Genes & Genomics (2012) 34: 305-320
DOI 10.1007/s13258-011-0213-z

RESEARCH ARTICLE

Genetic diversity, population structure and genome-wide marker-trait association analysis emphasizing seed nutrients of the USDA pea (*Pisum sativum* L.) core collection

Soon-Jae Kwon · Allan F. Brown · Jinguo Hu · Rebecca McGee · Chasity Watt · Ted Kisha · Gail Timmerman-Vaughan · Michael Grusak · Kevin E. McPhee · Clarice J. Coyne

Received: 29 October 2011 / Accepted: 10 January 2012 / Published online: 15 June 2012
© The Genetics Society of Korea and Springer 2012

Abstract
Genetic diversity, population structure and genome-wide marker-trait association analysis was conducted for the USDA pea (*Pisum sativum* L.) core collection. The core collection contained 285 accessions with diverse phenotypes and geographic origins. The 137 DNA markers included 102 polymorphic fragments amplified by 15 microsatellite primer pairs, 36 RAPD loci and one SCAR (sequence characterized amplified region) marker. The 49 phenotypic traits fall into the categories of seed macro- and micro-nutrients, disease resistance, agronomic traits and seed characteristics. Genetic diversity, population structure and marker-trait association were analyzed with the software packages PowerMarker, STUCTURE and TASSEL, respectively. A great amount of variation was revealed by the DNA markers at the molecular level. Identified were three sub-populations that constituted 56.1%, 13.0% and 30.9%, respectively, of the USDA *Pisum* core collection. The first sub-population is comprised of all cultivated pea varieties and landraces; the second of wild *P. sativum* ssp. *elatius* and *abyssinicum* and the accessions from the Asian highland (Afghanistan, India, Pakistan, China and Nepal); while the third is an admixture containing alleles from the first and second sub-populations. This structure was achieved using a stringent cutoff point of 15% admixture (q-value 85%) of the collection. Significant marker-trait associations were identified among certain markers with eight mineral nutrient concentrations in seed and other important phenotypic traits. Fifteen pairs of associations were at the significant levels of $P \leq 0.01$ when tested using the three statistical models. These markers will be useful in marker-assisted selection to breed pea cultivars with desirable agronomic traits and end-user qualities.

Keywords Genetic diversity; Legume; Population structure; Association study; Core collection

S.-J. Kwon and A. F. Brown contributed equally to this work.

S.-J. Kwon · J. Hu · T. Kisha · C. J. Coyne (✉)
USDA-ARS Western Regional Plant Introduction Station, Mail Stop 646402, Washington State University, Pullman, WA 99164-6402, USA
e-mail: clarice.coyne@ars.usda.gov

A. F. Brown
Plants for Human Health Institute, North Carolina State University, North Carolina Research Campus, Suite 4229, 600 Laureate Way, Kannapolis, NC 28081, USA

R. McGee
USDA-ARS, Grain Legume Genetics Physiology Research, 303 Johnson Hall, Washington State University, Pullman, WA 99164, USA

C. Watt
Department of Crop and Soil Sciences, Johnson Hall 291D, Washington State University, Pullman, WA 99164, USA

G. Timmerman-Vaughan
New Zealand Institute for Plant & Food Research, Christchurch 8140, New Zealand

M. Grusak
USDA-ARS, Children Nutrition Research Center, Room 11074, 1100 Bates St, Houston, TX 77030, USA

K. E. McPhee
Department of Plant Sciences, Loftsgard Hall 370G, North Dakota State University, Fargo, ND 58108, USA

Introduction

Fifteen years ago, Tanksley and McCouch (1997) illuminated the underutilization of plant germplasm in crop improvement. Later, Frary et al. (2000) demonstrated the high value of plant germplasm collections by cloning the gene *fw*2.2 underlying

tomato fruit size quantitative trait loci (QTL) from a wild tomato progenitor. However, the first breakthroughs regarding the effective use of germplasm for crop improvement occurred in Pritchard's statistical treatment of population structure (Pritchard et al., 2000) and Buckler's maize association mapping studies (Thornsberry et al., 2001; Remington et al., 2001). Assessing the genetic diversity and population structure within germplasm collections provides an important association mapping resource for crop improvement, as well as a novel resource for germplasm management. This focused genotypic and phenotypic characterization of germplasm is essential to increase germplasm utilization as a means of responding to an ever-growing list of challenges facing food production in the next century (FAO, 2011).

Germplasm core collections are limited sets of accessions chosen to represent the genetic variation of crop species and wild relatives with minimum repetition (Brown and Spillane, 1999). These abbreviated collections constitute pre-determined, genetically diverse sets of accessions that can be readily provided to researchers who might otherwise lack the resources to evaluate large numbers of potentially redundant accessions. Core collections also represent a genetic baseline for evaluating the originality of new entries, for providing a systematic way of prioritizing the maintenance of large collections (Chavarriaga-Aguirre et al., 1999), and for taking advantage of new opportunities for efficient utilization of plant germplasm for crop improvement (Glaszmann et al., 2010). The USDA *Pisum* core collection (UPCC) consists of 285 accessions from 57 countries collected or donated between 1933 and 1986 (Supplementary Table S1). The collection was constructed initially using passport data and analyzed flower color (Simon and Hannan, 1995) and was later refined using RAPD markers (Coyne et al., 2005a). The UPCC has been evaluated extensively for a variety of agronomic and morphological traits -- seed weight, disease and pest resistances, stem and root traits, protein and micronutrient content -- and the results of these evaluations have been published in peer-reviewed publications (Jermyn and Slinkard, 1977; Kraft et al., 1998; Tedford and Inglis, 1999; McPhee and Muehlbauer, 2001; Malvick and Percich, 1999; McPhee et al., 1999; McPhee and Muehlbauer, 1999; Grünwald et al., 2003; Grusak et al., 2004; McPhee, 2005; Coyne et al., 2005b). This information is now available through the USDA Agricultural Research Service's Germplasm Resources Information Network (GRIN) National Plant Germplasm System database (www.ars-grin.gov/npgs/).

Traditional approaches to describing genetic diversity within core collections have generally involved the use of passport data, such as geographical origin or breeding pedigrees (Brown, 1989). These methodologies have been largely displaced by diversity estimates generated from phenotypic or genotypic data using cluster analysis or ordination (Bretting and Widrlechner, 1995). The use of quantitative phenotypic measurements, particularly those with low heritability, for diversity estimates has generated considerable controversy, as the measurements are often dependent on particular environments, thus rendering the relationship between the phenotype and genotype ambiguous (Smith and Smith, 1992). Selectively neutral molecular markers avoid these issues, but their value in the construction and use of core collections is dependent on the largely unknown joint distribution of the markers and target alleles of interest (Schoen and Brown, 1993). Verifying that these estimates faithfully represent the phenotypic diversity of traits of interest within the reserve collection can be difficult, as the true range and the variance of many agronomic traits are generally unknown. While experimental results are mixed, a number of studies have shown low correlations between the genetic variance of populations and the molecular distance of the parental material (Charcosset and Moreau, 2004). Hoey et al. (1996) suggested that the most reasonable initial approach was to include as many character types as possible in order to gain an unbiased and representative sample of the genome; however, non-significant correlations between diversity estimates generated with morphological characteristics and molecular markers have generally discouraged researchers from adopting this approach (Tar'an et al., 2005).

Core collections that accurately reflect the range of genetic diversity and phenotypic expression have the potential to serve as platforms for association studies that identify statistically significant relationships between polymorphic markers and genes of economic and biological merit (Myles et al., 2009). Association mapping techniques based on the linkage disequilibrium (LD), both for genome-wide and candidate-gene approaches, has recently emerged as an alternative approach to mapping QTLs and provides a powerful tool for dissecting quantitative traits in plants (Rafalski 2010). In plants, LD-based association mapping started with the model plant *Arabidopsis* and has now been extended to other crops (Abdurakhmonov and Abdukarimov, 2008). These tools were used to identify marker-trait associations in plant germplasm populations such as potato (Gebhardt et al., 2004), maize (Yu et al., 2005) and tomato (Mazzucato et al., 2008).

A rich literature has emerged surrounding the molecular diversity studies on *Pisum* germplasm collections, including studies of primarily national and regional collections (Lee et al., 1990; Samec et al., 1998; Ford et al., 2002; Simioniuc et al., 2002; Baranger et al., 2004; Tar'an et al., 2005; Ghafoor et al., 2005; Lazaro and Aguinagalde, 2006; Choudhury et al., 2007; Le Clerc et al., 2006; Esposito et al., 2007; Smýkal et al., 2008; Nisar et al., 2009; Dribnokhodova and Gostimsky, 2009; Martín-Sanz et al., 2011) and studies pertaining to the *Pisum* species and sub-species diversity (Lu et al., 1996; Ellis et al., 1998; Hoey et al., 1996; Pearce et al., 2000; Vershinin et al., 2003; Jing et al., 2005; Jing et al., 2007; Kosterin et al., 2010). These studies used a wide range of marker classes, including simple sequence repeats (SSRs), random amplified polymorphic DNA (RAPDs), inter-simple sequence repeat

(ISSR), amplified fragment length polymorphisms, restriction fragment length polymorphisms, allozymes, sequence-related amplified polymorphisms, retrotransposon-based markers and gene-based markers. Baranger et al. (2004) discussed their collection as primarily western European; however, their analysis was the first significant contribution to understanding global genetic diversity in pea using 148 accessions from 16 countries. The genetic diversity of 1,221 pea landraces from China (Zong et al., 2008) was analyzed using 21 SSRs, and the analysis was expanded to 1,234 global pea accessions using the same set of SSRs presented (Zong et al., 2009). The following year, Jing et al. (2010) published a larger global *Pisum* collection study of 3,020 accessions using retrotransposon-based markers. A combined Bayesian analysis of published studies on *Pisum* genetic diversity extended the structural view of the global pea genetic diversity to 4,429 accessions, which were included in a review published by Smýkal et al. (2011).

