Theor Appl Genet (2009) 118:193–204
DOI 10.1007/s00122-008-0887-z

ORIGINAL PAPER

Analysis of a diverse global *Pisum* sp. collection and comparison to a Chinese local *P. sativum* collection with microsatellite markers

Xuxiao Zong · Robert J. Redden · Qingchang Liu · Shumin Wang · Jianping Guan · Jin Liu · Yanhong Xu · Xiuju Liu · Jing Gu · Long Yan · Peter Ades · Rebecca Ford

Received: 20 September 2007 / Accepted: 2 September 2008 / Published online: 25 September 2008
© Springer-Verlag 2008

Abstract Twenty-one informative microsatellite loci were used to assess and compare the genetic diversity among *Pisum* genotypes sourced from within and outside China. The Chinese germplasm comprised 1243 *P. sativum* genotypes from 28 provinces and this was compared to 774 *P. sativum* genotypes that represented a globally diverse germplasm collection, as well as 103 genotypes from related *Pisum* species. The Chinese *P. sativum* germplasm was found to contain genotypes genetically distinct from the global gene pool sourced outside China. The Chinese spring type genotypes were separate from the global gene pool and from the other main Chinese gene pool of winter types. The distinct Chinese spring gene pool comprised genotypes from Inner Mongolia and Sha'anxi provinces, with those from Sha'anxi showing the greatest diversity. The other main gene pool within China included both spring types from other northern provinces and winter types from central and southern China, plus some accessions from Inner Mongolia and Sha'anxi. A core collection of Chinese landraces chosen to represent molecular diversity was compared both to the wider Chinese collection and to a geographically diverse core collection of Chinese landraces. The average gene diversity and allelic richness per locus of both the micro-satellite based core and the wider collection were similar, and greater than the geographically diverse core. The genetic diversity of *P. sativum* within China appears to be quite different to that detected in the global gene pool, including the presence of several rare alleles, and may be a useful source of allelic variation for both major gene and quantitative traits.

Communicated by M. Morgante.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0887-z) contains supplementary material, which is available to authorized users.

X. Zong · Q. Liu · S. Wang · J. Guan · J. Liu · Y. Xu · X. Liu · J. Gu · L. Yan
Institute of Crop Sciences/The National Key Facility for Crop Gene Resources and Genetic Improvement,
Chinese Academy of Agricultural Sciences,
Beijing 100081, People's Republic of China

X. Zong · Q. Liu
Department of Plant Genetics and Breeding,
China Agricultural University, Beijing 100094,
People's Republic of China

R. J. Redden
Australian Temperate Field Crops Collection,
Grains Innovation Park, The Department of Primary Industries,
Private Bag 260, Horsham, VIC 3401, Australia

P. Ades · R. Ford (✉)
BioMarka, Faculty of Land and Food Resources,
The University of Melbourne, Melbourne, VIC 3010, Australia
e-mail: rebeccaf@unimelb.edu.au

Introduction

Field pea (*Pisum sativum*) is grown widely throughout the world; the seed and pods are consumed as a rich protein food source and both the growing plant and straw may be used for stock feed. Adapted to a wide range of climates and altitudes, accessions of *Pisum* have been collected and maintained within several major well characterized collections worldwide. These include but are not limited to collections held at: John Innes Centre (JIC), UK; Nordic Gene Bank (NGB), Sweden; United States Department of Agriculture (USDA), USA; International Center for Agricultural Research in the Dry Areas (ICARDA), Syria and Vavilov Institute, Russia.

Many studies have been conducted on *Pisum* germplasm collections to investigate genetic and trait diversities. Protein and molecular markers were employed to assess genetic diversity among 148 *Pisum* accessions that represented both primitive and modern cultivated forms (Baranger et al. 2004). Cluster analysis enabled discrimination of pea types corresponding to end-use such as fodder, food and animal feed, and separated the wild and primitive forms. More specifically, the spring-sown feed types were clearly differentiated from the winter-sown feed types. They also reported that released cultivars and breeding lines contained far less diversity than the rest of the collection that represented the wider *Pisum* genepool.

In other studies, Keneni (Keneni et al. 2005) assessed a collection of 148 Ethiopian *P. sativum* genotypes for 12 agronomically important traits and detected significant differences in all, except for the number of seeds per pod. Based on these traits, clusters of genotypes were formed but there was no obvious relationship between agronomic clusters and geographic origin. A more recent diversity study among 24 elite Indian *P. sativum* genotypes using 60 random amplified polymorphic DNA (RAPD) markers separated groups corresponding to tall and dwarf type varieties. The similarity detected between pairs of accessions ranged from 60 to 87% (Choudhury et al. 2007). In another study of 14 Indian and one UK *P. sativum* accession, that differed for seed characters and geographic origin, the range of detected pair-wise similarity was somewhat broader, from 26 to 79% using 11 RAPD primers (Yadav et al. 2007).

Meanwhile, a large Chinese collection of *Pisum* sp. accessions is held at the National Genebank of China (NGC), Institute of Crop Sciences (ICS), Chinese Academy of Agricultural Sciences, Beijing, China. This collection comprises accessions accumulated since the 1950s from 28 Chinese provinces. Genetic variability within this collection has not previously been assessed, nor has the Chinese germplasm been compared to that from outside China. As such, the Chinese accessions may represent potentially novel and previously untapped *Pisum sativum* ssp. *sativum* genetic material for the rest of the world. The identification of potentially novel germplasm will be invaluable in the search for desirable or novel traits for application to field pea improvement programs. An understanding of the geographical distribution of genetic diversity within the Chinese *P. sativum* collection, similar to that obtained for wild lentil germplasm by Ferguson (Ferguson et al. 1998a), would help guide the targeting of specific areas for future collection missions, to increase diversity in ex situ collections and to target areas for in situ conservation. Also, development of a core collection, representative of the larger Chinese *P. sativum* germplasm collection will enable its more efficient conservation and dissemination.

First detailed by Frankel and Brown (1984), a core collection aims to represent and maintain the maximum possible genetic diversity that may be detected within a larger comprehensive germplasm collection, by selecting a smaller subset of representative genotypes. More recently, van Hintum (1999) modified the core collection definition to "a germplasm collection optimally representing specific genetic diversity". Core collections of legume species have been defined using random sampling and stratified sampling strategies in lentil (Erskine and Muehlbauer 1991) and using evolutionary, agroecological and molecular data in common and wild bean (Tohme et al. 1995, 1996). Baranger (Baranger et al. 2004) found that a core of 43 accessions could explain 96% of the diversity detected by protein and molecular markers in the total collection of 148 field pea accessions.

