ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2011) **9**(2); 276–280
doi:10.1017/S1479262111000426

Development and application of EST-SSRs for diversity analysis in Ethiopian grass pea

M. Ponnaiah[1]*, E. Shiferaw[1], M. E. Pè[1] and E. Porceddu[2]

[1]*Scuola Superiore Sant'Anna, Piazza Martiri della Libertà, 33-56127 Pisa, Italy and* [2]*DABAC, University of Tuscia, Via S. C. De Lellis, 01100 Viterbo, Italy*

Abstract

To date, very limited molecular knowledge and molecular tools are available on grass pea (*Lathyrus sativus* L.). Herein, we present a genetic diversity analysis of 20 grass pea accessions collected from various regions of Ethiopia by means of seven newly developed *Lathyrus* expressed sequence tag-derived simple sequence repeat (EST-SSR) markers and four cross-transferable EST-SSRs derived from *Medicago truncatula* L. Forty-five alleles were detected among all the analyzed accessions. The number of alleles/locus ranged from two to seven with an average of four alleles/locus. The observed heterozygosity (H_o) ranged from 0.320 to 0.504, while the expected heterozygosity (H_e) ranged from 0.354 to 0.470. 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 accession. The model-based cluster analysis grouped the accessions into three clusters grouping accessions, irrespective of their collection regions.

Keywords: expressed sequence tag-derived simple sequence repeat; genetic diversity; *Lathyrus*

Introduction

Grass pea (*Lathyrus sativus* L.) is a legume crop widely cultivated in several arid and semi-arid countries, especially in Ethiopia, covering 9% of total pulse growing area (CSA, 2007).

DNA markers, particularly simple sequence repeats (SSRs) have proven to be very powerful tools in a variety of genetic studies in plants (Katti *et al*., 2001). The presence of SSR in gene coding regions and their availability of extensive full and partial cDNA/expressed sequence tag (EST) sequences for many plant species provide an effective way to develop SSR markers directly from EST (EST-SSR) (Eujayl *et al*., 2004; Holton *et al*., 2002; Kantety *et al*., 2002). This strategy allows to overcome crucial limitations associated to the development of SSR markers, especially in crops for which molecular information is scant, such as grass pea.

Here, we present the first assessment of genetic variability and genetic structure among Ethiopian grass pea accessions using EST-SSRs.

Material and methods

Plant material

A total of 240 plants, representing 20 grass pea accessions from different regions of Ethiopia, were analyzed. The GenElute Plant Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO, USA) was used to isolate genomic DNA from 2-3-week-old leaves.

EST-SSR marker development

Nineteen new EST-SSRs were designed from the 65 *L. sativus* ESTs deposited from public database (http://www.ncbi.nlM.nih.gov/dbEST), using Batchprimer3 software (http://probes.pw.usda.gov/cgi-bin/batchprimer3/batchprimer3.cgi).

* Corresponding author. E-mail: m.ponnaiah@sssup.it

In addition, 24 EST-SSR from *M. truncatula*, which proven to be transferable to other legume species (Gutierrez *et al.*, 2005) were selected to be used in grass pea.

PCR reaction and fragment analysis

PCR was performed in a final reaction volume of 15 µl containing 30 ng genomic DNA, 5 × PCR buffer, 0.2 mM each of dNTPs, 0.5 unit Go*Taq*® polymerase (Promega), 0.3 µl each of forward and reverse primers and 0.02 mM labelled M13 primer (6-FAM/VIC/PET/NED) (Schuelke, 2000). Amplicons were analyzed using an ABI3130xl genetic analyzer (Applied Biosystems).

Data analysis

Allele size was determined as base pairs using GeneMapper® Software v3.7 (Applied Biosystems). The allelic data were subjected to diversity analysis within and among the accessions using 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. Analysis of molecular variance (AMOVA) and other population statistics were measured using GenAlEx 6.1 (Peakall and Smouse, 2006) and population structure was examined using STRUCTURE 2.3.1 (Pritchard *et al.*, 2000; Falush *et al.*, 2003).