In this study, the focus is on distilling the molecular diversity of a global core collection of primarily landrace *Pisum* germplasm collected or donated over six decades in the twentieth century. The objective was to use the USDA *Pisum* core collection to unravel the genetic basis of quantitative and qualitative traits utilizing an association analysis. The population was characterized by 24 molecular markers, and 49 phenotypic traits were studied to (1) investigate the genetic diversity; (2) estimate the levels of population structure; and (3) evaluate this collection for association analysis.

Materials and Methods

Plant material

285 accessions representing approximately 10% of the UPCC catalogued during the past century was obtained from the Western Regional Plant Introduction Station, United States Department of Agriculture (USDA) Agricultural Research Service, Pullman, WA, USA. Accessions were collected or donated from 57 countries between 1933 and 1986 (see Supplementary Table 1S). Eleven accessions of the *Pisum sativum.* ssp. e*latius*, one accession of *P. sativum.* ssp. *abyssinicum*, one accession of *P. sativum.* ssp. *arvense* and six accessions identified as *P. s.* ssp *sativum* were included in the sample.

Molecular data collection and analysis

DNA was isolated from ten field-grown plants of each accession using a modified CTAB procedure (Murray and Thompson, 1980). Genotypes were collected using fifteen microsatellite primer pairs selected due to their high number of published polymorphisms (Burstin et al., 2001; Gilpin et al., 1997), as described in Loridon et al. (2005) (Table 1). As many of the accessions within the USDA *Pisum* collection represent landraces and unimproved semi-wild material, mixtures of homogeneity were expected, and SSR fragments were scored as present or absent using Gene Profiler version 4.05 (Scanalytics, Fairfax, VA).

RAPD genotyping was conducted using primers obtained from University of British Columbia (designated: UBC-fragment size) and from Operon Technologies (Alameda, CA, USA) (designated: kit-fragment size) (Table 1). PCR conditions were as described in Pilet-Nayel et al. (2002). One SCAR (sequence characterized amplified region) marker Y15_999*Fw* linked to Fusarium wilt race 1 was also used (Okubara et al., 2005).

Collection of phenotypic data

Phenotypic data included in the study for association analysis was collected at multiple locations and from publications in peer-reviewed journals (Table 2). The traits include concentration of seed macro- and micronutrients (protein, calcium, magnesium, potassium, phosphorus, iron, zinc, manganese, copper, nickel, boron and molybdenum) (Grusak et al., 2004; Coyne et al., 2005b); disease resistance ratings for Fusarium root rot caused by *Fusarium solani* (Mart.) Sacc. f. sp. *pisi* (F. R. Jones) W. C. Snyder & H. N. Hans (Grünwald et al., 2003); Fusarium wilt race 1 and race 2 caused by *Fusarium oxysporum* Schlecht. f. sp. *pisi* (C.J.J. Hall) Snyder and Hansen (McPhee et al., 1999); Aphanomyces root rot caused by *Aphanomyces euteiches* Drech., (Malvick and Percich, 1999); Ascochyta blight caused by *Mycosphaerella pinodes* (Berk. & Bloxam.) (Kraft et al., 1998) and pea cyst nematode *Heterodera goettingiana* Liebscher (Tedford and Inglis, 1999); agronomic and morphological traits (taxon, stem basal branching, seeds per pod, seed coat color, seed position per pod, seed pattern color, plant height, hilum color, flowers per peduncle, flower color, days to maturity, days to flower, biomass, 100 seed weight, cotyledon color and pod wall neoplasia) (McPhee and Muehlbauer, 1999; Coyne et al., 2005a); root and stem traits (McPhee and Muehlbauer, 1999; McPhee, 2005); and taxon. The data was downloaded from the USDA Agricultural Research Service Germplasm Resources Information Network (GRIN; www.ars-grin.gov/npgs) (Table 2). The final phenotypic value of each accession was calculated by average per each investigated year. The distributions of 49 traits of the USDA Pisum core collection are described in Supplementary Table 1S, and descriptive statistics for the 49 phenotypic traits are described in Table 3.

Data analysis

Characterization of the population genetic structure of the UPCC was accomplished using the software package STUCTURE 2.3.3 (Pritchard et al., 2000; Falush et al., 2003; Falush et al., 2007; Hubisz et al., 2009) that utilizes a Bayesian algorithm to assign accessions to putative populations (*k*) in Hardy-Weinberg equilibrium and that can provide an estimate

Table 1. Marker and primer details of fifteen microsatellites, eight RAPDs and one SCAR marker with the number of alleles and polymorphic information content (PIC) detected in the USDA *Pisum* core collection.

Marker (NCBI Accession)	Type	Sequence[a]	Chromosome	Previously described[b]	Alleles	PIC
PSU81288 (U81288)	EST-SSR	F:cgccatggagcttagcttcc R:cgagtagatagaagaagatgc	Composite map (LG I)[d]	3	6	0.15
PSGAPA1 (X15190)	EST-SSR	F:gacattgttgccaataactgg R:ggttctgttctcaatacaag	Composite map (LG V)	6	3	0.17
PSAJ3318 (AJ223318)	EST-SSR	F:cagtggtgacagcagggccaag R:cctacatggtgtacgtagacac	DP x JI296 (LG III)	3	5	0.20
AF016458 (AF016458)	EST-SSR	F:cactcataacatcaactatctttc R:cgaatcttggccatgagagttgc	Composite map (LG I)	5	6	0.17
AA430902 (AA430942)	EST-SSR	F:ctggaattcttgcggtttaac R:cgttttggttacgatcgagcta	Gilpin et al 1997 (LG IA/II)	4	5	0.20
PSU51918 (U51918)	EST-SSR	F:gtcgtaacagatcaatatggc R:cgatagtgagagtggcggttg	Composite map (LG I)	4	5	0.20
PEAOM14A (M69105)	EST-SSR	F:ggtgccctagcatttgtctg R:tagtaacaaccgcgctcaaa	Not mapped	3	3	0.04
PS11824 (Y11824)	EST-SSR	F:accaccaccaccgagaagat R:tttgtggcaatggagaaaca	Not mapped	2	2	0.04
PSU81287 (U81287)	EST-SSR	F:agagacaccggaagatcgag R:catccccatagccaccac	DP x JI296 (LG III)	2	2	0.24
AF004843 (AF029243)	EST-SSR	F:ccatttctggttatgaaaccg R:ctgttcctcattttcagtggg	Composite map (LG VII)	4	4	0.29
P446a2 (AA430935)	EST-SSR	F:tttttctagagcaattcatcctcc R:agccatacaattattgcctaaagg	Gilpin et al 1997 (LG VI)	2	3	0.14
PSBLOX13.2 (X78581)	EST-SSR	F:ctgctatgctatgtttcacatc R:ctttgcttgcaacttagtaacag	Bordat et al 2011 (LG III)	6	7	0.19
PSAS (Y13321)	EST-SSR	F:ggtgataactatttggctcatc R:gtagatttctccattcacctg	DP x JI296 (LG III)	3	3	0.16
PEACPLHPPS (L19651)	EST-SSR	F:gtggctgatcctgtcaacaa R:caacaaccaagagcaaagaaaa	Not mapped	4	5	0.21
PSARGDECA (Z37540)	EST-SSR	F:ctgttcctctttcaagcactcc R:gggaaagcaaagcatgcggatc	Not mapped	4	5	0.15
			Mean	3.6	4.3	0.17
R12	RAPD	acaggtgcgt	Térèse x K586[e] (LG III)		5	0.33
R13	RAPD	ggacgacaag	Térèse x K586 (LG III, IV)		5	0.32
V17	RAPD	gaagccagcc	Térèse x K586 (LG III, IV)		2	0.31
Y14	RAPD	ggtcgatctg	Térèse x K586 (LG III, IV)		8	0.30
L13	RAPD	accgcctgct	Térèse x K586 (LG III, VII)		1	0.37
J12	RAPD	gtcccgtggt	Térèse x K586 (LG III)		6	0.26
UBC467	RAPD	agc acg ggc a	Not mapped		4	0.36
UBC498	RAPD	gac agt cct g	Not mapped		6	0.33
			Mean		4.6	0.32
Y15 999*Fw* (DQ189096)	SCAR	F-atgagggtagcgcttcattg R-gccctttgttgtctcacctg	LG III[f]		2	0.33

[a] F = forward sequence; R = reverse sequence
[b] Fragments obtained for the different genotypes need to be checked by mapping or sequencing to determine allelism
[c] P446 described in Gilpin et al. (1997)
[d] Each marker was mapped by Loridon et al. (2005)
[e] RAPD markers were mapped by Laucou et al. (1998)
[f] SCAR marker mapped 4.6 cM from *Fw* by Okubara et al. (2005)

of the degree of admixture of the accessions, which can in turn be utilized as a matrix of co-factors in structured association programs. The average estimated log probability of the data Pr($x|k$), ideally should plateau at the most appropriate level of k. Values of k=1 to 10 are reported here and represent the average probability of 20 runs. The appropriate lengths of the program's burn-in (initiation) period and run time (actual number of simulations) were determined empirically to be 20,000 and 100,000, respectively. The default model of the program was utilized with the admixture option selected and correlated allele frequencies included between populations. In addition to the estimated log probability, the

Table 2. Description of the 49 qualitative and quantitative data of nutrients, disease/pest resistances, morphology/agronomy, phenology, and production traits used in generating marker-trait associations for 285 pea accessions of the USDA *Pisum* core collection.