Measures of variation at the molecular level are very suited to assessing relationships among individuals and constructing representative smaller collections. This may be achieved by detecting differences within the abundant and often degenerate DNA repeat sequences of the genome. Molecular variation in the John Innes Centre *P. sativum* core collection was assessed by comparing Ty1-*copia* retrotransposable element sequences (Ellis et al. 1998). The usefulness of the Ty1-*copia* sequence was proposed due to it being structurally conserved but having high insertion site polymorphism. Pearce (Pearce et al. 2000) went on to further characterize the Ty1-*copia* retrotransposable element sequences in pea. Diversity among accessions within the John Innes *Pisum* core collection (Matthews and Ambrose 1995) was assessed by examining differences in the insertional sequence of the *PDR1* element. Using sequence specific amplification polymorphism (SSAP) within various retrotransposon element variable regions, a detailed *Pisum* phylogeny was developed (Pearce et al. 2000).

Similarly, the potentially high rate of evolution due to "slippage" in pre-transcription binding events means that short sequence repeat (SSR) or microsatellite loci provide excellent markers for discriminating closely related genotypes within the *Pisum* species. In an initial study by Lu (Lu et al. 1996), PCR-based microsatellite markers were generally found to be more informative for assessing genetic relationships within *Pisum* than were restriction fragment length polymorphism (RFLP) type markers. Subsequently, Burstin (Burstin et al. 2001) retrieved 171 microsatellite sequences for *P. sativum* from the published databases and used these to determine the genetic relationships among 12 *P. sativum* genotypes. Three distinct genetic groups were revealed comprising accessions from Afghanistan as well as spring sown and fodder-type.

An abundance of *P. sativum* microsatellite locus-specific primer pairs have been generated through the construction of a genomic library enriched for repeat motifs by the Agrogène® consortium (Moissy Cramayel, France). A subset was

used for genetic mapping of resistance to Aphanomyces root rot in the USA (Pilet-Nayel et al. 2002) and to determine genetic relationships among commonly grown Australian *P. sativum* cultivars and relative *P. fulvum* genotypes (Ford et al. 2002).

In this study, we report the use of 21 microsatellite loci to characterise genetic diversity and relationships among a collection of 2120 diverse *Pisum* genotypes. Specifically, the aims of this research were to (1) determine the position of a Chinese *Pisum sativum* collection within a larger collection representative of global *Pisum* accessions acquired from diverse sources in Australia and worldwide, including wild species and subspecies, (2) more specifically, determine the relatedness of Chinese *P. sativum* ssp. *sativum* germplasm to *P. sativum* ssp. *sativum* from outside China and (3) construct a core collection of genotypes representative of the molecular and geographical diversity within the wider Chinese collection.

Materials and methods

Plant materials

Pisum genotypes included in this study were chosen based on broad geographic origin. These were sourced from within the Australian Temperate Field Crops Collection (ATFCC), Department of Primary Industries (DPI), Horsham, Victoria, Australia (224 accessions) and the National Genebank of China (NGC), Institute of Crop Sciences (ICS), Chinese Academy of Agricultural Sciences, Beijing, China as representative of diverse global collections. Also included were 133 elite field pea breeding lines from the Victorian Department of Primary Industries (VicDPI) breeding program, 345 external world sourced accessions from NGC, and 72 vegetable pea accessions from diverse geographic origins (774 *P. sativum* genotypes in total). In addition, wild relative genotypes that were classified as belonging to *P. sativum* subsp. *abyssinicum*, *P. sativum* subsp. *asiaticum*, *P. sativum* subsp *elatius*, *P. sativum* subsp. *transcaucasicum*, *P. sativum* var. *pumilio*, *P. sativum* var. *arvense* or *P. fulvum* were included (103 genotypes). Finally, from within the ICS, 1243 *P. sativum* genotypes were included. These were collected within China, across 28 provinces and represented a geographic subset of one-third of the 3,650 entire ‘Chinese local’ collection. In total, 2120 *Pisum* genotypes were selected for molecular analyses (Table 1; Supplementary Table A).

Genomic DNA extraction

Plants were grown in plots at the CAAS Bashang germplasm nursery and leaf tissues were directly stored in liquid N_2 upon collection. Young leaf tissue was collected and bulked from 20 seedlings of each of the 2120 genotypes. Total genomic DNA was extracted from leaf tissue (100–200 mg) by grinding in liquid nitrogen and applying the adapted CTAB method of Taylor (Taylor et al. 1995). DNA quality and quantity were assessed with a spectrophotometer and on 1.4% agarose gel, stained with ethidium bromide and visualized under UV-light. Genomic DNA stocks were subsequently diluted to working solutions of 10 ng/μL.

PCR and microsatellite marker genotyping

Initially, 111 microsatellite primers pairs (*Pisum* microsatellite consortium, Agrogène®, France) were screened to identify those able to amplify clear, reproducible poly-allelic amplicons across a subset of twelve genotypes originating from geographically diverse regions and comprising representatives of both the cultivated and wild relative *Pisum* species and subspecies (Table 2).

Each PCR reaction (25 μL) contained 40 ng of DNA, 0.4 μM of each primer, 240 μM each of dATP, dCTP, dGTP and dTTP (Invitrogen, Australia), 3.0 mM $MgCl_2$, buffer (20 mM Tris–HCl pH 8.4, 50 mM KCl) and 1.0 U of *Taq* DNA polymerase (Invitrogen, Australia). Amplification was conducted in a MJ Research PTC-220 thermocycler with the following cycling profile: initial denaturation step at 94°C for 3 min followed by 35 cycles of 94°C for 30 s, optimized annealing between 59 and 68°C for 30 s and extension at 72°C for 1 min with a final extension at 72°C for 5 min. Amplified fragments were resolved on 1.4% agarose gel as previously mentioned. Selected primer sets were subsequently screened across the entire 2120 samples using optimized primer annealing temperatures and the amplicons were resolved on polyacrylamide gel (5%) and stained with silver (Promega, USA). Allele identities were assigned and approximate allele sizes were determined by comparison with a 50 bp step ladder (Promega, USA). Informative measures, including observed number of alleles, effective number of alleles (Kimura and Crow 1964) and Shannon’s information index (Lewontin 1972) were calculated for each of the 21 SSR loci.

