Correspondence of DNA Markers for Genetic Diversity Studies among Different Apricot Genotypes from Cold Arid Deserts of Ladakh

Gyan P. Mishra
Defence Institute of High Altitude Research, C/o 56 APO, Leh
Email: gyan.gene@gmail.com

ABSTRACT

The phylogenetic relationships of 36 locally grown apricot (Prunus armeniaca L.) genotypes which are collected from nine sampling sites from two valleys viz. Nubra (9,600 ft) and Leh (11,500 ft) of trans-Himalayan region were analyzed using 31 PCR markers (20 RAPDs and 11 ISSRs). RAPD analysis yielded 139 fragments, of which 136 were polymorphic, with an average of 6.8 polymorphic fragments per primer. Out of 20 ISSR primers used, 11 amplified and produced 58 bands, of which 56 were polymorphic, with an average of 5.09 polymorphic fragments per primer. The primers based on (CT) n produced maximum number of bands (nine) while, (AT)n and many other motifs gave no amplification. Clustering of genotypes within groups was not similar when RAPD and ISSR derived dendrogram were compared, whereas the pattern of clustering of the genotypes remained more or less the same in RAPD and combined data of RAPD+ISSR. The results of PCA analysis were comparable to the cluster analysis. These analyses, allowed us to identify the groups corresponding to the two apricot collection sites.

Keywords: Prunus armeniaca; Apricot; Genetic diversity; ISSR; RAPD; Trans-Himalaya

1. INTRODUCTION

Apricot (Prunus armeniaca L.) is the most important fruit crop of cold arid regions of India which cover 3,200 hectare with a total production of 5,200 MT (Dwivedi and Attrey, 2002). Wild apricot popularly called as Zardalu, appears to be originated in Indian Himalayas. It is believed to have been introduced in cold arid Ladakh via Baltistan though some experts opined that it was introduced directly from China via Tibet. Classical approaches for identification and analysis of genetic variability in fruit crops are based on morphological, physiological and agronomic traits (UPOV, 1976). But, due to low polymorphic levels has led to search for alternatives like DNA markers, which are independent of environmental conditions and are unaffected by developmental stages of the plant. The evaluation of genetic diversity and construction of linkage maps would promote the efficient use of genetic variations in the breeding program (Paterson, et al., 1991).

The PCR technology has offered new marker systems for diagnosis of genetic diversity in large-scale studies (Saiki, et al., 1988). Over the last 15 years, PCR technology has led to the development of two simple and quick techniques called Random Amplified Polymorphic DNA (RAPD) and Inter Simple Sequence Repeats (ISSR). These markers have been used both for DNA fingerprinting (Martín and Sánchez-Yélabo, 2000; Blair, et al., 1999) and population genetic studies (Wolfe, et al., 1998).

In the present study, RAPD and ISSR marker techniques were used to determine the genetic variability in apricot accessions from two valleys, i.e., Leh and Nubra of Ladakh region in India. To our knowledge, no report has been published on the genetic diversity, population structure and gene flow among the populations of Prunus armeniaca from the trans-Himalayan region with molecular markers like RAPD and ISSR for diversity analysis. In this study, we investigated the genomic relationship among the different populations of Prunus armeniaca and their relationship in cold arid desert of trans-Himalayan region, to provide insight to facilitate conservation management of the populations.

2. MATERIALS AND METHODS

2.1 Plant Materials

The plant materials were obtained from two valleys (Leh and Nubra) with altitude ranging 9,600 m (Nubra) to 11500 m (Leh) from the cold arid desert of Trans-Himalayas (Table 1). Thirty-six locally grown genotypes collected from 9 different villages (4 genotypes from each village) were used in the present investigation. The interval between samples was 100-500 m, the pair wise distance between populations was 5–35 km, whereas, the pair wise distance between valley divisions was 50–250 km.