Trait	GRIN code	Desciption	Reference and source [a]
Nutrients	Mineral_B	Total seed B concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_Ca	Total seed Ca concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_Cu	Total seed Cu concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_Fe	Total seed Fe concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_K	Total seed K concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_Mg	Total seed Mg concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_Mn	Total seed Mn concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_Mo	Total seed Mo concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_Ni	Total seed Ni concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_P	Total seed P concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Mineral_Zn	Total seed Zn concentration	Grusak et al. 2004, http://www.ars-grin.gov/cgi-bin/npgs/html/desclist.pl?177
	Protein	Total seed protein concentration	Coyne et al. 2005, Jermyn&Slinkard 1977, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177065
Disease/pest resistances	Ascochyta	Ascochyta blight	Kraft et al., 1998, 1991-1992†, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177028
	Fuswilt1	Fusarium Wilt Race1	McPhee et al. 1999, 1993-2006, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177022
	Fuswilt2	Fusarium Wilt Race2	McPhee et al. 1999, 1996-2001, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177032
	Fusrootrot	Fusarium root rot	Grünwald et al 2003; 1996-2002, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177026
	Rootrot	Aphanomyces root rot	Malvick & Percich 1999, 1990-1993, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177027
	Nematodelf	Pea cyst nematode	Tedford & Inglis 1999; http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177024
Morphology/ agronomy	Branching	Stem basal branching	McPhee & Muehlbauer, 1999; 2008-2010*, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177084
	Seedspod	Seeds per pod	Coyne et al, 2005; 1994-2010, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177012
	Sdcoatcol	Seed coat color	Coyne et al, 2005; 2002, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177014
	Seedsurf	Seed surface	Coyne et al, 2005; 2002, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177013
	Sdpospod	Seed position per pod	Coyne et al, 2005; 2003, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177064
	Sdpatcolor	Seed pattern color	Coyne et al, 2005; http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177031
	Planthgt	Plant height final	Coyne et al, 2005; 1994-2010, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177001
	Hilumcolor	Hilum color	Coyne et al, 2005; 2002, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177016
	Podwallneo	Pod wall neoplasia	Coyne et al. 2005, 1993-2010, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177062
	Flowpedunc	Flowers per peduncle	Coyne et al, 2005; 1992-2010, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177005

Table 2. (Continued).

Trait	GRIN code	Desciption	Reference and source [a]
	Flowercol	Flower color	Coyne et al, 2005; 1992-2010, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177004
	Cotylcolor	Cotyledon color	Coyne et al, 2005; 2002, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177015
	Internode	Average internode length	McPhee & Muehlbauer 1999; http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177044
	Stemdiam	Average Diameter	McPhee & Muehlbauer 1999, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177045
	Crushforce	Force to crush internode (Newton's)	McPhee & Muehlbauer 1999, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177046
	Shearingforce	Force to shear internode (Newton's)	McPhee & Muehlbauer 1999, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177047
	Rtbiomass	Root biomass (mg)	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	Rootshoot	Root shoot length (mm)	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	Taproot	Taproot length (mm) on 14th day	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	Shootdrywt	Root shoot dry weight (mg)	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	Rootdrywt	Root dry weight (mg)	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	R/s_ratio	Root shoot ratio	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	Rtlength	Total root length (cm)	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	Rtsurface	Root surface area (cm^{-2})	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	Rtdiameter	Average root diameter (mm)	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	Rtvolume	Total root volume (cm^{-2})	McPhee 2005, http://www.ars-grin.gov/cgi-bin/npgs/html/eval.pl?492924
	Taxon	Species records in GRIN	http://www.ars-grin.gov/cgi-bin/npgs/html/tax_search.pl
Phenology	Daysmature	Days to maturity	Coyne et al, 2005; 1996-2008, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177040
	Daysflower	Days to flower	Coyne et al, 2005; 1992-2008, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177034
Production	Biomass	Biomass (kg ha^{-1}) Dry Vegetation	Coyne et al, 2005; 1996-1997, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177036
	Seedwgt	100 seed weight (g)	Coyne et al, 2005; 1996-2003, http://www.ars-grin.gov/cgi-bin/npgs/html/desc.pl?177029

[a] The year the trait was investigated in the field or greenhouse, details in web site cited.

following ad hoc statistics suggested by Evanno et al. (2005) are also reported: the rate of change of the log probability of data with respect to the number of clusters [L'(k) = Pr($x|k$) - Pr($x|k$-1)]; the absolute value of the rate of change of the likelihood distribution [| L"k| = |L'(k+1) - L'(k)|]; and the absolute value of the rate of change divided by standard deviation of the 20 original simulations [$\Delta k = \mu$ | L"k|/s [L(k)]. These statistics have proved useful in interpreting the results from simulations where the log likelihood scores failed to reach a terminus or obvious plateau (Evanno et al., 2005).

The cluster analysis was constructed using the UPGMA (unweighted pair group method with arithmetic mean) method based on the allele-sharing distance by PowerMarker version 3.25 (Liu and Muse, 2005) and displayed using the software Mega4 (Tamura et al., 2007). The (polymorphic information content (PIC) was calculated using the equation PIC = $1-\sum P^2_i - \sum 2 P^2_i P^2_j$,where $\sum P^2_i$ is the sum of each squared ith haplotype frequency (Botstein et al.,1980).

The hypothesis of association of molecular markers with various phenotypic data in the presence of population structure was tested using the software program TASSEL 3.0.1 (Bradbury et al., 2007; Yu et al., 2005). First, a single factor analysis (SFA) of variance that did not consider population was performed using each marker as the independent variable and comparing the mean performance of each allelic class. This was done using the general linear model (GLM) function in TASSEL. In the next TASSEL analysis, Q GLM was used based on the chosen Q-matrix derived from STRUCTURE. The number of permutation run was set as 10,000 to obtain the permutation-based test of marker significance and the experiment-wise P value for marker significance. The Q+K MLM method used a kinship matrix and the population structure Q matrix. The K matrix was also based on the data for the 24 molecular markers and consisted of pairwise kinship coefficients for all pairs of lines in each population. The SPAGeDi software (Hardy and Vekemans, 2002) was used to calculate kinship coefficients. Linkage disequilibrium was calculated using TASSEL 3.0.1.

Table 3. Descriptive statistics for the 49 phenotypic traits used in the marker-trait association analysis.

Trait	GRIN code	Type	Scale	Range		Mean	SD	Variance
				Min	Max			
Nutrients	Mineral_B	Continuous	Concentration in the seed (ppm)	4.19	14.06	7.8	1.7	2.9
	Mineral_Ca	Continuous	Concentration in the seed (ppm)	311	2566	808.1	373.5	139645.1
	Mineral_Cu	Continuous	Concentration in the seed (ppm)	1.37	13.8	4.5	1.9	3.5
	Mineral_Fe	Continuous	Concentration in the seed (ppm)	23.16	105.2	50.4	12.6	164.1
	Mineral_K	Continuous	Concentration in the seed (ppm)	7126	20065	12495.4	1722.7	3358907
	Mineral_Mg	Continuous	Concentration in the seed (ppm)	1058	2473	1685.5	192.4	44589.1
	Mineral_Mn	Continuous	Concentration in the seed (ppm)	8.04	54.26	16.3	5.4	29.2
	Mineral_Mo	Continuous	Concentration in the seed (ppm)	5.87	56.47	23.2	8.2	67.8
	Mineral_Ni	Continuous	Concentration in the seed (ppm)	0.29	11.89	2.6	1.9	3.5
	Mineral_P	Continuous	Concentration in the seed (ppm)	2505	7655	5092.1	987.5	1030004
	Mineral_Zn	Continuous	Concentration in the seed (ppm)	16.1	106.63	41.9	12.1	150
	Protein	Continuous	Seed protein concentration (%)	13.2	30.93	22.1	3	10.2
Disease/ pest resistances	Fuswilt1	Ordinal	1 = resistant, 2 = mixed, 3 = susceptible	1	3	1.6	0.8	0.6
	Fuswilt2	Continuous	% of plant damage	0	100	44.4	32.8	1069.2
	Rootrot	Continuous	% of plant damage	36	100	83.2	10.3	127.7
	Nematodelf	Continuous	% of leaf damage	2	5	3.8	0.5	0.3
	Fusrootrot	Continuous	0.0 to 5.0	2	5	4.1	0.8	0.7
	Ascochyta	Continuous	% plant damage	25	100	57.1	19.1	368.9
Morphology/ agronomy	Branching	Ordinal	Number of basal branches	1	5	2.1	0.8	0.7
	Seedspod	Ordinal	Number of seed per pod	3	9	5.7	1.1	1.3
	Sdcoatcol	Ordinal	0 = non-pigmented: 1 = mixed, 2 = pigmented	0	2	0.4	0.6	0.4
	Seedsurf	Ordinal	1= round, 2 = mixed, 3 = wrinkled, 4 = other	1	3	1.5	0.7	0.5
	Sdpospod	Ordinal	Number of seed per pod	2	10	5.7	1.1	1.3
	Sdpatcolor	Ordinal	1 = black, 2 = brown, 3 = blue, 4 = grey, 5 = green, 6 = mixed	1	6	2.7	2.2	4.6
	Hilumcolor	Ordinal	1 = clear, 2 = mixed, 3 = pigmented	1	3	1.6	0.7	0.5
	Podwallneo	Ordinal	1 = yes, 2 = no	1	2	1.2	0.4	0.1
	Flowpedunc	Ordinal	Number of flowers per peduncle.	1	3	1.8	0.4	0.2
	Flowercol	Ordinal	1 = white, 2 = pigmented, 3 = mixed	1	3	1.9	1	0.9
	Cotylcolor	Ordinal	1 = yellow, 2 = mixed, 3 = green, 4 = other	1	3	1.2	0.6	0.4
	Rootshoot	Continuous	Root shoot length (mm)	62.0	380.0	241.0	83.0	6888.9
	Shootdrywt	Continuous	Root shoot dry weight (mg)	12.5	104.5	56.1	18.4	336.8
	Taproot	Continuous	Tap root length (mm)	181.0	432.3	332.8	40.0	1601.7
	Rtdiameter	Continuous	Root diameter (mm)	0.4	0.8	0.6	0.1	0.0
	Rtvolume	Continuous	Root volume area (cm^2)	0.1	0.9	0.4	0.2	0.0
	Rootdrywt	Continuous	Root dry weight (mg)	6.5	56.6	24.5	8.4	71.0
	Rtbiomass	Continuous	Root biomass (mg)	19.6	146.0	80.6	24.7	612.3
	R/s_ratio	Continuous	Root to shoot ratio	0.2	0.9	0.4	0.1	0.0
	Rtsurface	Continuous	Root surface area (cm^2)	9.4	62.2	30.7	10.6	111.6
	Rtlength	Continuous	Root length (cm)	53.7	355.8	175.5	58.6	3428.5
	Stemdiam	Continuous	Stem diameter (mm)	1.4	4.1	2.8	0.6	0.4
	Stemlength	Continuous	Average internode length (cm)	1.1	5.8	3.2	0.9	0.8
	Crushforce	Continuous	Force to crush internode - Newton's	3.5	36.8	15.6	4.9	24.0
	Shearingforce	Continuous	Force to shear internode - Newton's	8.6	64.9	28.0	10.7	115.0
	Taxon[a]	Ordinal	Identified by subspecies	1	5	1.2	0.9	0.7
	Planthgt	Continuous	Plant height (cm)	31	196	99.3	29.8	903.7
Phenology	Daysmature	Continuous	Number of days to maturity	99	125	114.9	6	304.6
	Daysflower	Continuous	Number of days to first flower	41	85	56.5	7.2	60.4
Production	Biomass	Continuous	Biomass (kg/ha)	336	7631	4459.4	1384.2	1945214
	Seedwgt	Continuous	100 seed weight (g)	2	41	17.1	6	36.4