Molecular variation among the 2120 *Pisum* genotypes

Clustering, and principle components analysis (PCA) were used to visualise the genetic relationships within the diverse global *Pisum* collection and to compare germplasm within the Chinese local *P. sativum* collection, while differences between a priori geographic groups of genotypes were tested by analyses of molecular variance (AMOVA). The size of the data set made using conventional genetic software either too ponderous or impossible, so the genetic

Table 1 Summary of germplasm used for molecular analyses (The full germplasm list is available on request to the corresponding author)

Source	*Pisum* species	Geographic source[b]
Globally diverse germplasm from The ATFCC, the ICS and VicDPI	*Pisum sativum* subsp. *sativum* (774)	Afghanistan (8), Albania (2), Armenia (2), Australia (148), Austria (1), Belarus (1), Bolivia (1), Bulgaria (17), Burundi (1), Canada (8), Chile (4), China (7), Colombia (1), Czechoslovakia (4), Denmark (2), Egypt (1), Estonia (1), Ethiopia (42), France (13), Georgia (9), Germany (49), Greece (6), Hungary (5), ICARDA (8), Iceland (1), India (17), Iran (2), Israel (2), Italy (2), Japan (6), Kazakhstan (7), Kenya (1), Kyrgyzstan (3), Libya (1), Madagascar (1), Malaysia (1), Mexico (1), Mongolia (1), Morocco (3), Myanmar (1), Nepal (29), Netherlands (5), New Zealand (7), Pakistan (8), Peru (1), Philippines (1), Poland (12), Portugal (2), Romania (7), Russian Federation (21), Rwanda (1), Spain (30), Sudan (1), Sweden (6), Syria (8), Tajikistan (3), Tanzania (1), Tunisia (1), Turkey (22), Uganda (1), Ukraine (3), United Kingdom (31), United States (142), Uzbekistan (7), Yugoslavia (4), Zaire (1), Zambia (1), Unknown[a] (27)
	Pisum sativum subsp. *abyssinicum* (17)	Ethiopia (5), Israel (1), Morocco (8), Poland[b] (1, WT 1207), United Kingdom[b] (1, JI 250), Unknown[a] (1)
	Pisum sativum subsp. *asiaticum* (27)	Afghanistan (15), Algeria (3), Egypt (2), Russian Federation (1), Tajikistan (5), Unknown[a] (1)
	Pisum sativum subsp. *elatius* (11)	Georgia (5), Israel (1), Palestine (1), Poland[b] (1, WT 12076), Russian Federation (1), Unknown[a] (2)
	Pisum sativum subsp. *transcaucasicum* (8)	Russian Federation (5), Sweden[b] (2, WT 807, L 1489), Unknown[a] (1)
	Pisum sativum var. *arvense* (20)	Canada (1), China (1), Finland (17), Turkey (1)
	Pisum sativum var. *pumilio* (7)	Israel (1), Palestine (1), Russian Federation (1), Turkey (3), Unknown[a] (1)
	Pisum fulvum (13)	Israel (5), Turkey (1),United Kingdom[b] (2, WT 302, WT 303), Unknown[a] (5)
Chinese local germplasm from ICS	*Pisum sativum* subsp. *sativum* (1243)	Winter sowing (449): Anhui (35), Chongqing (2), Fujian (1), Guangdong (7), Guangxi (14), Guizhou (37), Henan (49), Hubei (46), Hunan (6), Jiangsu (2), Jiangxi (2), Shanghai (12), Sichuan (177), Taiwan (3), Yunnan (55), Zhejiang (1) Spring sowing (794): Beijing (15), Gansu (33), Hebei (2), Heilongjiang (1), Inner Mongolia (263), Liaoning (7), Ningxia (3), Qinghai (189), Shanxi (19), Sha'anxi (209), Tibet (45), Xinjiang (8),

[a] Note that the geographic origin of several genotypes was either not known due to incomplete passport data or conflicted among database records (*unknown*)

[b] Source equates to origin where passport data is complete, but if incomplete equates to donor genebank (Suspect sources have suffix[b] and respective synonyms)

Table 2 *Pisum* genotypes used for initial microsatellite primer screening

Accession no.[a]	Geographic source[b]	Taxon
ATC968	Italy	*P. sativum* subsp. *sativum*
ATC1502	Turkey	*P. sativum* subsp. *sativum*
ATC2702	Afghanistan	*P. sativum* subsp. *sativum*
ATC2706	Mongolia	*P. sativum* subsp. *sativum*
ATC3988	Kazakhstan	*P. sativum* subsp. *sativum*
ATC4056	Morocco	*P. sativum* subsp. *sativum*
ATC4322	Chile	*P. sativum* subsp. *sativum*
G000041	Henan province, China	*P. sativum* subsp. *sativum*
G000909	Yunan province, China	*P. sativum* subsp. *sativum*
ATC36	Ethiopia	*P. sativum* subsp. *abyssinicum*
ATC1372	Poland[b] (WT 12076)	*P. sativum* subsp. *elatius*
ATC3343	United Kingdom	*P. fulvum*

[a] *ATC* Australian Temperate Field Crops Collection, Australia and *G* Institute of Crop Sciences, China

[b] Source equates to origin where passport data is complete, but if incomplete equates to donor genebank (Suspect sources have suffix[b])

analyses were carried out using more general statistical software. The genotypic data were first arranged into a matrix with columns for every allele at every locus, and elements equal to the number of copies of that allele (i.e., 0, 1 or 2). Principal components analysis and ordination of the genotypes based on the variance-covariance matrix of the data was carried out using NTSYSpc version 2.20 (Rohlf 2006).

For use in cluster analysis and AMOVA, a matrix with elements equal to half the pair-wise, squared Euclidean distance between genotypes (Smouse and Peakall 1999) was also calculated. A dendrogram of the genotypes base on this squared distance matrix was produced by hierarchical clustering using the unweighted pair-group method with arithmetic averages (UPGMA) in NTSYSpc version 2.20 (Rohlf 2006).

AMOVAs were conducted among subpopulations of *P. sativum* ssp. *sativum* that were a priori assigned, based on geographical origin, to determine any genetic compartmentalisation. Calculation of Φ_{ST} from the squared genetic distance matrix and testing of its significance by AMOVA was done using code written in the interactive matrix language (IML) of SAS version 9.1 (SAS Institute 2004). The AMOVA structure used is that for genotypic data; one group of populations and no within individual variation. Significance of Φ_{ST} was tested by randomly permuting genotypes between populations. Several AMOVAs were conducted on different subsets of the collection and to compare different groups including: (a) between germplasm sources outside and within China, (b) between winter and spring sown Chinese germplasm, (c) between all Chinese provinces excluding Inner Mongolia and Sha'anxi, (d) between only winter sown provinces, excluding Inner Mongolia and Sha'anxi, (e) between spring sown provinces, excluding Inner Mongolia and Sha'anxi and (f) between just Inner Mongolia and Sha'anxi provinces.

