

Mol Breeding
DOI 10.1007/s11032-011-9662-y

Exploring the genetic diversity of Ethiopian grass pea (*Lathyrus sativus* L.) using EST-SSR markers

E. Shiferaw · M. E. Pè · E. Porceddu ·
M. Ponnaiah

Received: 19 August 2010 / Accepted: 19 October 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Expressed sequence tags (ESTs) in public databases and cross-species transferable markers are considered to be a cost-effective means for developing sequence-based markers for less-studied species. In this study, EST-simple sequence repeat (SSR) markers developed from *Lathyrus sativus* L. EST sequences and cross-transferable EST-SSRs derived from *Medicago truncatula* L. were utilized to investigate the genetic diversity among grass pea populations from Ethiopia. A total of 45 alleles were detected using eleven EST-SSRs with an average of four alleles per locus. The average polymorphism information content for all primers was 0.416. The average gene diversity was 0.477, ranging from 0.205 for marker *Ls942* to 0.804 for *MtBA32F05*. F_{ST} values estimated by analysis of molecular variance were 0.01, 0.15, and 0.84 for among regions, among accessions and within accessions respectively, indicating that most of the variation (84%) resides within accessions.

Model-based cluster analysis grouped the accessions into three clusters, grouping accessions irrespective of their collection regions. Among the regions, high levels of diversity were observed in Gojam, Gonder, Shewa and Welo regions, with Gonder region showing a higher number of different alleles. From breeding and conservation aspects, conducting a close study on a specific population would be advisable for genetic improvement in the crop, and it would be appropriate if future collection and conservation plans give due attention to under-represented regions.

Keywords *Lathyrus sativus* · EST-SSR · Genetic diversity

Introduction

Grass pea, *Lathyrus sativus* L. ($2n = 14$), belongs to the family Leguminosae, subfamily Papilionoideae, and is the only species widely cultivated as a food crop in the genus *Lathyrus*, whereas others species are cultivated to a lesser extent for both food and forage. It can survive in harsh environmental conditions, making the crop suitable for drought-prone areas of Africa and Asia (Campbell 1997). A strong and penetrating root system allows grass pea to grow in a wide range of soil types including very poor soils and heavy clays (Tiwaril and Campbell 1996), while its ability to utilize remnant water and soil nutrients makes the species adapted to low or zero inputs. In addition, it

Electronic supplementary material The online version of this article (doi:[10.1007/s11032-011-9662-y](https://doi.org/10.1007/s11032-011-9662-y)) contains supplementary material, which is available to authorized users.

E. Shiferaw · M. E. Pè · M. Ponnaiah (✉)
Scuola Superiore Sant'Anna, Piazza Martiri Della Libertà,
33, 56127 Pisa, Italy
e-mail: m.ponnaiah@sss sup.it

E. Porceddu
DABAC, University of Tuscia, Via S. C. De Lellis, 01100
Viterbo, Italy

effectively nodulates with *Rhizobium leguminosarum* (Yadav and Bejiga 2006), leaving the soil enriched in nitrogen for the following crop (Duke 1981; Campbell et al. 1994). Compared with other legumes, grass pea also has better resistance to many pests, including storage insects (Palmer et al. 1989).

Since it is a crop that can survive in harsh weather conditions, grass pea can serve as a survival food in difficult times. These appealing peculiarities make grass pea an interesting crop and encourage a thorough and extensive characterization of its germplasm. Unfortunately grass pea, together with many other legumes, belongs to the ‘research orphan crops’ and little information is available on it, particularly at the molecular level. DNA-based markers utilized so far in this species include random amplification of polymorphic DNA (RAPD) (Croft et al. 1999), amplified fragment length polymorphism (AFLP) (Tavoletti and Iommarini 2007) and very recently simple sequence repeat (SSR) (Lioi et al. 2011). Restriction fragment length polymorphism (RFLP) (Chtourou-Ghorbel et al. 2001) and inter-simple sequence repeat (ISSR) (Belaid et al. 2006) markers have also been utilized to evaluate the genetic relationships between different species in the genus. Microsatellites or SSRs are markers that consist of tandemly repeated units of short nucleotide motifs that are 1–6 bp long and can be amplified using polymerase chain reaction (PCR) (Weber and May 1989). Microsatellites are widely distributed throughout the genome, are codominant, usually multiallelic and hyper-variable (Katti et al. 2001). They have a high mutation rate which makes them highly polymorphic. SSR markers have been used in genome mapping, gene tagging, DNA fingerprinting, characterization of germplasm and cytogenetics research (reviewed by Gupta and Varshney 2000). The development of SSR markers is expensive and laborious, hence their application is largely limited to economically important crops and model species (Zane et al. 2002).