[a] Subspecies followed by number of accessions: *P. sativum*: 1, *P. sativum* subsp. *sativum*: 2, *P. sativum* var. *arvense*: 3, *P. sativum* subsp. *abyssinicum*: 4, *P. sativum* subsp. *elatius*: 5

Results

Characterization of molecular markers

The number of microsatellite alleles detected ranged from two (PS11824 and PSU81287) to seven (PSBLOX13.2) (Table 1). A total of 64 alleles were detected, with a mean of 4.3 alleles per microsatellite. The average PIC was 0.17, ranging from 0.14 (P446a) to 0.29 (AF004843). Nine of the 15 microsatellites previously described (eight from Burstin et al., 2001; one from Gilipin et al., 1997) revealed additional alleles (1-3), whereas PSGAPA1 was reduced from six to three alleles in this study. On average, one additional allele was identified per microsatellite. RAPD primers chosen were previously demonstrated to produce multiple polymorphic bands and to represent mapped loci across the pea genome (Laucou et al., 1998). After excluding faint, difficult-to-score bands, the eight RAPD primers produced 37 robust polymorphic fragments. The average PIC of RAPD makers was 0.32, ranging from 0.26 (J12) to 0.37 (L13). The one SCAR marker, Y15_999*Fw* linked to *Fusarium wilt* race 1, produced absence or presence of amplicons with 0.33 of PIC value.

Population structure

Association analysis requires population structure to be taken into account in order to avoid false positive associations (Yu et al., 2005). An analysis of population structure and genetic distance confirmed significant population structure in this UPCC. The population genetic structure was analyzed using data from the 24 molecular markers (Figure 1). The log likelihood score did not plateau at a single value of "*K*"; instead, it continued to increase at relatively constant increments (Figure 2A). The second order statistics suggested by Evanno et al. (2005) (Figures 2B-D) supported the hypothesis of K=2. The three populations (group 1, group 2 and the admixed group) represented 37 (13.0%), 160 (56.1%), and 88 (30.9%) accessions of the UPCC based on the standard q-value of 85%, respectively. At hypothesized levels of K>2, the relative proportion of accessions assigned to each population became symmetric (1/*K*), an indication of a lack of valid population structure (Hubisz et al., 2009). Furthermore, at K>2 the populations did not correspond to any prior published studies, a criterion that is often used when examining inbred individuals (Stich et al., 2005; Hubisz et al., 2009).

Population group 1, the smallest (38 accessions), had the estimated ancestry membership (*q*) of greater than 0.85 (blue bar). This group includes ten *Pisum sativum* ssp *elatius*, one *P. sativum* ssp *abyssinicum*, one *P. sativum* var *arvense* and those accessions from the Asian highlands. The admixed group, which had *q* values of less than 0.85, contained 87 (30.5%) accessions of the *Pisum* core. Population group 2 constitutes approximately half of the core collection (56.1%) and is the predominant group.

Genetic diversity and cluster analysis

The genetic distance matrix generated from the combined data set ranged from a low of 0.0280 between accessions PI 356973 and PI 356974 (two accessions from India collected in the same year at approximately the same location) to a high of 0.5147 between PI 273207 (*P. sativum* ssp *sativum*) and PI 125839 (an adapted Afghanistan cultivar). The cluster analysis was constructed by using the UPGMA method based on the allele-sharing distance (Fig. 1). At the highest level of divergence, six clusters can be distinguished. Clade-a contained 20 accessions of Asian origin, 14 accessions of European origin, and one accession of Sudanese origin. Interestingly, clade-a was grouped into group 1 in the population structure analysis (Figure 1). This clade was divided into sub-clades (a-1, a-2). Clade-a-1 includes nine accessions of *P. sativum* ssp *elatius* and one of *P. sativum*, while clade-a-2 includes 20 accessions from Afghanistan, India, Pakistan, China, Nepal, Iran and the United Kingdom. Excluding two accessions from Iran and the United Kingdom, the origins of these accessions in clade-2-a have been described as potential areas of primary or secondary diversity. The tendency of accessions from these highland areas to cluster together has been noted previously (Ellis et al., 1998; Pearce et al., 2000). Clades b, c, d, e and f included 48, 60, 66, 62 and 12 UPCC, respectively. Additionally, each accession in the cluster analysis was displayed using three different colors (blue, green and red) according to population structure. As mentioned above, all clade-a accessions exactly matched population group 1. The q-value of the 20 accessions in clade-a-2 is higher than 0.95 (blue bars). Clade-f included nine accessions from the admixed group and three from group 2; clade-b consisted of 36 accessions from the admixed group and 12 from group 2; 25% of both clades was comprised of accessions from group 2. 65% of clade-c and 69.4% of clade-e were comprised of accessions from group 2 as well. 98.5% of clade-d consisted of accessions from group 2. All accessions in clade-d had a q-value of greater than 0.95 (red bars) in the population structure analysis. This genetically divided population is in strong agreement with clade-a-2, which could be predictive of another geographical origin.

Association mapping

Associations between 24 genotypic (15 SSR, eight RAPD and one SCAR marker) and 49 phenotypic traits (Table 3) of UPCC were determined next by (1) single factor analysis (SFA); (2) structured association analysis using a general linear model where population membership served as covariates (Q GLM); and (3) a composite approach where the average relationship is estimated by kinship (K) and implemented in a mixed linear model (Q+K MLM) method. Table 4 presents the average significance levels for $P < 0.01$ for all markers for each of the analyses. Using SFA, we observed that 21 markers (37 alleles) were significantly associated with pheno-