Private alleles were identified between several population groups: Between *P. sativum* accessions from within China versus outside China; between Chinese spring sown versus Chinese winter sown *P. sativum* accessions; and between the cultivated versus wild germplasm accession. This was achieved using GenAlEx version 6.1 software (http://www.anu.edu.au/BoZo/GenAlEx/).

Formulation and testing of a Chinese *P. sativum* core collection based on molecular versus geographic data

A dendrogram was constructed from the microsatellite data matrix and principle component analysis was conducted to visualise genetic relationships among only the 1243 Chinese local genotypes as described above. The dendrogram was cut at a suitable, subjectively chosen point along the genetic distance scale to identify a meaningful clustering structure and to select individuals within separate clusters for a representative core collection.

In order to determine that the Chinese core collection selected on microsatellite data was an adequate representation of the larger Chinese collection of 1243 genotypes, the microsatellite allele frequencies and fixation indices were compared. This was done using FSTAT version 2.9.3.2 (Goudet 1995) and POPGENE version 1.32 (Yeh and Boyle 1997). Another Chinese *P. sativum* core collection was selected from the 1243 genotypes based solely on geographic origin, with genotypes randomly selected as representative of each of the provinces. This comprised 146 genotypes, selected randomly and representative of all Chinese provinces.

Results

Microsatellite primer screening and PCR optimization

From the 111 Agrogène® primer pairs screened, 21 were chosen for reliable amplification and for genotyping the entire 2120 *Pisum* accessions (Table 1). In total, 115 alleles were detected across the 21 microsatellite loci. The 21 SSRs were chosen from initial screening on 12 genotypes (Table 2). The number of alleles observed per SSR locus across the 12 geographically diverse genotypes ranged from one to four with a corresponding range of effectiveness and information provided (Table 3).

Table 3 Informative measures of the 21 microsatellite loci assessed among 12 *Pisum* genotypes used for initial microsatellite primer screening

[a]Locus	[b]Na	[c]Ne	[d]I	Chromosomal location
PB14	4	3.57	1.33	LG-VII
PSAA18	4	3.60	1.33	LG-II, LG-V
PSAA175	4	3.60	1.33	LG-III, LG-V
PSAC58	3	3.00	1.10	LG-V
PSAC75	4	3.00	1.24	LG-I
PSAA219	1	1.00	0.00	LG-IV
PSAD83	4	2.29	1.07	LG-II
PSAD270	5	4.50	1.56	LG-III
PSAA456	1	1.00	0.00	LG-VII
PSAB23	5	4.45	1.55	LG-V
PSAB47	4	3.60	1.33	LG-I, LG-V
PSAA497	4	4.00	1.39	LG-V
PSAD280	2	2.00	0.69	LG-V
PSAB72	1	1.00	0.00	LG-II
PSAB109	3	2.00	0.87	LG-II
PSAB141	3	2.27	0.95	LG-III
PSAB161	2	1.32	0.41	N/a
AD100	2	1.80	0.64	N/a
AD134	4	2.29	1.07	LG-II, LG-III, LG-VII
AA303	1	1.00	0.00	N/a
AA315	2	2.00	0.69	LG-IV
Mean	3.0000	2.54	0.88	
St. dev.	1.3416	1.16	0.53	

[a] SSR primer pairs provided through the Agrogene®, France, *Pisum sativum* microsatellite consortium

[b] *NA* Observed number of alleles

[c] *NE* Effective number of alleles (Kimura and Crow 1964)

[d] *I* Shannon's information index (Lewontin 1972)

Genetic diversity and PCA analyses among the 2120 *Pisum* sp. genotypes, to compare world diversity with the 1243 Chinese local *P. sativum* genotypes

Cutting the UPGMA-generated dendrogram at a distance of 0.40 along the genetic scale (0–1.0) provided 214 distinct clusters among the 2120 genotypes (see Supplementary material Fig. 3). Selection of a single genotype from within each of the 214 clusters produced a core collection with only 10% of the individuals from the larger collection.

PCA clearly demonstrated the molecular differentiation of the *Pisum* genotypes sourced from within China to those sourced from outside China and held in the ATFCC and ICS collections (labeled as 'other' in Fig. 1). PCA showed clear differentiation of the spring- and winter-sown Chinese *P. sativum* genotypes. Further molecular discrimination was observed using only the first two PCA axes within the Chinese local *P. sativum* genotype collection whereby groups of genotypes originating from Sha'anxi and Inner Mongolia were formed, as well as groups of winter and spring sown genotypes (Fig. 2).

AMOVA analyses demonstrated highly significant molecular difference between all of the *P. sativum* ssp. *sativum* geographical subpopulations assessed, and at all levels ($P = 0.001$). Almost as much variation was detected within provinces or regions within China (minimum $V_W = 24.52$) as was detected within populations when germplasm was sourced outside China as compared to that sourced within China ($V_W = 25.51$; Table 4).

Construction of a Chinese *P. sativum* core collection

Placement of a vertical transect at a distance of 0.38 along the genetic scale (0–1.0) of the microsatellite diversity and UPGMA-generated dendrogram of the 1243 Chinese local *P. sativum* genotypes revealed 146 separate clusters (data not shown). Subsequently, a single genotype was randomly selected from within each cluster to produce a microsatellite-based core collection of 146 genotypes. This represented approximately 12% of the Chinese germplasm and included all 28 provinces (Table 5).

Characteristics of the Chinese microsatellite-based *P. sativum* core collection

The average gene diversity and allelic richness per locus for the 146 genotype microsatellite-based Chinese core collection was similar to or greater (0.67 and 4.81) than that detected within the entire 1243 genotype Chinese collection (0.66 and 4.51), and the 146 genotype geographic origin-based Chinese core collection (0.64 and 4.42). The average number of alleles sampled per locus was the same (4.95) in both the microsatellite-based core and the entire Chinese collection, but greater than that of the geographic-origin core (4.57; Table 6). This indicated that the microsatellite-based core (146 genotypes) was a better representative of the larger Chinese collection (1243 genotypes) than the geographic origin-based core (146 genotypes).

