IPGRI) et de la Division de la production végétale et de la protection des plantes de l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO)

Distribution: Le *Bulletin des ressources phytogénétiques* paraît une fois par an en un volume regroupant quatre numéros publiés en mars, juin, septembre et décembre. Il est distribué gratuitement aux bibliothèques des banques de gènes, universités, services gouvernementaux, instituts de recherche, etc. s'intéressant aux ressources phytogénétiques. Il est aussi envoyé sur demande à tous ceux pouvant démontrer qu'ils ont besoin d'un exemplaire personnel de cette publication.

Types de documents publiés

Articles: Un article contient les résultats de travaux nouveaux et originaux qui apportent une contribution importante à la connaissance du sujet dont traite l'article. Les articles, qui doivent être d'une longueur raisonnable, sont d'abord examinés par le Comité de rédaction qui en évalue la portée et la validité, puis par un expert qui en examine le contenu et l'intérêt scientifiques.

Brèves communications: On entend par brève communication un texte contenant, sous une forme abrégée, les résultats de travaux présentant un intêrêt pour tous ceux qui s'occupent de ressources phytogénétiques. Elle contient en particulier des comptes rendus des missions d'acquisition de matériel génétique.

Autres documents: Le *Bulletin des ressources phytogénétiques* publie d'autres types de rapport tels que des documents de synthèse, des études critiques et des articles commentant des problèmes actuels concernant les ressources phytogénétiques. Le Bulletin publie une revue de livres ainsi qu'une section intitulée Nouvelles et Notes. Les auteurs sont invités à envoyer leurs suggestions pour les livres à passer en revue ainsi que des contributions aux Nouvelles et Notes.

Présentation: En premier lieu, les documents doivent être soumis dactylographiés ou par courrier électronique. La version définitive doit être présentée en fichier de courrier électronique ou sur disquettes compatibles MS-DOS. Prière d'adresser les manuscrits présentés pour être publiés et d'autres communications sur des questions de rédaction au Bureau de rédaction de l'IPGRI.

Boletín de Recursos Fitogenéticos

Objetivos y temas: El *Noticiario de Recursos Fitogenéticos* publica documentos en inglés, francés y español que tratan de los recursos genéticos de plantas útiles, fruto de nuevos trabajos, estudios históricos, revisiones y análisis críticos relacionados con la diversidad genética, investigaciones etnobotánicas y ecogeográficas, estudios de herbarios, actividades de colección, caracterización y evaluación, documentación, conservación, y prácticas en bancos de germoplasma.

Dirección: El *Noticiario de Recursos Fitogenéticos* se publica bajo los auspicios conjuntos del Instituto Internacional de Recursos Fitogenéticos y la Dirección de Producción y Protección Vegetal de la Organización de las Naciones Unidas para la Agricultura y la Alimentación.

Distribución: El *Noticiario de Recursos Fitogenéticos* aparece como un volumen anual compuesto por cuatro números, que se publican en marzo, junio, septiembre y diciembre. Se distribuye gratuitamente a las bibliotecas de bancos de germoplasma, facultades universitarias y servicios gubernamentales, centros de investigación, etc. que se interesan en los recursos fitogenéticos. También pueden obtener este noticiario las personas que demuestren necesitar una copia personal.

Tipos de documentos

Artículos: Los artículos divulgarán los resultados de trabajos nuevos y originales que contribuyan de modo importante al conocimiento del tema tratado. Dichos artículos, que deberán tener una longitud razonable, serán examinados por el Comité de Redacción en cuanto a su pertinencia e idoneidad y posteriormente un experto juzgará su contenido y validez científicos.

Comunicaciones breves: Las comunicaciones breves informarán de modo conciso sobre los resultados de trabajos de interés para las personas que se ocupan de los recursos fitogenéticos. Las comunicaciones breves incluirán, en particular, resúmenes sobre las misiones de adquisición de germoplasma.

Otros documentos: El *Noticiario de Recursos Fitogenéticos* publicará otros tipos de informes, como documentos de trabajo, análisis críticos, y documentos que examinen cuestiones de actualidad relacionadas con los recursos fitogenéticos. El Noticiario publicará una reseña de libros así como una sección de Noticias y Notas. Las propuestas de libros para reseñar y las contribuciones a la sección de Noticias y Notas serán bien acogidas.

Presentación: Los documentos deben entregarse, incialmente, en forma de texto mecanografiado o a través del correo electrónico. La versión final debe presentarse como un archivo de correo electrónico o en disquete compatible con el sistema operativo MS-DOS. Los manuscritos para publicar y otras comunicaciones sobre asuntos relativos a la redacción deberán dirigirse a la Oficina de Redacción del IPGRI.

Plant Genetic Resources Newsletter

No. 119 Supplement, December, 1999

Contents