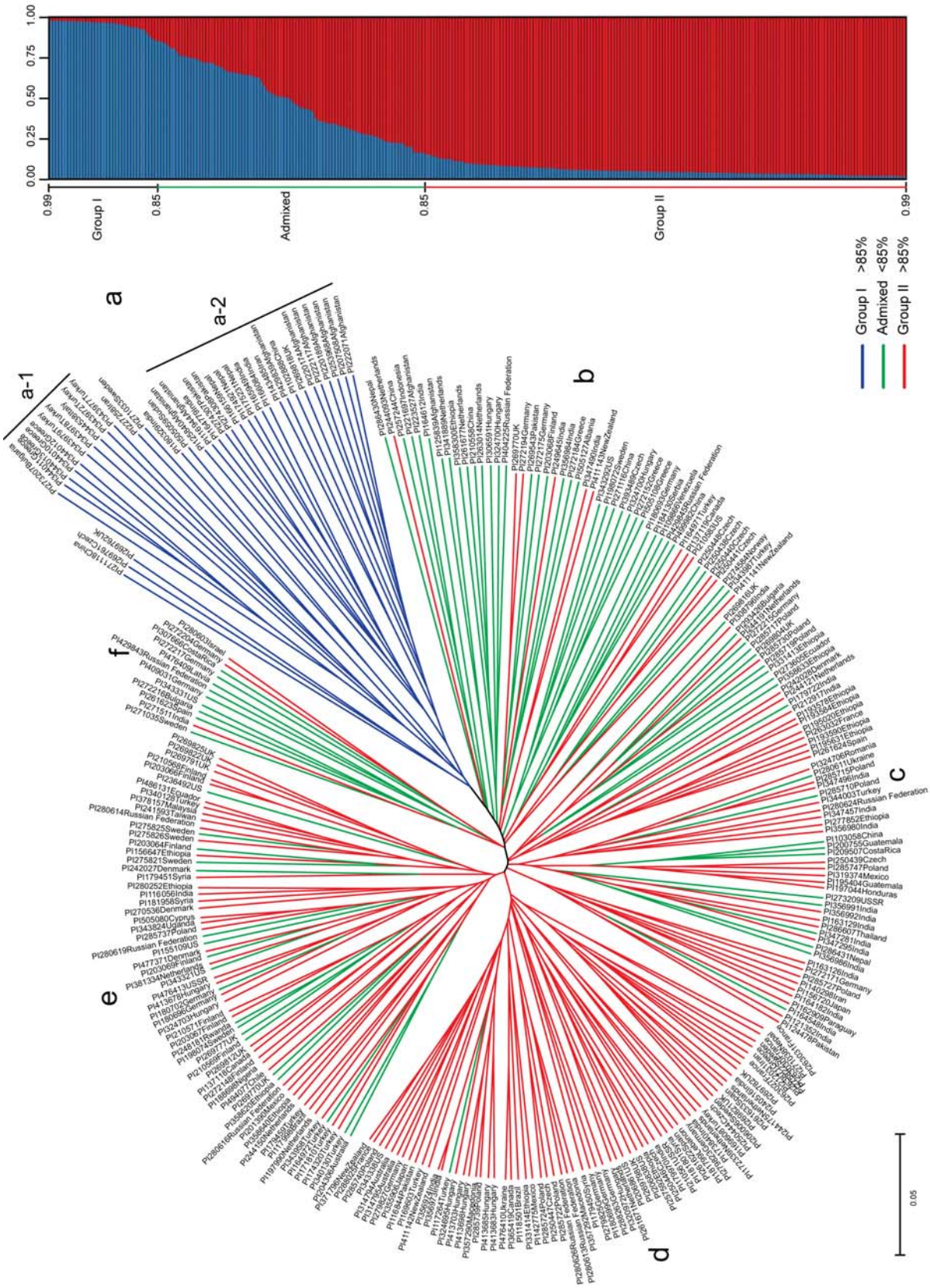

Figure 1. Cluster analysis and population structure generated from 285 accessions of the USDA pea (*Pisum*) core collection using 24 molecular markers.

typic traits. A total 72 significant marker-trait associations (SMTA) (P <0.01) were detected using SFA (Table 4). R13 was associated with 20 traits, while UBC498, PSAS, PEAOM14A, Y15_999*Fw*, UBC467, PSGAP1, PSU81287, AF016458, Y14, PSU81288 and V17 were associated with one trait each. The lowest P-value detected using SFA was AA430902B (P = 4.99E-06) associated with taxon trait, while the highest P-value was J12_1100 (P = 0.0093) associated with seed Fe concentration trait. In Q GLM, 16 markers (29 alleles) were observed SMTA from all of phenotypic traits. P446 was associated with 15 traits, and PEAOM14A, PSGAP1, PSU81287, R12, UBC498, V17 and Y14 were associated with one trait each. The lowest P-value of SMTA occurred in AF004843_B (P = 7.29E-05) associated with total seed P concentration trait, while the highest P-value of SMTA occurred in UBC467_600 (P = 0.0098) associated with plant height trait. According to the Q+K MLM method, using a Q and K matrix, twelve markers (20 alleles) were observed to be SMTA in 29 phenotypic traits. R13 contained 16 SMTAs, while PSU81288, AF016458, PSU81287 and PSU51918 each contained one SMTA. The lowest P-value of SMTA was observed in AA430902B (P=8.00E-05) associated with taxon trait, whereas the highest P-value of SMTA was observed in P446A (P=0.0096) associated with seed Ni concentration trait. The two model approaches (GLM and MLM) were compared for all traits, as the significance criteria there were different for each model. In our study, for all molecular alleles, 29 and 20 SMTAs were defined using GLM and MLM methods, respectively. Therefore, using the MLM method, the number of SMTAs detected decreased by 31%. 17 identical SMTAs were shared by the two models and fulfilled the significance criteria in both (P<0.01). In the overall average using these three models (SFA, GLM and MLM), P446 was observed as having the most (four) SMTAs, while seven markers (UBC498, PEAOM14A, PSGAP1, R12, PSU81287, Y14 and V17) contained one SMTA. The lowest P-value was observed in AA430902B (P = 8.15E-05) associated with taxon trait, while the highest P-value was observed in P446A (P = 0.0095) associated with seed Ni concentration trait in the overall average (Table 4). Ultimately, the three procedures revealed 28 SMTAs pertaining to kinship and/or population structure for this core collection (Table 4). Interestingly, SMTAs were identified for eight of the seed mineral nutrient concentrations, including Ca, Cu, K, Mo, Ni and P. Three of the disease and pest resistances, including Fusarium wilt race 1, Aphanomyces root rot and resistance to the pea cyst nematode, had significant SMTAs.

Discussion

The USDA *Pisum* core collection has been assembled to represent a broad subset of available genetic diversity. The range of the genetic distance scores (0.0280 to 0.5147) suggests that redundancy has for the most part been successfully eliminated from the core collection during the course of its various refinements. The results also suggest that a multifaceted population genetic structure is present in the UPCC, influenced

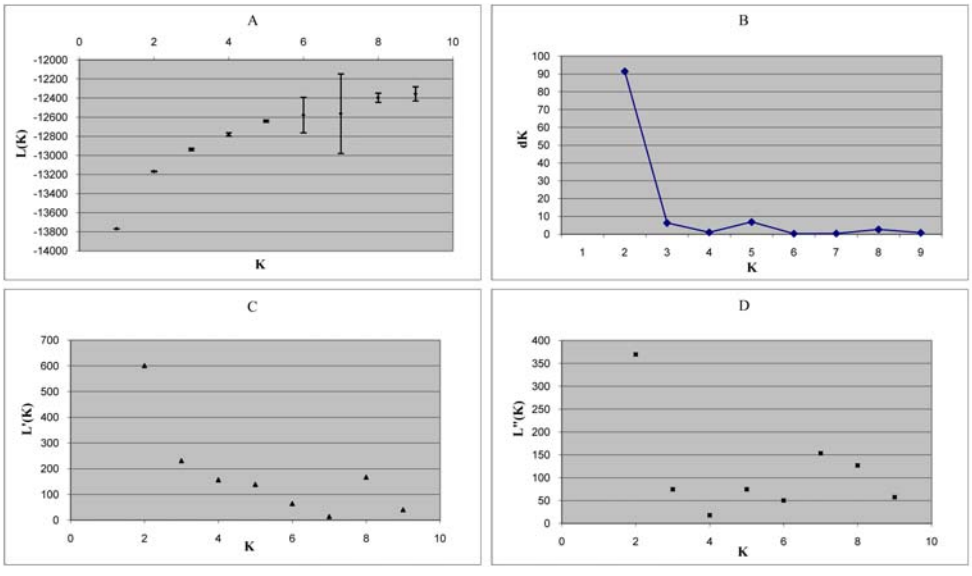

Figure 2. Results of the population genetic sub-structure analysis generated from 15 microsatellite markers and 285 accessions of USDA *Pisum* core collection. A) The average estimated log probability of the data Pr(x | k) for k=1 to 10. B). The rate of change of the log probability of data with respect to the number of clusters [L'(k) = Pr(x | k) - Pr(x | k-1)]. C) The absolute value of the rate of change of the likelihood distribution [| L"k| = |L'(k+1) - L'(k)|]. D) The absolute value of the rate of change divided by standard deviation of the 20 original simulations. [$\Delta k = \mu$ | L"k|/ s [L(k)].

Table 4. Significance of tests for association analysis between 24 molecular makers and 49 phenotypic traits using three statistical approaches. Traits /markers in bold are significant at $p < 0.01$ for all three tests.