Genetic distinction of germplasm from within cultivated groups of different geographic origin and wild sources

Several private alleles were revealed among individuals within cultivated germplasm from different geographic origins, both within and outside China. One of the private alleles, only found in the cultivated Chinese germplasm, occurred at a frequency of 0.217. Also, two alleles were found that were entirely different among but conserved within either the winter or spring sown Chinese cultivated accessions. Private alleles were also found conserved among cultivated or wild germplasm sources with frequencies

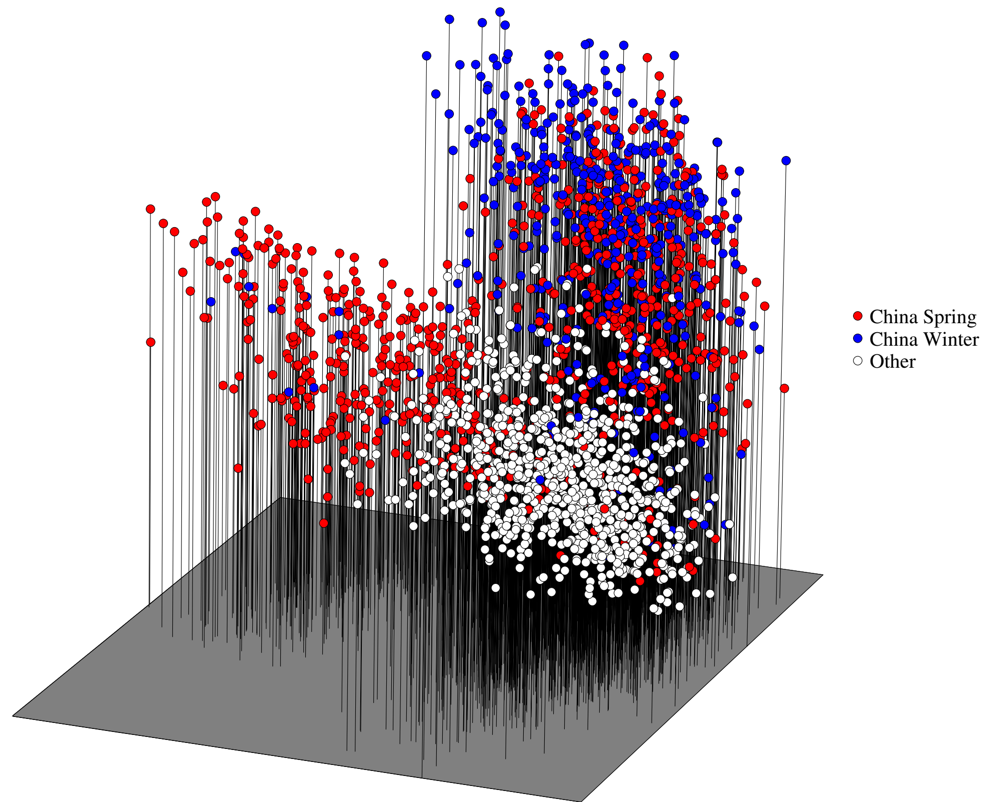

Fig. 1 Principal components analysis of molecular variation among *Pisum* genotypes within and outside China

ranging from 0.001 to 0.137. This indicated clear genetic differences among the cultivated gene pools within and outside China and the potential of genetic isolation. Clear differences were also observed between cultivated and wild accessions (Table 7).

Discussion

General observations

Microsatellite markers were successfully employed to visualise the genetic variation within a diverse *Pisum* sp. collection and a novel set of Chinese local *P. sativum* ssp. *sativum* genotypes and the relationships between them. The Chinese local *P. sativum* genotypes were genetically distinct from the *P. sativum* ssp. *sativum* and the broader *Pisum* sp. germplasm held within the ATFCC. Thus, the local Chinese *P. sativum* ssp. *sativum* genotypes represent unique germplasm, not previously accessed or utilised outside China.

Chinese *P. sativum* gene pools

Notably, several genetic groupings were observed within the Chinese collection. These included groups representing winter- (CW) and spring-sown types (CS), which had minimal overlap, together with very distinct clusters representing the Sha'anxi and Inner Mongolian-sourced landraces (SIM). The geographic origins of the SIM genotypes were different to those of the CS genotypes, which ranged from the Xinjiang and Qinghai provinces in north-western China to the Heilongjiang province in the north-east. Therefore, the discrimination of the molecular groupings was associated with broad geographic origin, since winter-sown genotypes are generally sourced from the 'southern' provinces of China and spring-sown genotypes are sourced from 'northern' provinces.

The occurrence of distinct genetic groups of field pea within China detected with microsatellite markers has remarkable features and possible evolutionary implications. *P. sativum* is commonly regarded as having originated in the Turkey–Iran Fertile Crescent (Ambrose 1995), with infrequent domestication events at this location as evidenced by limited chromosomal differentiation (Zohary 1973, 1999). However, the significant differentiation of the Chinese local *P. sativum* germplasm may indicate genetic isolation with directional selection, further introgression from wild relatives, or even independent domestication of this species outside the Fertile Crescent. The wider genetic diversity of the gene pool in China and its distinctiveness