The presence of SSRs in gene coding regions and the availability of full and partial cDNA/EST (expressed sequence tag) sequences for many plant species provide an effective way to develop novel SSR markers directly from ESTs (Holton et al. 2002; Kantety et al. 2002; Choudhary et al. 2009), thus allowing the limitations to be overcome, especially in crops for which molecular information is scant. Molecular marker development using publicly available EST sequences can be achieved

for species for which substantial sequencing data are available. For many crop species, however, this is not the case and consequently this approach cannot be exploited. However, the cross-species transferability of molecular markers is an alternative approach, making it possible to use primers developed for other related species, check for their transferability and use them for the species of interest. Since EST-SSR markers are derived from transcribed regions of the DNA, they are expected to be more conserved and have a higher rate of transferability than genomic SSR markers (Scott et al. 2000). The availability of well studied model species facilitates the finding of cross-species transferable markers to be used in genetic studies of less-studied species. Like genomic SSRs, EST-SSRs (genic-SSRs) are useful for many applications in genetic studies including diversity analysis (Gupta et al. 2003; Thiel et al. 2003).

Grass pea is an important crop of economic significance in Ethiopia. It is the fifth most important pulse crop in the country and covers about 9% of the total pulse growing area (Tsegaye et al. 2005; CSA 2007). Diversity is a potential resource and a guarantee for improving a species of interest. It is also the source of new genes for combating threats to agricultural production caused due to biotic or abiotic factors (Frankel et al. 1995; Gepts 2006). Information about the level and extent of polymorphism of conserved materials is hence of great value for utilization of conserved genetic materials. In this study we took advantage of the available *L. sativus* EST sequences in the public domain to develop EST-SSRs in *L. sativus*; additionally, we utilized EST-SSR markers that were developed in the model species *Medicago truncatula* L., validating their cross-transferability to grass pea. The work then focused on studying the genetic variation and population structure among grass pea accessions collected from different regions of Ethiopia in order to understand the present diversity status and acquire information that will contribute to future collection, management and utilization of the crop.

Materials and methods

Plant material

Twenty grass pea accessions comprising a total of 240 individual plants (12 plants per accession) from

different regions of Ethiopia were kindly provided by the Institute of Biodiversity Conservation, IBC/Ethiopia (Electronic Supplementary Material Table 1; Electronic Supplementary Material Fig. 1). The GenElute Plant Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA) was used to isolate genomic DNA from 2 to 3-weeks-old grass pea leaves. DNA quality and quantification were estimated on 1% agarose gel electrophoresis and ethidium bromide staining, using lambda DNA as a reference.

EST-SSR marker development

Nineteen primer pairs were designed from the 178 *L. sativus* ESTs deposited in the public database (<http://www.ncbi.nlm.nih.gov/dbEST>), using Batch-primer3 software (<http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi>). SSR-containing sequences consisting of di-, tri-, and tetranucleotide repeats with a minimum of seven, four, or three sub-units, respectively, were selected (Electronic Supplementary Material Table 2). In addition, 24 EST-SSRs from *M. truncatula*, which were proven to be transferable to other legume species (Gutierrez et al. 2005), were selected and used (Electronic Supplementary Material Table 3).

PCR reaction and fragment analysis

PCR was performed in a final reaction volume of 15 μ l containing 30 ng genomic DNA, 5 \times PCR buffer, 0.2 mM each of dNTPs, 0.5 unit *GoTaq*[®] polymerase (Promega), 0.3 μ l each of forward and reverse primers, 0.02 mM labeled M13 primer (6-FAM/VIC/PET/NED) (Schuelke 2000). Amplicons were analyzed using an ABI3130xl genetic analyser (Applied Biosystems).