Trait	GRIN code	Marker	SFA[a]	R^2	Q GLM[b]	R^2	Q+K MLM[c]	R^2	Overall average†
Nutrients	**Ca**	**PS11824B**	**	0.031	**	0.031	**	0.031	**
	Cu	**P446A**	**	0.029	**	0.027	**	0.026	**
	Fe	J12_1100	**	0.027	**	0.031	*	0.023	*
	K	**UBC498_600**	**	0.031	**	0.030	*	0.025	**
	Mo	**R13_850**	***	0.051	*	0.020	**	0.036	**
		AF004843A	**	0.034	***	0.047	**	0.026	**
	Ni	**P446A**	**	0.035	**	0.033	*	0.022	**
		PSAS_B	**	0.025	*	0.021	*	0.022	*
		PEAOM14AA	**	0.033	**	0.032	*	0.016	*
	P	**PSU51918C**	**	0.028	**	0.026	**	0.034	**
		AF004843B	***	0.061	***	0.055	**	0.028	**
	Protein	AF004843B	**	0.028	**	0.030	ns	0.012	*
Disease/pest resistances	**Fuswilt1**	**PSU51918E**	**	0.039	**	0.034	*	0.021	**
		UBC467_600	*	0.025	*	0.029	**	0.034	**
	Rootrot	**PSARGDECA_F**	**	0.034	**	0.041	**	0.029	**
	Nematode(Leaf)	**PSGAP1_A**	**	0.245	**	0.239	*	0.170	**
		PS11824B	***	0.190	***	0.191	**	0.206	***
		J12_1300	**	0.128	**	0.123	ns	0.064	*
	Fusrootrot	AA430902_C	**	0.026	ns	0.012	*	0.015	*
		PSARGDECA_A	**	0.028	ns	0.007	*	0.022	ns
	Ascochyta	R13_850	***	0.064	*	0.018	***	0.026	*
		AF004843B	ns	0.012	**	0.032	ns	0.004	ns
		PEACHLHPPS_B	**	0.038	*	0.019	**	0.033	*
Morphology/ agronomy	**Branching**	**R13_300**	**	0.036	*	0.022	**	0.027	**
	Seedsurf	R12_800	*	0.019	**	0.023	ns	0.012	*
		AF004843B	**	0.032	*	0.014	ns	0.013	*
	Sdpospod	J12_950	**	0.032	**	0.034	*	0.021	*
		PSU81287_B	**	0.028	**	0.030	**	0.029	**
	Sdpatcolor	R13_300	**	0.076	ns	0.010	*	0.050	ns
		AA430902_C	**	0.073	ns	0.007	ns	0.041	ns
	Hilumcolor	R12_1100	***	0.054	*	0.014	*	0.024	*
		AF016458_B	*	0.023	**	0.030	*	0.021	*
	Podwallneo	AA430902_B	*	0.020	ns	0.013	**	0.027	*
	Flowpedunc	PSU81288_C	*	0.020	**	0.025	*	0.021	*
	Flowercol	**PSARGDECA_E**	**	0.038	**	0.023	**	0.031	**
		R12_1100	***	0.043	ns	0.005	*	0.017	ns
		R13_850	**	0.032	ns	0.008	*	0.016	*
		J12_850	**	0.036	ns	0.005	*	0.022	ns
		PEACHLHPPS_D	***	0.051	ns	0.005	**	0.032	ns
		Y15_999Fw	**	0.036	ns	0.009	ns	0.013	ns
	Cotylcolor	**J12_850**	***	0.046	**	0.027	*	0.025	**
		P446B	**	0.033	**	0.029	**	0.040	**
		PEACHLHPPS_A	**	0.025	*	0.016	ns	0.002	ns
		AF004843B	**	0.030	*	0.021	ns	0.007	ns
	Rootshoot	**P446C**	**	0.033	**	0.044	**	0.034	**
		R13_850	***	0.049	*	0.020	**	0.038	*
	Shootdrywt	P446C	*	0.023	**	0.036	*	0.023	*
		R13_850	***	0.054	ns	0.011	**	0.041	*
	Taproot	P446C	*	0.023	**	0.035	*	0.025	*
		R13_850	***	0.058	*	0.017	**	0.044	*
	Rtdiameter	P446C	*	0.025	**	0.037	*	0.027	*
		R13_850	***	0.052	ns	0.012	***	0.050	***
	Rtvolume	P446C	*	0.025	**	0.037	*	0.027	*
		R13_850	***	0.052	ns	0.012	***	0.049	*

Table 4. (Continued).

Trait	GRIN code	Marker	SFA[a]	R^2	Q GLM[b]	R^2	Q+K MLM[c]	R^2	Overall average†
	Rootdrywt	P446C	*	0.024	**	0.037	*	0.023	*
		R13_850	***	0.054	ns	0.012	**	0.041	*
	Rtbiomass	PEACHLHPPS_D	**	0.046	ns	0.005	*	0.028	ns
		PSARGDECA_A	**	0.044	ns	0.001	*	0.030	ns
		R13_850	**	0.045	ns	0.001	*	0.034	ns
	R/Sratio	P446C	*	0.025	**	0.037	*	0.027	*
		R13_850	***	0.052	ns	0.012	***	0.049	*
	Rtsurfacea	P446C	*	0.024	**	0.007	*	0.023	*
		R13_850	***	0.054	ns	0.012	**	0.042	*
	Rtlength	P446C	*	0.024	**	0.038	*	0.023	*
		R13_850	***	0.063	ns	0.012	**	0.047	*
	Average diameter(stem)	**AF016458_F**	**	0.030	**	0.032	**	0.030	**
		P446C	*	0.023	**	0.030	*	0.023	*
	Average length(stem)	R13_850	**	0.027	ns	0.007	**	0.027	ns
	Crushing force(stem)	P446C	*	0.017	**	0.027	*	0.020	*
		R13_850	***	0.051	ns	0.008	**	0.029	*
		PS11824A	**	0.026	*	0.020	*	0.022	*
	Shearing force(stem)	R13_850	***	0.041	ns	0.011	**	0.033	*
	taxon	**AA430902_B**	***	0.071	***	0.042	***	0.056	***
		R13_850	***	0.069	ns	0.008	**	0.029	*
		R12_1100	***	0.044	ns	0.011	*	0.020	*
		AF004843B	*	0.017	**	0.032	ns	0.013	*
	Planthgt	UBC467_600	**	0.025	**	0.024	ns	0.014	*
		AF004843B	**	0.031	**	0.029	*	0.019	*
Phenology	**Daysmature**	**AA430902_E**	**	0.225	**	0.227	**	0.225	**
	Daysflower	**Y14_2000**	**	0.029	**	0.036	*	0.024	**
		R12_1100	**	0.028	*	0.016	*	0.025	*
		P446C	*	0.024	**	0.032	*	0.020	*
Production	**Biomass**	**PEACHLHPPS_B**	**	0.035	*	0.021	**	0.029	**
		AA430902_F	**	0.033	**	0.033	**	0.031	**
		UBC467_600	*	0.026	**	0.028	**	0.029	**
		PSU81288_F	**	0.033	**	0.026	**	0.030	**
		AF004843B	*	0.021	***	0.043	ns	0.012	*
		J12_950	**	0.034	*	0.022	*	0.022	*
		PSARGDECA_A	**	0.031	ns	0.007	*	0.025	ns
		R13_850	**	0.037	ns	0.006	*	0.021	ns
	Seedwgt	**V17_900**	**	0.036	**	0.019	*	0.026	**
		PEACHLHPPS_D	**	0.037	ns	0.001	*	0.018	ns

[a] SFA: single factor analysis of variance

[b] Q GLM: general linear model using the Q population structure matrix

[c] Q+K MLM: mixed linear model using the Q population structure matrix and the K kinship matrix.

* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

† Overall: average p-value of 3 methods.

on one level by the inclusion of unadapted sub-species and on another by what appears to be specific sets of breeding considerations.

Our study describes the population structure and genetic diversity of the USDA *Pisum* core collection. This approach has provided insight that would not have been obtained by any single technique. The analysis of the population genetic structure using 15 microsatellites, eight RAPDs and one SCAR marker identified three populations within the UPCC. Population group 1, 13.0% of the core collection, consisted of the wild *P. sativum* ssp. *elatius* and *abyssinicum* and the accessions from the Asian highland. Population group 2 and the admixed group constituted 56.1% and 30.9% of the *Pisum* core, respectively. A similar result was described by Jing et al. (2010), in which 3,029 *Pisum* germplasm samples from the John Innes *Pisum* germplasm (JIPG) collection were genotyped using 45 retrotransposon-based insertion polymorphism (RBIP) markers, and the population structure revealed three major groups corresponding approximately to landrace, cultivar and wild *Pisum*. In addition, Jing et al. (2010) observed

that several possible values of K were resolved into 3, 7 and 11 sub-groups based on the ΔK and Ln P(X|K), many of which correlate with taxonomic, domestication-related phenotypic traits and geographical data. In our study, the similar pattern also emerged when K=3 and 5. This indicated that the UPCC could be divided into three and/or five sub-groups of genotypes. These sub-groups were primarily linked to taxonomic, domestication-related phenotypic traits and geographical data. The differences in sub-ΔK between JIPG and our collection were caused by the diversity of the collections used and differences in genotyping methods.

The cluster analysis using a combined dataset comprised of microsatellites and RAPDs agreed with the results of the population structure, identifying six clades that corresponded to the bulk of the core collection. Twelve accessions of *Pisum* sub-species of European origin were clustered together in clade-a-1. According to Baranger et al. (2004), all *Pisum* sub-species (*P. sativum* ssp. *abyssinicum* and *P. sativum* ssp. *elatius*) were clustered together with *P. sativum* accessions that originated from Afghanistan. In addition, Jing et al. (2010) documented that *P. sativum* ssp. *elatius* and *P. sativum* ssp. *abyssinicum* were located in same branch in a distance-based estimation with *P. fulvum.* Ellis et al. (1998) suggested that there were three recognizable main groups of *Pisum*: *P. sativum* ssp. *abyssinicum, P. fulvum* and the bulk of the *Pisum* germplasm. While noting that *P. sativum* ssp. *elatius* was not as distinct as *P. sativum* ssp. *abyssinicum*, Ellis et al. (1998) did suggest that *P. sativum* ssp. *elatius* appeared to have some sub-group structure. Our results, however, suggest that while there appears to be some integration between *P. sativum* ssp. *elatius* and the main body of *Pisum* germplasm, there also appears to be a core of *P. sativum* ssp. *elatius* accessions that are at least as distinct as *P. sativum* ssp. *abyssinicum*. Although our genotype analysis used a limited number of markers to cover the entire *Pisum* genome, the average genetic distance among *P. sativum* ssp. *elatius* was 0.177; on the other hand, the average genetic distance between *P. sativum* ssp. *elatius* and *P. sativum* ssp. *abyssinicum* was 0.37.

Clade-a-2, which consists primarily of accessions from the Asian highlands, corresponds to 18 accessions of *P. sativum* from Afghanistan, India, Pakistan, China and Nepal. These accessions have similar phenotypes associated with developed germplasm (i.e. flower color; smooth seed surface with distinct markings, such as purple speckles and/or brown marbling; and relatively low seed weights). Furthermore, the accessions from the Asian highlands appear to be as distinct as either of the sub-species (*P. sativum* ssp. *elatius* and *P. sativum* ssp. *abyssinicum*). The population genetic structure analysis suggests that this cluster is genetically closer to the wild sub-species than to the main body of the core collection. The geographic origin of the accessions included in this cluster roughly corresponds to a putative center of diversity of *Pisum*. As noted previously, these accessions share a number of traits. The distinctiveness of these accessions could be due to the geographic isolation of the region of origin or to selection required to meet unique features associated with this region. Regardless, the cluster appears to be unique and forms a well-explored genotypically evolutionary branch of *Pisum* (Smýkal et al., 2011).