2.2 DNA Extraction

Total genomic DNA was extracted from frozen leaves
(5g) by the CTAB method (Saghai-Marooof, et al., 1984) with minor modifications, which included the use of 200 mg per sample polyvinyl pyrollidone (PVP, MW: 40,000).

2.3 Random Amplified Polymorphic DNA Amplification

Twenty random decamer primers from IDT Tech, USA, (Table 2.2) were used for amplification following the protocol of Williams, et al. (1990). Amplification reaction were performed in volumes of 25 µl containing 10 mM Tris-HCl (pH 9.0), 1.5 mM MgCl2, 50 mM KCl, 200 µM of each dNTPs, 0.4 µM primer, 20 ng template DNA and 0.5 unit of Taq polymerase (Sigma). DNA amplification was performed using a Biometra Gradient (Germany), thermal cycler. The first cycle consisted of denaturation of template DNA at 94 °C for 5 min., primer annealing at 37 °C for 1 min, and primer extension at 72 °C for 2 min. In the next 40 cycles the period of denaturation was reduced to 1 min. at 92 °C, while the primer annealing and primer extension time remained the same as in the first cycle. The last cycle consisted of only primer extension (72 °C) for 7 min. The PCR products were stored at 4 °C before analysis.

2.4 Inter Simple Sequence Repeats Amplification

The ISSR primers were obtained from Applied Biosciences, and PCR amplification was performed in a 25 ml reaction volume containing 100 mM Tris-HCl (pH 8.3), 15 mM MgCl2, 10 mM each of dNTP, 0.4 mM of primer, 0.01% gelatin, 1 unit of Taq polymerase and 25 ng of genomic DNA. Initial denaturation for 5 min at 94°C was followed by 40 cycles of 1 min at 94°C, 1 min at specific annealing temperature (±5°C of Tm). 2 min at 72°C and a 10 min final extension step at 72°C (Table 2.2).

2.5 Agarose Gel Electrophoresis

Amplification products were electrophoresed on 1.5 per cent agarose gel (Life Science Technologies, USA) at constant voltage (50 V) in 1X TBE for approximately 2 h, visualised by staining with ethidium bromide and documented on a gel documentation system.

2.6 Data Collection and Analysis

The banding patterns obtained from RAPD and ISSR were scored as present (1) or absent (0), each of which was treated as an independent character. Gel-Pro analyzer version 3-1 software was used to score RAPD and ISSR profiles. Jaccard’s similarity coefficient (J) was used to calculate similarity between pairs of accessions. Where, [J= nxy / (nx - nxy)], nxy is the number of bands common to accession A and accession B; nx the total number of bands present in all samples and ny the number of bands not present in A and B but found in other samples. The similarity matrix was subjected to cluster analysis by unweighted pair group method for arithmetic mean (UPGMA) and a dendrogram was generated using the program NTSYS-PC (Rohlf, 1992). FreeTree software (Pavlicek, et al., 1999) was used for construction of phylogenetic trees on the basis of distance data and for bootstrap analysis of the trees robustness. POPGENE software was used to calculate Nei’s unbiased genetic distance among different species with all markers, including monomorphic markers.

Data for observed number of alleles (Na), effective number of alleles (Ne), Nei’s genetic diversity (H), Shannon’s information index (I), number of polymorphic loci (NPL) and percentage polymorphic loci (PPL) across all the nine populations were also analyzed (Zao, et al., 2006). Within species diversity (Hs) and total genetic diversity (Ht) (Nei, 1978) were calculated within the species and within four major groups (as per their collection site) by POPGENE software. The RAPD and ISSR data were subjected to a hierarchical analysis of molecular variance (AMOVA), as described by Excoffier, et al. (1992). GenAlEx software was used to calculate a principal coordinates analysis (PCA) that plots the relationship between distance matrix elements based on their first two principal coordinates (Peakall and Smouse, 2001). Regression (R²) between two matrices obtained with two marker types was estimated using Nei’s genetic diversity. According to Prevost and Wilkinson (1999) the resolving power (Rp) of a primer is: Rp = £ IB where IB (band informativeness) takes the value of: 1–[2x (0.5–P)], P being the proportion of the 36 genotypes containing the band.