Data analysis

Allele size was determined as base pairs using GeneMapper[®] software v3.7 (Applied Biosystems). Allelic data were used to calculate the number, range and distribution of amplified alleles to determine variation level in the accessions studied. Allele frequency, gene diversity and polymorphic information content (PIC) of each EST-SSR marker were computed using the software PowerMarker 3.25 (Liu and Muse 2005). MICRO-CHECKER 2.2.1 (Van

Oosterhout et al. 2004) was used to check for potential genotyping errors such as allelic dropouts, stuttering or null alleles. Within each accession the frequency of null alleles was determined using Brookfield's estimate (Brookfield 1996), and allele and genotype frequencies were then adjusted accordingly. Numbers of alleles per locus, percentage of polymorphic loci, observed and expected heterozygosity were analyzed using GenAIEx 6.1 (Peakall and Smouse 2006) for each accession and accessions pooled by region of collection.

Population structure was examined using Structure 2.3.1 (Pritchard et al. 2000; Falush et al. 2003). The analysis was performed considering both the admixture model and the correlated allele frequencies between populations, with values of *K* set from 1 to 10 without incorporating population information into the analyses. *K* is the probable maximum population number that is assumed to represent and to contribute to the genotypes of sampled individuals. Each run was performed with 20,000 burn-in iterations and 200,000 subsequent Monte Carlo Markov Chain (MCMC) runs. To check the consistency of the results between runs with the same *K*, ten replicates were run for each assumed *K* value. The approach suggested by Evanno et al. (2005) was adopted to calculate the most likely value of *K* based on the second-order rate of change of the likelihood function with respect to *K* (ΔK). Once the number of genetic clusters was established, each individual was assigned to a cluster, and the overall membership of each sampled individual in the cluster was estimated. The genetic structure of the grass pea accessions was further investigated by analysis of molecular variance AMOVA using GenAIEx 6.1 (Peakall and Smouse 2006).

Results

Microsatellite validation and variability

From *L. sativus* ESTs, out of 19 SSR-containing sequences, three were duplicates and were excluded. Six of the primer pairs failed to give amplification products, one marker amplified a single fragment of more than the expected range and nine of the primer pairs gave amplification products of the expected range. From a total of 24 *M. truncatula*-derived EST-SSRs, 12 gave amplified fragments of the expected size in grass

pea. These and the nine *L. sativus*-derived EST-SSRs were subsequently tested for their ability to detect polymorphism in five grass pea accessions from different regions, each represented by 12 individuals. Polymorphism information content (PIC) values, which provide an estimate of the discriminatory power of each EST-SSR locus, were computed and polymorphic EST-SSRs were chosen. Eleven EST-SSRs (seven *L. sativus*- and four *M. truncatula*-derived EST-SSRs) that allowed polymorphism discovery either between or within accessions were utilized in the diversity analysis.

The 11 polymorphic EST-SSRs detected a total of 45 alleles in the 240 individual plants genotyped. The number of alleles per locus ranged from two (*Ls942*) to seven (*MtBA32F05*) and averaged four. PIC ranged from 0.184 (*Ls942*) to 0.776 (*MtBA32F05*) with a mean value of 0.416 (Table 1). The most informative markers were *MtBA32F05* and *MtBA10B02* with PIC values of 0.776 and 0.639, respectively. Rare alleles (frequencies <0.05) were observed in all markers except *Ls942*, the highest being in *Ls074*. Rare alleles represent 35% of the total alleles identified. The correlation coefficient between gene diversity (GD) and the number of alleles was high: $r = 0.825$ ($P < 0.05$). Allele frequencies were re-adjusted within populations to account for null alleles and diversity analysis was performed using the adjusted data. There was no evidence for large allele drop-out or stuttering for any of the SSR loci.

Diversity among accessions

Percentage of polymorphic loci, mean number of alleles per locus, Shannon's information index and observed and expected heterozygosity were calculated for each accession (Table 2). The effective number of alleles per locus ranged from 1.76 to 2.26 with an average of 1.96. The percentage of polymorphic loci ranged from 80 to 100% with mean 95.5%. Observed and expected heterozygosity ranged from 0.320 (accession 9) to 0.504 (accession 15) and from 0.354 (accession 20) to 0.470 (accession 6), respectively. Mean observed and expected heterozygosity were 0.404 and 0.419, respectively. Shannon's information index averaged 0.704, and showed the same trend as expected heterozygosity, with accession 20 showing the least diversity (0.595) and accession 6 being the most diverse (0.814).