An important feature of population structure analysis is that lines can be divided genetically pure and/or admixed line in a given plant population. The combination of population structure and cluster analysis methods provided an effective means of examining the gene flow and history of the germplasm collection. LD is the genetic phenomenon of nonrandom association of alleles at different loci (Flint-Garcia et al., 2003). The nonrandom association has been observed not only between alleles of loci on different chromosomes, but also between alleles of loci on the same chromosome (Hagenblad and Nordborg, 2002; Stich et al., 2005; Tenaillon et al., 2001). Allele frequency and recombination between sites, as well as the effective population size, are important factors in LD (Weir, 1996). According to Loridon et al. (2005), PSU81288, PSU51918 and AF016458 were located in the same linkage group (LG I) at 77.0, 94.1 and 145.6 cM, respectively. In our LD analysis, r^2 between PSU81288 and PSU51918 (17.1 cM) was 0.301, between PSU51918 and AF016458 (51.5 cM) was 0.119, and between PSU81288 and AF016458 (68.6 cM) at 0.047. Although extensive LDs in self-pollinated species such as rice, barley and *Arabidopsis*, intrachromosomal LDs of up to 50 cM with r^2>0.05 are rarely reported and many studies have indicated the LD decay of 1 cM or less in self-pollinated species (Malysheva-Otto et al., 2006; Neumann et al., 2011; Nordborg et al., 2002; WeiGuo et al., 2009). A recent LD analysis using more than 3 million SNPs in the model legume *Medicago truncatula*, r^2 between each pair of SNPs fell to 50% within the initial 3kb and to less than 0.03 within 5kb (Branca et al., 2011). Moreover, when Jing et al. (2007) examined LD decay with 39 dispersed loci using sequence diversity in *Pisum* cultivars, extensive LDs were also observed. Although we analyzed LDs using limited markers and loci, the alleles showing high association may shed some light on germplasm management and subsequent breeding programs utilizing the USDA *Pisum* core collection.

Single factor analysis of variation, a traditional QTL statistical method, identified 37 loci associated with eight seed nutrient concentrations; five disease/pest resistances; and 27 morphological traits. The Q GLM model that utilized population structure identified 48 SMTAs. These SMTAs were reduced to 38 when population structure and kinship were included in the MLM model. This reduction in significant associations using the Q+K MLM method is generally consistent with results in maize (Yu et al., 2005). Although limited genotype data were used in this study, our population genetic structure and cluster analysis agreed with many previous studies involving *Pisum* germplasm. The most important facet of our research was that it constitutes the first attempt to apply associ-

ation analysis to the management of *Pisum* germplasm. These results revealed the significant marker/trait associations between 24 molecular markers and 49 phenotype data within the USDA *Pisum* core collection. Most notable were the associations identified for eight of the seed mineral nutrients considered in this study. Disease and pest resistance SMTAs were identified for Fusarium wilt race 1, Aphanomyces root rot and the pea cyst nematode. Future research efforts will focus on locating positive alleles for other traits linked to improved field performance, specifically phonological (days to flower and days to maturity) and production-related (seed weight and biomass) traits.

Marker-trait studies in pea germplasm could provide a useful alternative to linkage mapping in the detection of marker-phenotype associations to be used in the implementation of marker-assisted selection and, eventually, in genomic selection for pea crop improvement. More research is needed to bring various pea germplasm populations, collections, cores, mini-cores, reference sets, etc., together for efficient utilization in crop improvement research (Glaszmann et al., 2010; Smýkal et al., 2011). Recent publications concerning a pea SNP set (Deulvot et al., 2010), the *in silico* placement of 5,460 unigenes on the pea linkage map (Bordat et al., 2011), pea transcriptomes (Franssen et al., 2011) and the formation of appropriate association mapping populations of now well-described pea germplasm are important steps toward fully utilizing the genetic diversity within this valuable germplasm in crop improvement.

Acknowledgements The authors wish to acknowledge their funding sources: USDA-ARS post-doctoral fellowships (S-JK and AFB); USDA-ARS Pacific West Summer Internship (CW); USDA Foreign Agricultural Service Scientific Cooperation Exchange Program (GTV, CJC) and funding from CRIS Projects #5348-21000-017-00D (JH, SJK, TK, CJC) and #5348-21000-024-00D (RJM). For technical assistance, the authors wish to thank Landon Charlo and Michael Cashman for their morphological data collection, as well as Leon Razai and Lisa Taylor for their laboratory assistance.

References

Abdurakhmonov IY and Abdukarimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int. J. Plant Genomics 574927.

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

Botstein D, White RL, Skolnick M, and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Amer. J. Human Genet. 32: 314-331.

Bordat A, Savois V, Nicolas M, Salse J, Chauveau A, Bourgeois M, Potier J, Houtin H, Rond C, Murat F, Marget P, Aubert G and Burstin J (2011) Translational genomics in legumes allowed placing *in silico* 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. Genes Genom. Genet. 1: 93-103.

Branca A, Paape TD, Zhou P, Briskine R, Farmer AD, Mudge J, Bharti AK, Woodward JE, May GD, et al. (2011) Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. Proc. Natl. Acad. Sci. USA 108: E864-E870.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, and Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Bretting PK and Widrlechner MP (1995) Genetic markers and plant genetic resource management. In Plant Breeding Reviews, J. Janick, ed., John Wiley & Sons, New York, NY, pp. 11-86.

Brown AH (1989) Core collections: a practical approach to genetic resources management. Genome 31: 818-824.

Brown AH and Spillane C (1999) Implementing core collections - principles, procedures, progress, problems and promise. In Core collections for today and tomorrow, R.C. Johnson and T. Hodgkin, eds., Crop Science Society of America, Madison, WI, pp. 1-10.

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

Charcosset A and Moreau L (2004) Use of molecular markers for the development of new cultivars and the evaluation of genetic diversity. Euphytica 137: 81-94.

Chavarriaga-Aguirre P, Maya MM, Tohme J, Duque MC, Iglesias C, Bonierbale MW, Kresovich S, and Kochert G (1999) Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collection and to assess the usefulness of DNA-based markers to maintain germplasm collections. Mol. Breed. 5: 263-273.

Choudhury PR, Tanveer H, and Dixit GP (2007) Identification and detection of genetic relatedness among important varieties of pea (*Pisum sativum* L.) grown in India. Genetica 130: 183-191.

Coyne CJ, Brown AF, Timmerman-Vaughn GM, McPhee K, and Grusak MA (2005a) USDA-ARS refined pea core collection for 26 qualitative traits. Pisum Genet. 37: 3-6.

Coyne CJ, Grusak MA, Razai L, and Baik BK (2005b) Variation for pea seed protein concentration in the USDA Pisum core collection. Pisum Genet. 37: 5-9.

Deulvot C, Charrel H, Amandine M, Jacquin F, Donnadieu C, Lejeune-Hénaut I, Burstin J, and Aubert G (2010) High-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. BMC Genomics 11: 468.

Dribnokhodova OP and Gostimsky SA (2009) Allele Polymorphism of Microsatellite Loci in Pea *Pisum sativum* L. Lines, Varieties, and Mutants. Rus. J. Genet. 45: 788-793.

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, and Ambrose MJ (1998) Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. Mol. Gen. Genet. 260: 9-19.

Esposito MA, Martin EA, Cravero VP, and Cointry E (2007) Characterization of pea accessions by SRAP's markers. Scientia Hort. 113: 329-335.

Evanno G, Regnaut S, and Goudet J (2005) Detecting the number

of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611-2620.

Falush D, Stephens M, and Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Falush D, Stephens M, and Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes 7: 574-578.

Flint-Garcia SA, Thornsberry JM, and Buckler ES (2003) Structure of linkage disequilibrium in plants. Ann. Rev. Plant Biol. 54: 357-374.

Food and Agricultural Organization of the United Nations (2011) http://www.fao.org/

Ford R, Le Roux K, Itman C, Brouwer JB, and Taylor PWJ (2002) Diversity analysis and genotyping in *Pisum* with sequence tagged microsatellite site (STMS) primers. Euphytica 124: 397-405.

Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, and Weber APM (2011) Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing BMC Genomics 12: 227.

Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, and Tanksley SD (2000) *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science 289: 85-88.

Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, and Schüler K (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. Mol. Breed. 13: 93-102.

Ghafoor A, Ahmad Z, and Anwar R (2005) Genetic diversity in *Pisum sativum* and a strategy for indigenous biodiversity conservation. Pak. J. Bot. 37: 71-77.

Gilpin BJ, McCallum JA, Frew TJ, and Timmerman-Vaughan GM (1997) A linkage map of the pea (*Pisum sativum* L.) genome containing cloned sequences of known function and expressed sequence tags (ESTs). Theor. Appl. Genet. 95: 1289-1299.

Glaszmann JC, Kilian B, HD Upadhyaya HD, and Varshney RK (2010) Accessing genetic diversity for crop improvement. Curr. Opin. Plant Biol. 13: 167-173.

Grünwald NJ, Coffman VA, and Kraft JM (2003) Sources of partial resistance to Fusarium root rot in the *Pisum* core collection. Plant Disease 87: 1197-1200.

Grusak MA, Burgett CL, Knewtson SJB, Lopéz-Millán A-F, Ellis DR, Li C-M, Musetti VM, and Blair MW (2004) Novel approaches to improve legume seed mineral nutrition. Proceedings of the 5th AEP-2nd ICLGG Conference: pp. 37-38.