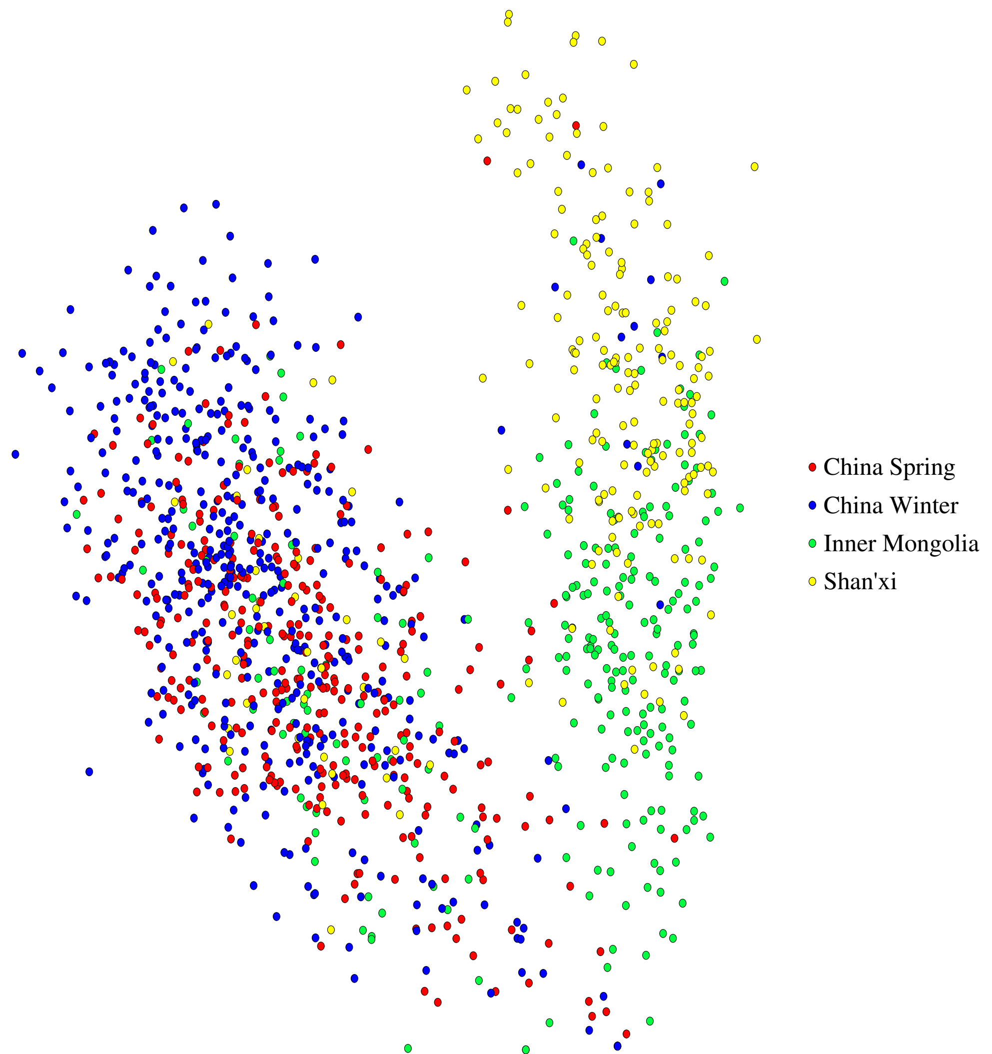

Fig. 2 Principal components analysis of molecular bivariation among *Pisum sativum* germplasm sourced within China

Table 4 Analyses of molecular variance between *P. sativum* ssp. *sativum* subpopulations within China and between germplasm sourced within and outside China

Population	Sub-populations	No. accessions	No. sub-populations	V_w	V_b	Φ_{st}
All *P. sativum* ssp. *sativum*	China vs. other	1987	2	25.51	2.36	0.09
China (Inner Mongolia and Sha'anxi excluded)	Winter vs. Spring sown	772	2	24.52	1.02	0.04
China (Inner Mongolia and Sha'anxi excluded)	Province	772	26	22.83	2.53	0.10
China winter-sown provinces (Inner Mongolia and Sha'anxi excluded)	Province	450	16	22.83	2.59	0.10
China spring sown provinces (Inner Mongolia and Sha'anxi excluded)	Province	322	10	22.82	1.88	0.08
Inner Mongolia and Sha'anxi only	Province	472	2	22.45	2.19	0.09

V_w Variance explained between sub-populations within the defined population, V_b Variance explained within a sub-population within the defined population, $\Phi_{st} = V_b/(V_b + V_w)$. Significant Φ_{st} values were greater than zero ($P < 0.01$)

from the rest of the world is remarkable given that the environmental and geographical diversity of regions where field peas are cultivated worldwide is at least as large as that within China (Redden et al. 2005).

Differentiation of gene pools in different regions within cultivated species has been noted in other crops (Islam et al. 2004; Rajaram 1999). This is thought to be associated with reproductive isolation, due to wide geographic separation, as well as to non-overlapping ecological adaptation. The distinct differentiation of the Chinese *P. sativum* genotypes may in part reflect historic isolation of agriculture in eastern Asia from that in southern Asia, Europe and northern

Table 5 Geographic representation within the Chinese *P. sativum* ssp. *sativum* germplasm, microsatellie-based core and the geographic origin-based core collections

Province	Chinese local collection (1243 genotypes)		Microsatellite-based Chinese core collection (146 genotypes)		Geographic origin-based Chinese core collection (146 genotypes)	
	No. of genotypes	% of collection	No. of genotypes	% of collection	No. of genotypes	% of collection
Winter sown						
Anhui	35	2.82	4	2.74	12	8.22
Chongqing	2	0.16	1	0.68	1	0.68
Fujian	1	0.08	1	0.68	1	0.68
Guangdong	7	0.56	2	1.37	1	0.68
Guangxi	14	1.13	1	0.68	2	1.37
Guizhou	37	2.98	5	3.42	13	8.90
Henan	49	3.94	7	4.79	9	6.16
Hubei	46	3.70	5	3.42	8	5.48
Hunan	6	0.48	2	1.37	1	0.68
Jiangsu	2	0.16	1	0.68	2	1.37
Jiangxi	2	0.16	1	0.68	2	1.37
Shanghai	12	0.97	3	2.05	1	0.68
Sichuan	177	14.24	22	15.07	12	8.22
Taiwan	3	0.24	3	2.05	1	0.68
Yunnan	55	4.42	11	7.53	11	7.53
Zhejiang	1	0.08	1	0.68	1	0.68
Sub-total	449	36.12	70	47.95	78	53.42
Spring sown						
Beijing	15	1.21	4	2.74	4	2.74
Gansu	33	2.65	3	2.05	12	8.22
Hebei	2	0.16	1	0.68	2	1.37
Heilongjiang	1	0.08	1	0.68	1	0.68
Inner Mongolia	263	21.16	15	10.27	11	7.53
Liaoning	7	0.56	2	1.37	1	0.68
Ningxia	3	0.24	1	0.68	2	1.37
Qinghai	189	15.21	19	13.01	9	6.16
Shanxi	19	1.53	2	1.37	7	4.79
Sha'anxi	209	16.81	17	11.64	14	9.59
Tibet	45	3.62	8	5.48	2	1.37
Xinjiang	8	0.64	3	2.05	3	2.05
Sub-total	794	63.88	76	52.05	68	46.58
Total	1243	100	146	100	146	100

Africa. Furthermore, the evolution of the very distinct SIM cluster within China raises questions on whether further introgression of wild relatives into the *P. sativum* gene pool may have occurred. Indeed, the related taxon *P. elatius* is sometimes classed within the species *P. sativum*, and pressed plant specimens from central Asia are present in the herbarium at the Vavilov Institute, St Petersburg, Russia (Redden personal communication, 2006). This suggests a need for collecting expeditions for wild relatives and remote landraces in northern and western China, and further investigation of the phenotypic and genetic characteristics of these gene pools.