3. RESULTS

3.1 RAPD Band Patterns

All the chosen primers amplified with the number of amplified fragments ranging from one (S25, 28 and 37) to 13 (S21) and which varied in size from 200-2,500 bp and yielded 3004 fragments. Out of 139 amplified bands, 136 were found polymorphic, with an average number of bands per primer and average numbers of polymorphic bands per primer as 6.95 and 6.8 respectively. Percentage polymorphism ranged from 75% (S29) to a maximum of 100% (18 primers), with an average of 97.84% polymorphism. The resolving power (Rp) of the RAPD primers ranged from 3.0 (S37) to 13.39 (S21) (Table 2.1). The total number of fragments amplified, number of polymorphic fragments, percentage of polymorphic loci (P%), and resolving power (Rp) obtained for each primer are shown in the Table 2.1 and comparative list is presented in the Table 2.3.

3.2 Genetic Variability Details using RAPD Markers

The respective values of Na, Ne, H, I, Ht, Hs, NPL and PPL were found maximum for Stok genotypes while minimum values are observed for Stakna (Na and I) and Hundar (Ne, H, Ht, NPL and PPL) genotypes (Table 3.1 and 3.2). When these parameters were studied for valley divisions (i.e. Leh and Nubra valleys) then all these respective values were found higher for Leh valley genotypes indicating that there is more variability in Leh valley than in Nubra valley for the samples taken in the present investigation (Table 3.3).
Mean coefficient of gene differentiation (Gst) value 0.3674 indicated that 63.26% of the genetic diversity resided within the population. Based on the Gst value, the mean estimated number of gene flow (Nm) between populations was found 0.4304 (Table 3.4). AMOVA helps in partitioning of the overall RAPD variations among groups and among populations within the group. Molecular variance among valley (11%), among population (12%) and within the population (77%) (Table 4) indicating that there are more variations with in the population. This is helpful in making strategy for germplasm collection and evaluation.

3.3 Random Amplified Polymorphic DNA
Derived Dendrogram Analysis

A dendrogram based on UPGMA analysis grouped the 36 genotypes into eight main clusters and Jaccard’s similarity coefficient ranging from 0.05 to 0.39 (Fig. 2). Cluster I, II, III, V and VI represents the genotypes from Leh valley while, cluster IV, VII and VIII have all the genotypes from Nubra valley. The results of PCA analysis were comparable to the cluster analysis (Fig. 3). The first three most informative PC components explained 34.82 per cent of the total variation. Three genotypes from Nubra valley, i.e., Khardung 1, Khalsar 3 and Hundar 4 (i.e., 21, 27 and 32) appear to be distinct from other genotypes in the PCA.

3.4 Inter Simple Sequence Repeats-Band Patterns

Out of 20, only 11 ISSR primers give rise to reproducible amplification products (Table 2.2). The sequences of these 20 primers seem to indicate that microsatellites more frequent in Apricot contain the repeated dinucleotides (AG)n, (GA)n, (TG)n, (CT)n, (AC)n, (GT)n, and trinucleotides (ACC)n, (CCG)n, (GCG)n. The number of bands produced with different repeat nucleotide were more with the (GT)n, (GA)n, (CT)n, and (AC)n primers (ISSR8, 6, 7 and 9). The primers that were based on the (GA)n, poly (CT)n and (GT)n motif produced more polymorphism on average (7 bands per primer) than the primers based on any other motifs used in the present investigation. The primer sequences which did not amplified in the present investigation contain the dinucleotides repeats as (AT)n, (GT)n, (TC)n, (TA)n, (CT)n; tri-nucleotides repeats (TGC)n and the repeated tetra-nucleotides sequences (CTAG)n and (TGCA)n (Table 2.2).