Results

Microsatellite validation and variability

Polymorphism screening was performed in five randomly chosen grass pea accessions. The screening revealed seven of the 19 *Lathyrus* EST and four of the 24 *Medicago* ESTs to be polymorphic and those 11 markers were utilized for the variation analysis.

From the 11 polymorphic EST-SSRs, a total of 45 alleles were detected in the 240 individual plants genotyped. The number of alleles/locus ranged from two (locus 942) to seven (locus MtBA32F05) and averaged four. Polymorphism information content (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 value of 0.776 and 0.639, respectively. Rare alleles (frequencies <0.05) were observed in all markers except marker Ls932, the highest being Ls074. Among the 45 alleles detected, rare alleles represent 35% of the alleles found in this analysis. The correlation coefficient between gene diversity (GD) and the number of alleles was high, $r = 0.825$ ($P < 0.05$).

Diversity among and within accessions

Allele frequencies were re-adjusted within populations to account for null alleles and diversity analysis was performed using on the adjusted data. Average value of effective number of alleles/locus, percentage of polymorphic loci, H_o and H_e, Shannon's information index (I) were 1.96, 95.5%, 0.404, 0.419 and 0.704, respectively; therefore, our data show that moderately high diversity exists among the accessions under the study.

Accessions were grouped as seven populations based on their collection site to measure the diversity among regions. Average of effective number of alleles/locus, percentage of polymorphic loci, I, H_o and H_e were 2.09, 97.4%, 0.760, 0.390 and 0.430, respectively. Regions 'Gojam, Welo and Gonder' showed higher values in

Table 1. Characteristics of the EST-SSR markers used

Marker name	Repeat motif	No. of alleles	Size range (bp)	Allele size (bp) (high frequency)	Major AF (%)	GD	H	PIC[a]
Ls989	(GT)8	4	157–165	159	45	0.628	0.297	0.551
Ls074	(TC)7	5	160–168	162	67	0.477	0.430	0.413
Ls617	(GTTG)3	3	164–172	168	54	0.500	0.375	0.379
Ls576	(ATG)5	3	173–179	176	86	0.237	0.150	0.212
Ls744	(TTC)4	4	132–147	144	80	0.325	0.391	0.279
Ls848	(CAT)4	4	158–173	170	80	0.321	0.378	0.274
Ls942	(CCAA)3	2	137–145	137	88	0.205	0.223	0.184
MtBA10B02	(TGG)7	5	338–368	344	39	0.695	0.435	0.639
MtBA52F10	(TC)17	5	117–125	119	55	0.556	0.300	0.470
MtBB52E08	(AG)7	3	79–91	81	58	0.504	0.630	0.399
MtBA32F05	(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; H, heterozygosity.
[a] Calculated over a set of 240 individual plants.

https://doi.org/10.1017/S1479262111000426 Published online by Cambridge University Press

diversity measures, whereas the 'Arsi and Hararge' region exhibited lower levels of diversity.

Genetic structure

Ls989 locus showed null alleles in most of the accessions and it was excluded from further analysis. STRUCTURE was run for $K = 1–10$ based on the distribution of remaining 41 alleles at ten EST-SSR loci among the 240 plants. STRUCTURE simulation produced the highest K value at $K = 3$.

STRUCTURE revealed also that cluster I is composed of individuals from Northern regions (Tigray, Gojam, Gonder and Welo). Cluster II comprised of individuals from all the growing regions, and cluster III consisted of individuals primarily from Shewa and Gojam, and a few representatives from Welo and Gondar. None of the clusters had individuals exclusively from one region only (Fig. 1).

AMOVA

AMOVA showed that the within-accession diversity explained most of the variation (84%). The mean *Φpt* value (analogous to F_{ST}), 0.15, indicated the presence of moderate level of differentiation among the accessions, and a low level of differentiation (1%) among regions (Supplementary Table S1, available online only at http://journals.cambridge.org).