Diversity among regions

To measure the diversity among regions, data of accessions belonging to the same region were pooled. The effective number of alleles per locus ranged from 1.76 to 2.35 with an average of 2.07 (Table 3). The percentage of polymorphic loci ranged from 90.91% to 100% with mean 97.4%. Shannon's diversity index ranged from 0.595 to 0.855 with mean 0.760. Observed and expected heterozygosity ranged from

Table 1 Characteristics of the EST-SSR markers used for germplasm analysis

Marker name	Repeat motif	No. of alleles	Size range (bp)	Allele size with high frequency (bp)	Major AF (%)	GD	H	PIC
<i>Ls989</i>	(GT)8	4	157–165	159	45	0.628	0.297	0.551
<i>Ls074</i>	(TC)7	5	160–168	162	67	0.477	0.430	0.413
<i>Ls617</i>	(GTTG)3	3	164–172	168	54	0.500	0.375	0.379
<i>Ls576</i>	(ATG)5	3	173–179	176	86	0.237	0.150	0.212
<i>Ls744</i>	(TTC)4	4	132–147	144	80	0.325	0.391	0.279
<i>Ls848</i>	(CAT)4	4	158–173	170	80	0.321	0.378	0.274
<i>Ls942</i>	(CCAA)3	2	137–145	137	88	0.205	0.223	0.184
<i>MtBA10B02</i>	(TGG)7	5	338–368	344	39	0.695	0.435	0.639
<i>MtBA52F10</i>	(TC)17	5	117–125	119	55	0.556	0.300	0.470
<i>MtBB52E08</i>	(AG)7	3	79–91	81	58	0.504	0.630	0.399
<i>MtBA32F05</i>	(AG)5	7	162–176	172	29	0.804	0.646	0.776
Mean		4				0.477	0.387	0.416

AF frequency of alleles, GD gene diversity, H heterozygosity, PIC polymorphism information content. All parameters were estimated using a set of 240 plants

Table 2 Diversity parameters of the 20 accessions of grass pea analyzed using EST-SSRs

Accession code	Accession no	P	Na	Ne	I	Ho	He
1	46019	100.00	2.63	2.00	0.711	0.429	0.431
2	234043	90.91	2.82	1.95	0.709	0.357	0.405
3	236700	100.00	2.82	1.98	0.726	0.372	0.421
4	235018	90.91	2.82	2.26	0.782	0.407	0.455
5	238241	90.91	2.91	1.91	0.717	0.404	0.418
6	238945	100.00	3.09	2.16	0.814	0.468	0.470
7	238955	100.00	2.73	1.91	0.689	0.345	0.408
8	236562	100.00	2.45	1.79	0.677	0.359	0.422
9	231325	90.91	2.73	1.97	0.686	0.320	0.392
10	238931	100.00	3.00	2.21	0.799	0.417	0.452
11	236568	100.00	2.54	1.99	0.708	0.389	0.428
12	46106	90.91	2.82	1.95	0.687	0.336	0.405
13	46015	100.00	2.54	1.88	0.696	0.451	0.430
14	238929	100.00	2.82	2.08	0.737	0.448	0.427
15	215706	100.00	2.36	1.95	0.692	0.504	0.441
16	46050	80.00	2.27	1.76	0.598	0.413	0.382
17	215246	100.00	2.54	1.86	0.653	0.469	0.402
18	207497	100.00	2.63	1.91	0.712	0.437	0.434
19	212742	90.91	2.81	1.97	0.692	0.424	0.427
20	241143	90.91	2.54	1.77	0.595	0.326	0.354
	Overall	95.5	2.69	1.96	0.704	0.404	0.419
	Mean (SE)	(1.25)	(0.073)	(0.054)	(0.025)	(0.015)	(0.013)

P percentage of polymorphic loci, *Na* number of different alleles, *Ne* number of effective alleles, *I* Shannon's information index, *Ho* observed heterozygosity, *He* expected heterozygosity