The 11 primers on an average produced 58 bands across 36 genotypes, of which 56 bands were polymorphic, accounting for 96.55%. Number of bands varied from three (ISSR 3) to seven (ISSR 6, 7 and 8), and sizes ranged from 200-2500 bp. Average numbers of bands and polymorphic bands per primer were 5.27 and 5.01 respectively. Percentage of polymorphism ranged from 80% (ISSR 9) to 100% (nine primers), with an average polymorphism of 96.5% across all the genotypes. Total no of bands amplified ranged from 45 (ISSR 3) to 174 (ISSR 8). The Rp ranged from 2.5 (ISSR 3) to 9.67 (ISSR 8). Primer amplification details as obtained for each ISSR primer are shown in Table 2.2 and comparative list is presented in Table 2.3.

3.5 Genetic Variability details using ISSR Markers

The respective values of Na, Ne, H, I, Ht, Hs, NPL and PPL across all the nine populations were found maximum for Stok genotypes while it is minimum for Khalsar and Hundar genotypes (Table 3.1 and 3.2). For valley divisions (i.e. Leh and Nubra valley) all these respective values were found higher for Leh valley genotypes indicating that there is more variability in Leh valley than in Nubra valley for the samples analyzed (Table 3.3).

Mean coefficient of gene differentiation (Gst) value 0.4552 indicated that 54.48 % of the genetic diversity resided within the population (Table 3.4). Molecular variance among valley (22%), among population (16%) and within the population (62%) (Table 4) indicating that there is more variations with in the population.

3.6 Inter Simple Sequence Repeats-Derived Dendrogram Analysis

The complete data was based on a total of 1367 bands. The 36 genotypes were clustered into five clusters where, cluster I, II and V represents all the genotypes from Leh valley while, cluster III, and IV contains all the genotypes from Nubra valley. Jaccard’s similarity coefficient ranged from 0.05 to 0.39 (Fig. 2). The results of PCA analysis were comparable to the cluster analysis. The first three most informative PC components explained 56.39 % of the total variation. Two genotypes i.e. Stok 3 and Hemis 3 (i.e., 7 and 11) appear to be distinct from other genotypes in the PCA.

3.7 Genetic Variability details for RAPD+ISSR Combined Data

The respective values of Na, Ne, H, I, Ht, Hs, NPL and PPL across all the nine populations were found maximum for Stok genotypes while these are minimum for Hundar genotypes (Table 3.1 and 3.2). When these parameters were analyzed for valley divisions, then the values were found higher for Leh valley genotypes (Table 3.3). Value for Ht and Hs were found to be 0.3452, 0.2097 respectively. Mean coefficient of gene differentiation (Gst) value 0.3922 indicated that 60.78% of the genetic diversity resided within the population (Table 3.4). Molecular variance among valley (15%), among population (16%) and within the population (72%) indicating that there are more variations with in the population (Table 4).

3.8 RAPD and ISSR Combined Data for Cluster Analysis

The UPGMA dendrogram obtained from the cluster analysis of RAPD and ISSR combined data gave near similar clustering pattern, with Jaccard’s similarity coefficient ranging from 0.12 to 0.46 (Fig. 2). The 36 genotypes were clustered into four major clusters. Cluster I and II represents all the genotypes from Leh valley while, cluster III, and IV contains all the genotypes from Nubra valley. Both RAPD and ISSR clusters showed partial similarity with combined data of
RAPD+ISSR. The results of PCA analysis were comparable to the cluster analysis. The first three most informative PC components explained 33.73% of the total variation. The two-dimensional ordination confirms the cluster analysis results showing that two genotypes i.e. Stok 4 and Khardung1 (8 and 21) were separated (Figure 3.3). Other genetic variation studies were also performed on ISSR and RAPD combined data which are represented in different tables (Tables 3.1, 3.2, 3.3, 3.4).