Discussion

Development of SSRs from EST databases has shown to be a feasible option for obtaining high-quality nuclear markers. For under-studied crops, this method is a relatively cheap way of developing sequence-based markers (Gupta *et al.*, 2003; Ellis and Burke, 2007). Thirty-seven per cent (7 out of the 19 newly designed *Lathyrus* EST-SSRs) of newly designed EST-SSR markers were polymorphic between the accessions under the study. Seventeen per cent (4 out of 24 *Medicago* EST-SSRs) were polymorphic under the study. Diversity analysis among the analyzed accessions showed the presence of a moderate level of diversity. Our analysis also confirmed the transferability of this type of markers also to related species, as demonstrated from the successful utilization of *Medicago* EST-SSRs. On the other hand, we also demonstrated the successful transferability of *Lathyrus* EST-SSRs to related species such as groundnut and green peas (data not shown).

High levels of heterozygosity were observed in Gojam, Gonder and Welo regions. Accessions from Gonder also showed high number of different alleles. Tadesse and Bekele (2003) reported the presence of a 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 estimates using EST-SSR markers were 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, but it could also be due to the actual low level of diversity present, since grass pea is not common in these two regions. Uneven distribution of alleles was observed among the analyzed samples, as revealed by the number of rare alleles (frequency ≤ 0.05), which accounted for 35% of the total number of alleles detected.

Population genetic structure across the analyzed accessions identified three groups in which individuals are clustered independently of their collection region, and it also showed admixture among accessions. This relatively low genetic differentiation among regions

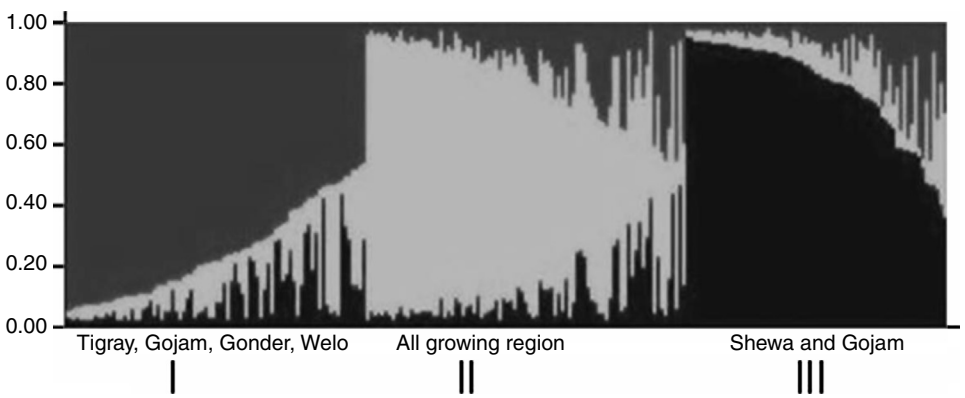

Fig. 1. Estimated population structure of grass pea landraces from Ethiopia. Summary plot of estimates by Q (estimated membership coefficients for each individual in the three clusters) as inferred from Structure.

https://doi.org/10.1017/S1479262111000426 Published online by Cambridge University Press

could be interpreted by gene flow due to movement of seeds – seed exchange among farmers being a mechanism used to enhance diversity of local germplasm and avoid crop failure. This results in an increase in distribution of alleles among different population, irrespective of geographical distance (Louette *et al.*, 1997). Grass pea reproductive biology might also have contributed to increase within-population variation. In fact, although the floral biology of grass pea favours self-pollination (Campbell, 1997; Yadav and Bejiga, 2006), there are records of substantial outcrossing, which is dependent on environmental and/or genetic factors (Chowdhury and Slinkard, 1997; Gutiérrez-Marcos *et al.*, 2006). Our result showed the existence of a significant level of variation among accessions, though most of the variation was present within populations. The current Ethiopian *ex-situ* collection of grass pea predominantly includes samples from the Shewa (45%) region. Based on present results, it might be useful to increase representative samples from other regions. Due to 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 at the public database will significantly increase. This could facilitate the development and application of a large set of molecular markers, such as EST-SSRs, which proved to be very informative and easy to develop. Due to modern genomic revolution, it is expected that the number of EST sequences getting deposited at the public database will increase exponentially and hence an increase in utilization of markers like EST-SSRs is highly recommended, as this study showed that the new EST-SSRs developed for the first time for grass pea are useful tools for the genetic diversity analysis in the species.