Table 3 Diversity parameters among seven regions of Ethiopia

Region	P	Na	Ne	I	Ho	He
Shewa	100	3.09	2.12	0.787	0.388	0.464
Tigray	100	3.36	2.04	0.781	0.403	0.442
Gojam	100	3.54	2.24	0.832	0.430	0.478
Welo	100	3.36	2.12	0.787	0.429	0.462
Gonder	100	3.63	2.35	0.855	0.434	0.474
Arsi	90.91	2.73	1.97	0.686	0.320	0.392
Hararge	90.91	2.54	1.76	0.595	0.326	0.354
Overall	97.40	3.18	2.07	0.760	0.390	0.430
[Mean (SD)]		(0.15)	(0.10)	(0.046)	(0.022)	(0.023)

P percentage of polymorphic loci, *Na* number of different alleles, *Ne* number of effective alleles, *I* Shannon's information index, *Ho* observed heterozygosity, *He* expected heterozygosity

0.320 to 0.434 and from 0.354 to 0.478, respectively. Mean observed and expected heterozygosity were 0.390 and 0.430, respectively. Regions Gojam, Welo and Gonder showed higher values in diversity measures, whereas Arsi and Hararge regions exhibited a low level of diversity.

Analysis of molecular variance (AMOVA)

Within and among components of total genetic variation were evaluated by AMOVA (Table 4). The results showed that the within-accession diversity explained most of the genetic diversity (84%). The

Table 4 Analysis of genetic differentiation among accessions of grass pea by AMOVA

Source of variation	<i>df</i>	SS	MS	Est. var.	%	<i>P</i> value
Among regions	6	120.357	20.059	0.086	1	0.004
Among populations	13	223.847	17.219	0.974	15	0.001
Within population	220	1,216.667	5.530	5.530	84	0.001
Total	239	1,560.871		6.590	100	

df Degrees of freedom, *SS* sum of squares deviations, *MS* mean squared deviation, *Est. Var.* estimates of variance components, % percentage of total variance contributed by each component

mean Φ_{pt} value (analogous to F_{ST}) of 0.15 indicated the presence of a moderate level of differentiation among accessions and a low level of differentiation (1%) among regions.

Genetic structure

Structure software was run for $K = 1-10$ based on the distribution of 41 alleles at ten EST-SSR loci among 240 grass pea individual plants; locus *Ls989*, which showed null alleles in most of the analyzed accessions, was removed from the analysis. Structure simulation demonstrated that the K value showed the highest peak at $K = 3$, suggesting that three populations could contain all individuals with the greatest probability. Hence a K value of 3 was selected to describe the genetic structure of the 20 accessions analyzed.

The estimated population structure of the accessions analyzed displayed partial membership to multiple clusters, with a few populations exhibiting distinctive identities. A graphic representation of the estimated membership coefficients to the three groups for each individual is given in Fig. 1a. Each individual is represented by a single vertical line broken into K segments, whose lengths are proportional to each of the K inferred clusters. Large admixture was observed among the accessions; however some of them (accessions 7, 8, and 11 from Shewa region), and accessions 13 and 16 from Gojam region appeared in distinct clusters, with less than 25% admixture from other clusters, suggesting that little gene flow occurred between these and other accessions.

Representation of the three clusters (Fig. 1b) indicated that cluster I was mainly composed individuals from Northern regions (Tigray, Gojam, Gonder and Welo). Cluster II contained individuals from all the growing regions, and cluster III consisted of individuals primarily from Shewa and Gojam, and a few

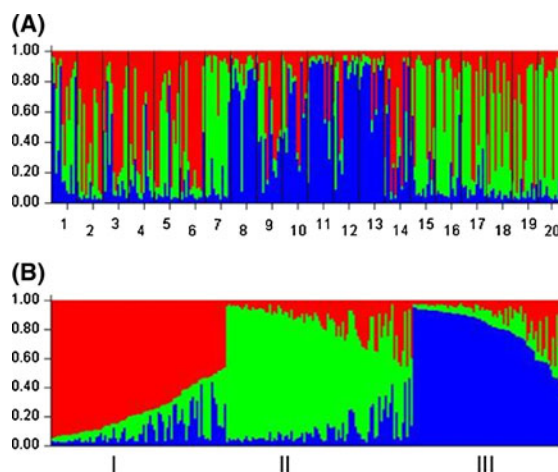


Fig. 1 a Estimated population structure of the grass pea landraces from Ethiopia ($K = 3$); each individual is represented by a single vertical line broken into K segments, with lengths proportional to each of the K inferred clusters. **b** Summary plot of estimated membership of 240 individuals in three clusters

representatives from Welo and Gondar regions. None of the clusters had individuals exclusively from one region.