A Chinese *P. sativum* core collection

Molecular variation within the Chinese local *P. sativum* collection was further employed to develop a microsatellite-based core collection, which when tested against the larger collection was shown to be highly representative. Such a core will be useful for future conservation, distribution and utilisation of the seemingly novel variation within this collection.

The lower allele diversity within the core based solely on geographic origin indicated that geographic origin alone was not the optimal way to select representative accessions.

Table 6 Gene diversity, allele richness and number of alleles sampled at each microsatellite loci assessed among the microsatellite-based core, the geographic origin-based core and the larger local Chinese genotype collection (1243)

Per locus measures	Gene diversity			Allelic richness			No. alleles sampled		
Locus	Microsatellite core	Geographic core	1243 collection	Microsatellite core	Geographic core	1243 collection	Microsatellite core	Geographic core	1243 collection
PB14	0.67	0.64	0.67	3.00	3.00	3.00	3	3	3
PSAA18	0.76	0.63	0.69	5.00	4.00	5.00	5	4	5
PSAA175	0.73	0.74	0.69	4.86	4.99	4.86	5	5	5
PSAC58	0.72	0.77	0.76	7.59	7.49	6.48	8	8	8
PSAC75	0.81	0.75	0.79	7.90	6.17	7.39	8	7	8
PSAA219	0.58	0.63	0.64	4.00	4.00	4.00	4	4	4
PSAD83	0.71	0.68	0.66	4.00	4.00	4.00	4	4	4
PSAD270	0.80	0.77	0.77	7.84	6.29	6.64	8	7	8
PSAA456	0.64	0.64	0.56	4.00	3.96	3.90	4	4	4
PSAB23	0.76	0.78	0.76	4.98	5.00	4.96	5	5	5
PSAB47	0.55	0.53	0.60	4.97	4.95	4.77	5	5	5
PSAA497	0.66	0.67	0.68	4.00	4.00	4.00	4	4	4
PSAD280	0.74	0.70	0.74	5.78	4.99	5.13	6	5	6
PSAB72	0.69	0.70	0.66	4.00	4.00	3.94	4	4	4
PSAB109	0.70	0.66	0.68	5.00	4.86	4.96	5	5	5
PSAB141	0.40	0.35	0.44	3.67	2.74	2.93	4	3	4
PSAB161	0.64	0.71	0.69	4.00	4.00	4.00	4	4	4
AD100	0.59	0.51	0.53	3.00	2.00	2.99	3	2	3
AD134	0.73	0.67	0.69	4.93	4.68	4.67	5	5	5
AA303	0.62	0.55	0.60	5.91	4.00	5.01	6	4	6
AA315	0.48	0.42	0.51	3.98	3.80	3.82	4	4	4
Mean	0.67	0.64	0.66	4.81	4.42	4.51	4.95	4.57	4.95

This may indicate that rare microsatellite alleles are found only in certain provinces and not evenly distributed across China. Thus, by selecting core members on origin alone, potentially rare but important genetic variation may be missed. Indeed, this is clear when examining the differences in amounts of variation identified from within each of the provinces (Table 4).

Novel *P. sativum* germplasm for future breeding objectives

The finding of genetically distinct gene pools (including private alleles) within domestic field pea potentially has significant implications for breeding and future genetic improvement. In wheat, two-way introgression between the winter and spring wheat gene pools led to wider access to allelic diversity for both disease resistance and for quantitative traits associated with adaptation and expression of increased yield (Fuentes-Davila et al. 1995; Rajaram 1999). In common bean, the Andean gene pool was found to be narrow, and large benefits should be gained from introgression from the Meso-American gene pool (Islam et al. 2004). In lentil, not only was there a founder effect and restricted genetic variability brought into the domestic gene pool, but further narrowing of the gene pool occurred with adaptation to specialized environments such as in Nepal and Bangladesh (Erskine 1997; Erskine et al. 1998; Ferguson et al. 1998b). A widening of the lentil gene pool in the breeding program for these countries has been very beneficial for yield improvement (ICARDA 2001; Erskine et al. 1998).

There has already been exchange of both pea germplasm from the ATFCC gene bank and the Australian pea breeding program to China, and a geographic core collection of Chinese landraces has been provided to Australia. In both countries, widening of their respective breeding gene pools with germplasm from the other has begun. This understanding of genetic diversity can now lead to a more systematic utilization of respective gene pools both for novel alleles for major gene traits such as disease resistances and abiotic stress tolerance, and quantitative expressions for adaptation and grain yield. These molecular studies provide an example of clarification of diversity relationships through use of

Table 7 Summary of private alleles present at SSR loci among populations

Population	SSR Locus	Allele	Frequency
P. sativum accessions from China versus the rest of the world			
China	PSAC75	7	0.009
China	PSAC75	8	0.217
China	AA303	5	0.005
China	AA303	6	0.001
Rest of the world	PB14	1	0.076
Rest of the world	PB14	5	0.006
Rest of the world	PB14	6	0.005
Rest of the world	PSAA18	6	0.058
Rest of the world	PSAA18	7	0.013
Rest of the world	PSAA175	6	0.005
Rest of the world	PSAA175	7	0.002
Rest of the world	PSAD83	5	0.023
Rest of the world	PSAD83	6	0.002
Rest of the world	PSAB72	5	0.007
Rest of the world	AD134	6	0.001
Chinese spring sown versus Chinese winter sown P. sativum accessions			
Chinese spring sown	AA303	6	0.001
Chinese winter sown	PSAD280	5	0.002
Cultivated versus wild accessions			
Cultivated	PB14	5	0.003
Cultivated	PB14	6	0.002
Cultivated	PSAA18	7	0.006
Cultivated	PSAC75	7	0.006
Cultivated	PSAC75	8	0.137
Cultivated	PSAA219	4	0.038
Cultivated	PSAA456	3	0.107
Cultivated	PSAA456	4	0.011
Cultivated	PSAD280	5	0.004
Cultivated	PSAD280	6	0.053
Cultivated	PSAB109	4	0.057
Cultivated	PSAB109	5	0.017
Cultivated	PSAB161	4	0.084
Cultivated	AA303	5	0.003
Cultivated	AA303	6	0.001
Wild	PSAA175	7	0.015
Wild	PSAD83	6	0.015

molecular markers, which would otherwise have been very difficult to demonstrate with phenotypic analysis alone.