Hagenblad J and Nordborg M (2002) Sequence variation and haplotype structure surrounding the flowering time locus *FRI* in *Arabidopsis thaliana*. Genetics 161: 289-298.

Hardy OJ and Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2:618-620.

Hoey BK, Crowe KR, Jones VM, and Polans NO (1996) A phylogenetic analysis of *Pisum* based on morphological characters, and allozyme and RAPD markers. Theor. Appl. Genet. 92: 92-100.

Hubisz MJ, Falush D, Stephens M, and Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Res. 9: 1322-1332.

Jermyn WA and Slinkard AE (1977) Variability of percent protein and its relationship to seed yield and seed shape in peas. Legume Res. 1: 33-37.

Jing R, Johnson R, Seres A, Kiss G, Ambrose MJ, Knox MR, Ellis TH, and Flavell AJ (2007) Gene-based sequence diversity analysis of field pea (*Pisum*). Genetics 177: 2263-2275.

Jing R, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis TH, and Flavell AJ (2005) Insertional polymorphism and antiquity of *PDR1* retrotransposon insertions in *Pisum* species. Genetics 171: 741-752.

Jing R, Vershinin A, Grzebyta J, Shaw P, Smykal P, Marshall D, Ambrose MJ, Ellis TH, and Flavell AJ (2010) The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. BMC Evol. Biol. 10: 44.

Kosterin OE, Zaytseva OO, Bogdanova VS, and Ambrose MJ (2010) New data on three molecular markers from different cellular genomes in Mediterranean accessions reveal new insights into phylogeography of *Pisum sativum* L. subsp. *elatius* (Bieb.) Schmalh. Genet. Res. Crop Evol. 57: 733-739.

Kraft JM, Dunne B, Goulden D, and Armstrong S (1998) A search for resistance in peas to *Mycosphaerella pinodes*. Plant Disease 82: 251-253.

Laucou V, Haurogn K, Ellis N, Rameau C (1998) Genetic mapping in pea. 1. RAPD-based genetic linkage map of *Pisum sativum*. Theor. Appl. Genet. 97: 905-915.

Lazaro A and Aguinagalde I (2006) Genetic variation among Spanish pea landraces revealed by inter simple sequence repeat (ISSR) markers: its application to establish a core collection. J. Agri. Sci. 144: 53-61.

Le Clerc V, Cadot V, Canadas M, Lallemand J, Guerin D, and Boulineau F (2006) Indicators to assess temporal genetic diversity in the French Catalogue: no losses for maize and peas. Theor. Appl. Genet. 113: 1197-1209.

Lee D, Ellis THN, Turner L, Hellens RP, and Cleary WG (1990) A *copia*-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic analysis. Plant Mol. Biol. 15: 707-722.

Liu K and Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128-2129.

Loridon K, McPhee K, Morin J, Dubreuil P, Pilet-Nayel ML, Aubert G, Rameau C, Baranger A, Coyne C, and Lejeune-Hénaut I (2005) Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). Theor. Appl. Genet. 111: 1022-1031.

Lu J, Knox MR, Ambrose MJ, Brown JKM, and Ellis THN (1996) Comparative analysis of genetic diversity in pea assessed by RFLP-and PCR-based methods. Theor. Appl. Genet. 93: 1103-1111.

Malvick DK and Percich JA (1999) Identification of *Pisum sativum* germ plasm with resistance to root rot caused by multiple strains of *Aphanomyces euteiches*. Plant Disease 83: 51-54.

Malysheva-Otto LV, Ganal MW, and Roder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). BMC Genetics 7: 6.

Martín-Sanz A, Caminero C, Jing R, Flavell AJ, and Pérez de la Vega M (2011) Genetic diversity among Spanish pea (*Pisum sativum* L.) landraces, pea cultivars and the world *Pisum* sp. core collection assessed by retrotransposon-based insertion polymophisms (RBIPs). Span. J. Agri. Res. 9: 166-78.

Mazzucato A, Papa R, Bitocchi E, Mosconi P, Nanni L, Negri V, Picarella ME, Siligato F, Soressi GP, Tiranti B, and Veronesi

F (2008) Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. Theor. Appl. Genet. 116: 657-669.

McPhee K (2005) Variation for seedling root architecture in the core collection of pea germplasm. Crop Sci. 45: 1758-1763.

McPhee KE and Muehlbauer FJ (1999) Stem strength in the core collection of *Pisum* germplasm. Pisum Genet. 31: 21-23.

McPhee KE and Muehlbauer FJ (2001) Biomass production and related characters in the core collection of Pisum germplasm. Genet. Res. Crop Evol. 48: 195-203.

McPhee KE, Tullu A, Kraft JM, and Muehlbauer FJ (1999) Resistance to Fusarium wilt race 2 in the *Pisum* core collection. J. Amer. Soc. Hort. Sci. 124: 28-31.

Murray MG and Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucl. Acids Res. 8: 4321-4325.

Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, and Buckler ES (2009) Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21: 2194-2202.

Neumann K, Kobiljski B, Denčić S, Varshney RK, and Börner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). Mol. Breed. 27: 37-58.

Nisar M, Ghafoor A, Khan MR, and Subhan M (2009) Genetic similarity of Pakistan pea (*Pisum sativum* L.) germplasm with world collection using cluster analysis and Jaccard's similarity index. J. Chem. Soc. Pak. 31: 138-144.

Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, and Oefner PJ (2002) The extent of linkage disequilibrium in Arabidopsis thaliana. Nat. Genet. 30: 190-193.

Okubara PA, Keller KE, McClendon MT, McPhee KE, Inglis DA, and Coyne CJ (2005) Y15_999Fw, a dominant SCAR marker linked to the Fusarium wilt race 1 (*Fw*) resistance gene in pea. Pisum Genet. 37: 32-35.

Pearce SR, Knox M, Ellis THN, Flavell AJ, and Kumar A (2000) Pea Ty1-copia group retrotransposons: transpositional activity and use as markers to study genetic diversity in Pisum. Mol. Gen. Genet. 263: 898-907.

Pilet-Nayel M, Muehlbauer F, McGee R, Kraft J, Baranger A, and Coyne C (2002) Quantitative trait loci for partial resistance to Aphanomyces root rot in pea. Theor. Appl. Genet. 106: 28-39.

Pritchard JK, Stephens M, and Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945.

Rafalski JA (2010) Association genetics in crop improvement. Curr. Opin. Plant Biol. 13: 174-180.

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley

J, Kresovich S, Goodman MM, and Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA 98: 11479-11484.

Samec P, Pošvec Z, Stejskal J, Našinec V, and Griga M (1998) Cultivar identification and relationships in *Pisum sativum* L. based on RAPD and isoenzymes. Biologia Plantarum 41: 39-48.

Schoen DJ and Brown AH (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc. Natl. Acad. Sci. USA 90: 10623-10627.

Simioniuc D, Uptmoor R, Friedt W, Ordon F, and Swiecicki W (2002) Genetic diversity and relationships among pea cultivars revealed by RAPDs and AFLPs. Plant Breed. 121: 429-435.

Simon CJ and Hannan RM (1995) Development and use of core subsets of cool-season food legume germplasm collections. HortSci. 30: 907.

Smith JSC and Smith OS (1992) Fingerprinting crop varieties. Ad. Agronomy 47: 3.

Smýkal P, Kenicer G, Flavell AJ, Corander J, Kosterin O, Redden RJ, Ford R, Coyne CJ, Maxted N, Ambrose MJ, and Ellis THN (2011) Phylogeny, phylogeography and genetic diversity of *Pisum* genus. Plant Genet. Res. 9: 4-18.

Smýkal P, Hýbl M, Corander J, Jarkovsk J, Flavell AJ, and Griga M (2008) Genetic diversity and population structure of pea (*Pisum sativum* L.) varieties derived from combined retrotransposon, microsatellite and morphological marker analysis. Theor. Appl. Genet. 117: 413-424.

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, and Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor. Appl. Genet. 111: 723-730.

Tamura K, Dudley J, Nei M, and Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24: 1596-1599.

Tanksley SD and McCouch SR (1997) Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. Science 277: 1063-1066.

Tar'an B, Zhang C, Warkentin T, Tullu A, and Vandenberg A (2005) Genetic diversity among varieties and wild species accessions of pea (*Pisum sativum* L.) based on molecular markers, and morphological and physiological characters. Genome 48: 257-272.

Tedford EC and Inglis DA (1999) Evaluation of legumes common to the Pacific northwest as hosts for the pea cyst nematode, *Heterodera goettingiana*. J. Nematol. 31: 155-163.

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, and Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA 98: 9161-9166.

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, and Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. 28: 286-289.

Vershinin AV, Allnutt TR, Knox MR, Ambrose MJ, and Ellis TH (2003) Transposable elements reveal the impact of introgression, rather than transposition, in Pisum diversity, evolution, and domestication. Mol. Biol. Evol. 20: 2067.

Weiguo Z, Eun-Jin P, Jong-Wook C, Yong-Jin P, Ill-Min C, Joung-Kuk A, and Gwang-Ho K (2009) Association analysis of the amino acid contents in rice. J. Integrative Plant Biol. 51: 1126-1137.

Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer, Sunderland, MA, pp. 445.

Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, and Holland JB (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203-208.

Zong X, Redden RJ, Liu Q, Wang S, Guan J, Liu J, Xu Y, Liu X, Gu J, Yan L, Ades P, Ford R (2009) Analysis of a diverse global *Pisum* sp. collection and comparison to a Chinese local *P. sativum* collection with microsatellite markers. Theor. Appl. Genet. 18: 193-204.

Zong X-X, Guan J-P, Wang S-M, and Liu Q-C (2008) Genetic diversity among Chinese pea (*Pisum sativum* L.) landraces as revealed by SSR markers. Acta Agron. Sin. 34: 1330-1338.