Discussion

EST-SSRs developed from *L. sativus* EST sequences and transferable EST-SSRs from *M. truncatula* were evaluated and used for diversity analysis in Ethiopian grass pea. The PIC value of each marker was used to assess their informativeness, which is determined by the number of alleles and their frequency distribution within a population. Selecting only highly polymorphic markers might result in overestimating the overall genetic diversity, and thus to reduce the bias, all the markers were used for the diversity assay. In an independent study, Lioi et al. (2011) used *L. sativus*

ESTs to develop EST-SSR markers and studied the diversity in Italian grass pea. Six of the EST-SSRs used were developed from similar ESTs used in our study; however there are differences in the target core motifs in three of them. In terms of total number of alleles, 25 alleles were detected in the Ethiopian samples using seven EST-SSRs of *L. sativus* origin, and 17 alleles in Italian samples using six EST-SSRs. The number of alleles detected per marker and diversity level depends on the number and origin of genotypes analyzed and hence it is not easy to compare the level of diversity between different studies; nevertheless both studies have confirmed the value of EST-SSRs in assessing diversity in grass pea. To date, there have been no genomic SSRs developed for grass pea, and hence it was not possible to compare the level of polymorphism detected with genomic and EST-SSRs. Some studies that compared genomic and EST-SSRs using the same set of genotypes indicated polymorphism detected with EST-SSRs to be lower than that of genomic SSRs (Gupta et al. 2003; Chabane et al. 2005).

Cross-species transferability of molecular markers is beneficial for species that lack sequence information. In grass pea, transferability of SSRs from *L. japonicus* has been tested and low rate of transfer have been reported (Lioi et al. 2011). Among the model legumes, transferability of *M. truncatula*-derived markers to other legumes (Gutierrez et al. 2005; Gupta and Prasad 2009), including grass pea (Chandra 2011), has been demonstrated. In the present study, 50% of *M. truncatula* EST-SSRs tested were transferable to grass pea. These EST-SSRs were proven to be transferable to faba bean, chickpea and field pea at a transferability rate of 43, 39 and 40%, respectively (Gutierrez et al. 2005). The transferability rate to grass pea was relatively higher probably because only those EST-SSRs that were proven to be transferable to other legume crops were used. A higher transferability rate of EST-based markers is expected because of the conservative nature of cross-species transferable markers (Scott et al. 2000).

Though cross-species transferability is an important feature of EST-based markers, the screening of many markers may be needed to obtain informative ones. In this work, out of nine *L. sativus*-derived primer pairs that gave amplification products of the expected range, seven (77%) were polymorphic; and out of 12 *M. truncatula*-derived EST-SSRs only four

(33%) were polymorphic in grass pea. This indicated that more polymorphic or informative markers were developed from *L. sativus* than transferable ones from *M. truncatula*. As a result of the establishment of second-generation high-throughput sequencing technologies, it is expected that in the near future the number of EST sequences from *Lathyrus* deposited in the public database will significantly increase. This could facilitate the development and application of a large set of molecular markers useful for genetic studies in the species.

Diversity analysis among the various accessions studied here showed the presence of a moderate level of diversity. High levels of heterozygosity were observed in Gojam, Gonder, Shewa and Welo regions. Accessions from Gonder also showed a high number of different alleles. Tadesse and Bekele (2003) reported the presence of significant variation among grass pea accessions from Ethiopia based on morphological data. Their study showed a higher variability in accessions from Gondar and Tigray regions. The lowest diversity using EST-SSR markers was observed in Arsi and Hararge regions. This might be due to the limited sampling, since these two regions were represented by only one accession each. However, this could also be due to the actual low level of diversity present, since grass pea is not common in these two regions. Rare alleles (frequency ≤ 0.05) accounted for 35% of the total number of alleles detected in the analyzed populations. In addition to contributing to overall genetic diversity, rare alleles might be of adaptive significance and hence need due attention in conservation strategy (Park et al. 2008; Roussel et al. 2004).

The result of the genetic differentiation analysis showed the presence of moderate level of differentiation (mean $F_{ST} = 0.15$, $P < 0.001$), indicating that most of the variation was due to differences that existed among individuals within accessions (84%). A similar result has been reported by Chowdhury and Slinkard (2000) for grass pea accessions from 10 different geographical regions in which the within-regions diversity accounted for the majority (90.7%) of the total genetic diversity.