In our study, Gojam, Gonder and Welo regions had higher level of diversity compared with the others. This is consistent with the results of Tadesse and Bekele (2003), whose studies of grass pea accessions from Ethiopia were based on morphological data. Most of the variations were due to the differences between individuals within accession (84%), with moderate high level of population differentiation (mean $F_{ST} = 0.15$, $P < 0.001$). This may be due to farmer's habit to mix seeds from different sources before sowing as a mean to avoid crop failure. The results are similar to those reported by Chowdhury and Slinkard (2000), who mentioned that 90.7% of the variability was due to within-region diversity. Our results showed the existence of moderate genetic variability in grass pea populations of Ethiopia mostly within accession.

Although still based on a limited number of molecular markers, our study represents the most accurate description to date of *L. sativus* germplasm from Ethiopia, where this crop is often the only choice available to harness in adverse cultural conditions. Modern breeding strategies for this important but neglected crop might take advantage of the information provided herewith.

References

CSA (2007) *Agricultural sample survey 2006/2007*. Report on area and production of crops volume I. Ethiopia: Central Statistics Authority.

Campbell CG (1997) *Grass pea*. Lathyrus sativus *L. Promoting the conservation and use of underutilized and neglected crops. 18*. Rome: Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute.

Chowdhury MA and Slinkard AE (1997) Natural outcrossing in grasspea. *J Hered* 88: 154–156.

Chowdhury MA and Slinkard AE (2000) Genetic diversity in grass pea (*Lathyrus sativus* L.). *Genetic Resources and Crop Evolution* 47: 163–169.

Ellis JR and Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity* 99: 125–132.

Eujayl I, Sledge MK, Wang L, May GD, Chekhovskiy K, Zwonitzer JC and Mian MAR (2004) *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *Theoretical and Applied Genetics* 108: 414–422.

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.

Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N and Balyan HS (2003) Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Molecular Genetics and Genomics* 270: 315–323.

Gutiérrez-Marcos JF, Vaquero F, Sáenz de Miera LE and Vences FJ (2006) High genetic diversity in a world-wide collection of *Lathyrus sativus* L. revealed by isozymatic analysis. *Plant Genetic Resources* 4: 159–171.

Gutierrez MV, Vaz Patto MC, Huguet T, Cubero JI, Moreno MT and Torres AM (2005) Cross-species amplification of Medicago truncatula microsatellites across three major pulse crops. *Theoretical and Applied Genetics* 110: 1210–1217.

Holton TA, Christopher JT, McClure L, Harker N and Henry RJ (2002) Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. *Molecular Breeding* 9: 63–71.

Kantety R, La Rota VM, Matthews DE and Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant and Molecular Biology* 48: 501–510.

Katti MV, Ranjekar PK and Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution* 18: 1161–1167.

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

https://doi.org/10.1017/S1479262111000426 Published online by Cambridge University Press

Louette D, Charrier A and Berthaud J (1997) *In situ* conservation of maize in Mexico: genetic diversity and maize seed management in a traditional community. *Economic Botany* 51: 20–38.

Peakall R and Smouse PE (2006) GENALEX6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

Tadesse W and Bekele E (2003) Phenotypic diversity of Ethiopian grass pea (*Lathyrus sativus* L.) in relation to geographical regions and altitudinal range. *Genetic Resources and Crop Evolution* 50: 497–505.

Van Oosterhout C, Hutchinson WF, Wills DPM and Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538.

Yadav SS and Bejiga G (2006) In: Brink M and Belay G (eds) Lathyrus sativus *L. Record from Protabase*. Wageningen: PROTA (Plant Resources of Tropical Africa), http://database.prota.org/search.htm