Acknowledgments The funding received within the Australian Center for International Agricultural Research project CS1/2000/035 is gratefully acknowledged.

References

Ambrose MJ (1995) From Near East Center of Origin, the prized pea migrates throughout the world. Diversity 11:118

Baranger AG, Aubert G, Arnau G, Lainé AL, Deniot G, Potier J, Weinachter C, Lejeune-Hénaut J, Lallemand J, Burstin J (2004) Genetic diversity within *Pisum sativum* using protein- and PCR-based markers. Theor Appl Genet 108:1309–1321

Burstin J, Deniot G, Potier J, Weinachter C, Aubert G, Baranger A (2001) Microsatellite polymorphism in *Pisum sativum*. Plant Breed 120:311–317

Choudhury PR, Tanveer H, Dixit GP (2007) Identification and detection of genetic relatedness among important varieties of pea (*Pisum sativum* L.) grown in India. Genetica. doi:10.1007/s10709-006-9005-9

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of Ty1-*copia* class retrotranspo-

sons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19
Erskine W (1997) Lessons for breeders from land races of lentil. Euphytica 93:107–112
Erskine W, Muehlbauer FJ (1991) Allozyme and morphological variability, outcrossing rate and core collection formation in lentil germplasm. Theor Appl Genet 83:119–125
Erskine W, Chandra S, Chaudhry M, Malik IA, Sarker A, Sharma B, Tufail M, Tyagi MC (1998) A bottleneck in lentil: widening its genetic base in South Asia. Euphytica 101:207–211
Ferguson ME, Ford-Lloyd BV, Robertson LD, Maxted N, Newbury HJ (1998a) Mapping the geographical distribution of genetic variation in the genus *Lens* for the enhanced conservation of plant genetic diversity. Mol Ecol 7:1743–1755
Ferguson ME, Robertson LD, Ford-Lloyd BV, Newbury HJ, Maxted N (1998b) Contrasting genetic variation amongst lentil landraces from different geographical origins. Euphytica 102:265–273
Ford R, Le Roux K, Itman C, Brouwer JB, Taylor PWJ (2002) Genome-specific sequence tagged microsatellite site (STMS) markers for diversity analysis and genotyping in *Pisum* species. Euphytica 124:397–405
Frankel OH, Brown AHD (1984) Current plant genetic resources—a critical appraisal. In: Genetics: new Frontiers, vol IV. Oxford and IBH Publ. Co., New Delhi
Fuentes-Davila G, Rajaram S, Singh G (1995) Inheritance of resistance to karnal bunt (*Tilletia indica* Mitra) in bread wheat (*Triticum aestivum* L.). Plant Breed 114:250–252
Goudet J (1995) FSTAT (ver. 1.2): a computer program to calculate F-statistics. J Hered 86:485–486
International Center for Agricultural Research in the Dry Areas (2001) Annual report 2001
Islam F, Beebe S, Munoz M, Tohme J, Redden RJ, Basford KE (2004) Using molecular markers to assess the effect of introgression on quantitative attributes of common bean in the Andean gene pool. Theor Appl Genet 108:243–252
Keneni G, Jarso M, Wolabu T, Dino G (2005) Extent and pattern of genetic diversity for morpho-agronomic traits in Ethiopian highland pulse landraces: 1. Field pea (*Pisum sativum* L.). Genet Res Crop Evol 52:539–549
Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. Genetics 49:725–738
Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:381–398
Lu J, Knox MR, Ambrose MJ, Brown JKM, Ellis THN (1996) Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. Theor Appl Genet 93:1103–1111
Matthews P, Ambrose MJ (1995) Development and use of a "core" collection for the John Innes *Pisum* Collection. In: Proceedings of 2nd European conference on grain legumes, Copenhagen, Denmark, pp 194–195
Pearce SR, Knox M, Ellis THN, Flavell AJ, Kumar A (2000) Pea Ty1-*copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. Molec Gen Genet 263:898–907
Pilet-Nayel L, Muehlbauer FJ, McGee RJ, Kraft JM, Baranger A, Coyne CJ (2002) Quantitative trait loci for partial resistance to Aphanomyces root rot in pea. Theor Appl Genet 106:28–39
Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. Heredity 82:561–573
Rajaram S (1999) Approaches for breaching yield stagnation in wheat. Genome 42:629–634
Redden B, Leonforte T, Ford R, Croser J, Slattery J (2005) Pea (*Pisum sativum* L.). In: Singh RJ, Jauhar PP (eds) Genetic resources, chromosome engineering, and crop improvement. Grain legumes, vol 1, Chap. 3. CRC Press, Boca Raton, pp 49–83
Rohlf F (2006) NTSYSpc: numerical taxonomy system (ver. 2.2). Exeter Publishing, Ltd., Setauket
Taylor PWJ, Fraser TA, Ko H-L, Henry R (1995) RAPD analysis of sugarcane during tissue culture. In: Terzi M, Cella R, Falavigan A (eds) Current issues in plant molecular and cellular biology. Kluwer Academic Int, Dordrecht, pp 241–246
Tohme J, Jones P, Beebe S, Iwanaga M (1995) The combined use of agroecological and characterization data to establish the CIAT *Phaseolus vulgaris* core collection. In: Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV (eds) Core collections genetic resources. Wiley, Chichester, pp 95–107
Tohme J, Gonzales OD, Beebe S, Duque MC (1996) AFLP analysis of gene pools of a wild bean core collection. Crop Sci 36:1375–1384
Van Hintum TJL (1999) The general methodology for creating a core collection. In: Johnson RC, Hodgkin T (eds) Core collections for today and tomorrow. International Plant Genetic Resources Institute, Rome, pp 10–17
Yadav VK, Kumar S, Panwar RK (2007) Measurement of genetic dissimilarity in fieldpea (*Pisum sativum* L.) genotypes using RAPD markers. Genet Res Crop Evol. doi:10.1007/s10722-006-9109-1
Yeh FC, Boyle TBJ (1997) POPGENE Microsoft Windows-based software for population genetic analysis. A joint project development by Francis C. Yeh, University of Alberta and Tim Boyle, Center for International Forestry Research, Bogor, Indonesia, 29 p
Zohary D (1973) The origin of cultivated cereals and pulses in the Near East. In: Wahrman J, Lewis KR (eds) Chromosomes today, vol 4. Wiley, New York, pp 307–320
Zohary D (1999) Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the Near East. Genet Res Crop Evol 46:133–142