Population genetic structure across the accessions analyzed identified three groups in which individuals are clustered independently of their collection region, and it also showed admixture among accessions. The low genetic differentiation among regions could be interpreted by gene flow due to movement of seeds.

Seed exchange among farmers is a mechanism used to enhance diversity of local germplasm which may result in an increase in the distribution of alleles among different populations irrespective of their geographical distance (Louette et al. 1997). Grass pea's reproductive biology might also have contributed to the distribution of alleles among different populations. Although the floral biology of grass pea favors self-pollination (Yadav and Bejiga 2006; Campbell 1997), there are records of substantial out-crossing, which is dependent on environmental and/or genetic factors (Chowdhury and Slinkard 1997; Gutiérrez-Marcos et al. 2006).

This study demonstrated that the EST-SSRs developed for grass pea are useful tools for studying genetic diversity. The results showed the existence of moderate genetic variability in grass pea populations of Ethiopia which mostly resides within accessions, and indicated that different regions harbor comparable levels of diversity. Although the study was based on a limited number of markers, this observation should be taken into account in planning future conservation and research programs for the species. From the breeding aspect, conducting a close study on a specific population would be advisable when carrying out genetic improvement in the crop. The current grass pea collection in the Ethiopian genebank contains predominantly samples from Shewa region (about 45%). Thus, it would be useful to increase representative samples from other regions to capture the maximum diversity.

Acknowledgments We wish to acknowledge the Institute of Biodiversity Conservation, IBC/Ethiopia for kindly providing the grass pea accessions. Special thanks go to Ms. Htay Htay Aung for her help in statistical analysis and critical suggestions and interpretation. This work was supported by the International Doctoral Programme in Agrobiodiversity of Scuola Superiore Sant' Anna di Pisa, Italy. We also thank the anonymous reviewers who provided constructive comments on the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Belaid Y, Chtourou-Ghorbel N, Marrakchi M, Trifi-Farah N (2006) Genetic diversity within and between populations of *Lathyrus* genus (Fabaceae) revealed by ISSR markers. *Genet Resour Crop Evol* 53:1413–1418
- Brookfield JF (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol Ecol* 5:453–455
- Campbell CG (1997) Grass pea. *Lathyrus sativus* L. Promoting the conservation and use of underutilized and neglected crops. 18. Institute of Plant Genetics and Crop Plan Research, Gatersleben/International Plant Genetic Resources Institute. Rome, Italy
- Campbell CG, Mehra RB, Agrawal SK, Chen YZ, Abd EL, Moneim AM, Kawaja HIT, Yadav CR, Tay JU, Araya WA (1994) Current status and future strategy in breeding grass pea (*Lathyrus sativus*). *Euphytica* 73:167–175
- Chabane K, Ablett GA, Cordeiro GM, Valkoun J, Henry RJ (2005) EST versus genomic derived microsatellite markers for genotyping wild and cultivated barley. *Gen Res Crop Evol* 52:903–909
- Chandra A (2011) Use of EST database markers from *M. truncatula* in the transferability to other forage legumes. *J Environ Biol* 32:347–354
- Choudhary S, Sethy NK, Shokeen B, Bhatia S (2009) Development of chickpea EST-SSR markers and analysis of allelic variation across related species. *Theor Appl Genet* 118:591–608
- Chowdhury MA, Slinkard AE (1997) Natural out crossing in grass pea. *J Hered* 88(2):154–156
- Chowdhury MA, Slinkard AE (2000) Genetic diversity in grass pea (*Lathyrus sativus* L.). *Genet Resour Crop Evol* 47:163–169
- Chtourou-Ghorbel N, Lauga B, Combes D, Marrakchi M (2001) Comparative genetic diversity studies in the genus *Lathyrus* using RFLP and RAPD markers. *Lathyrus Lathyrism Newsl* 2:62–68
- Croft AM, Pang ECK, Taylor PWJ (1999) Molecular analysis of *Lathyrus sativus* L. (grass pea) and related *Lathyrus* species. *Euphytica* 107:167–176
- CSA (2007) Agricultural sample survey 2006/2007. Report on area and production of crops volume I. Central Statistics Authority, Ethiopia
- Duke JA (1981) Handbook of legumes of world economic importance. Plenum Press, New York
- Evanno G, Regnaut S, 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, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Frankel OH, Brown AHD, Burdon JJ (1995) The conservation of plant biodiversity. Cambridge University Press, Cambridge, pp 10–15
- Gepts P (2006) Plant genetic resources conservation and utilization: The accomplishments and future of a societal insurance policy. *Crop Sci* 46:2278–2292
- Gupta S, Prasad M (2009) Development and characterization of genic SSR markers in *Medicago truncatula* and their transferability in leguminous and non-leguminous species. *Genome* 52(9):761–771
- Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant

- breeding with emphasis on bread wheat. *Euphytica* 113:163–185
- Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Gen Genomics* 270:315–323
- Gutierrez MV, Vaz Patto MC, Huguet T, Cubero JI, Moreno MT, Torres AM (2005) Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theor Appl Genet* 110:1210–1217
- Gutiérrez-Marcos JF, Vaquero F, Sáenz de Miera LE, Vences FJ (2006) High genetic diversity in a world-wide collection of *Lathyrus sativus* L. revealed by isozymatic analysis. *Plant Genetic Resour* 4:159–171
- Holton TA, Christopher JT, McClure L, Harker N, Henry RJ (2002) Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. *Mol Breed* 9:63–71
- Kantety R, La Rota VM, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* 48:501–510
- Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol* 18(7):1161–1167
- Lioi L, Sparvoli F, Sonnante G, Laghetti G, Lupo F, Zaccardelli M (2011) Characterization of Italian grass pea (*Lathyrus sativus* L.) germplasm using agronomic traits, biochemical and molecular markers. *Genet Resour Crop Evol* 58:425–437
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21(9):2128–2129
- Louette D, Charrier A, Berthaud J (1997) In situ conservation of maize in Mexico: Genetic diversity and maize seed management in a traditional community. *Econ Bot* 51(1):20–38
- Oosterhout Van, Hutchinson WF, Wills DPM, Shipley PF (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538
- Palmer VS, Kaul AK, Spencer PS (1989) International Network for the Improvement of *Lathyrus sativus* and the Eradication of Lathyrism (INILSEL): a TWMRF initiative. In: Spencer PS (ed) *The grass pea: threat and promise*. Third World Medical Research Foundation, New York, pp 219–223
- Park Yj, Dixit A, Ma KH, Lee JK, Lee MH, Chung CS, Nitta M, Okuno K, Kim TS, Cho EG, Rao VR (2008) Evaluation of genetic diversity and relationships within an on-farm collection of *Perilla frutescens* (L.) Britt. Using microsatellite markers. *Genet Resour Crop Evol* 55:523–535
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Roussel V, Koenig J, Beckert M, Balfourier F (2004) Molecular diversity in French bread wheat accessions related to temporal trends and breeding programmes. *Theor Appl Genet* 108:920–930
- Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnol* 18:233–234
- Scott KD, Egger P, Seaton G, Rosetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. *Theor Appl Genet* 100:723–726
- Tadesse W, Bekele E (2003) Phenotypic diversity of Ethiopian grass pea (*Lathyrus sativus* L.) in relation to geographical regions and altitudinal range. *Gen Res Crop Evol* 50:497–505
- Tavoletti S, Iommarini L (2007) Molecular marker analysis of genetic variation characterizing a grasspea (*Lathyrus sativus*) collection from central Italy. *Plant Breed* 126(6):607–611
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411–422
- Tiwari KR, Campbell CG (1996) Inheritance of neurotoxin (ODAP) content flower and seed coat colour in grass pea (*Lathyrus sativus* L.). *Euphytica* 91:195–203
- Tsegaye D, Tadesse W, Bayable M (2005) Performance of grass pea (*Lathyrus sativus* L.) somaclones at Adet, northwest Ethiopia. *Lathyrus Lathyrism Newsl* 4:5–6
- Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44:388–396
- Yadav SS, Bejiga G (2006). *Lathyrus sativus* L. Record from Protabase. In: Brink M, Belay G (eds) *PROTA (Plant Resources of Tropical Africa)*. Wageningen, Netherlands (<http://database.prota.org/search.htm>)
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11:1–16