3.9 Comparative Analysis of RAPD with ISSR Markers

Range of similarity index for RAPD markers (0.09-0.49), ISSR markers (0.05-0.39) and RAPD+ISSR marker data (0.12-0.46) indicating more diversity in case of RAPD (Figures 2.1, 2.2, 2.3). Clustering of genotypes within groups was not similar when RAPD and ISSR derived dendrograms were compared, whereas the pattern of clustering of the genotypes remained more or less the same in RAPD and combined data of RAPD+ISSR. Although with in the valley genotypic differences were there but the genotypes from different villages did not fall into a distinct clustering pattern. While genotypes from two valleys (i.e., Leh and Nubra valley) were present in distinct clusters in all the three cases (RAPD, ISSR and RAPD+ISSR) showing that there is more variation between valley genotypes.

4. DISCUSSION

4.1 RAPD and ISSR Markers

In this work we compared the applicability of ISSRs and RAPDs as genetic markers to characterize the 36 apricot genotypes (Fig. 1; Table 1) which were locally grown. However, no such reports on genetic diversity using molecular markers were available in the genus apricot from this region. The evolution of genotypes in distinct agro-climatic zones demonstrates significant levels of variation in response to the selection pressure in the zones (Singh, et al., 1998). It is, therefore, not surprising to find significant levels of polymorphism among the 36 genotypes of apricot with RAPD and ISSR markers.

RAPD markers were found more efficient with regards to polymorphism detection, as they detected 97.84% as compared to 96.5% for ISSR markers. Also, the average number of polymorphic bands per primer and total number of polymorphic bands are more for RAPD (6.8 and 136 respectively) than for ISSR (5.09 and 56 respectively) (Table 2.3). This is in contrast to the results as obtained for several other plant species like wheat (Nagaoka and Oghihara, 1997) and vigna (Ajibade, et al., 2000). The 20 RAPD and 11 ISSR primers in the present study yielded 192 polymorphic markers that unambiguously discriminated 36 genotypes into four clusters.

The regression test between the two Nei’s genetic diversity indexes was performed. This resulted in low regression between RAPD and ISSR based similarities (R² = 0.1996; Figure 4.1), moderate for ISSR and ISSR+RAPD (0.5366; Figure 4.2), while it is maximum for RAPD and ISSR+RAPD based similarities (0.8766; Figure 4.3). This shows that RAPD data is more close to ISSR+RAPD combined data. A possible explanation for the difference in resolution of RAPDs and ISSRs is that the two-marker techniques target different portions of the genome. There was some consensus between the RAPD and ISSR based grouping of the 36 apricot genotypes.

4.2 Dendrograms obtained using RAPDs and ISSRs

The differences found among the dendrograms generated by RAPDs and ISSRs could be partially explained by the different number of PCR products analysed (3004 for RAPDs and 1367 for ISSRs) reinforcing again the importance of the number of loci and their coverage of the overall genome, in obtaining reliable estimates of genetic relationships among apricot cultivars. Similar results have been observed by Loarce, et al. (1996) in barley. Another explanation could be the low reproducibility of RAPDs (Karp, et al., 1977). Also the phylogenetic analysis on the basis of RAPD and ISSR-derived dendrogram supports the fact that region specific variations are there, which is because of the multiple generations of selection carried out after their introduction.

The putatively similar bands originating for RAPDs in different individuals are not necessarily homologous, although they may share the same size in base pairs. This situation may lead to wrong results when calculating genetic relationships (Fernandez, et al., 2002). Dendrograms in the present study did not indicate very clear pattern of clustering for within valley samples but, a clear pattern was observed in all the three cases for between valley samples. Similar results were obtained in Azukibeans (Fernandez, et al., 2002) and groundnut (Dwivedi, et al., 2001). The genetic closeness with in the Leh valley and Nubra valley cultivars can be explained by the high degree of commonness in their genotypes. Similar results were also observed in blackgram (Gaffor, et al., 2001).

In all the dendrograms, Nimu (2,3,4) and Thoise (2,3,4) genotypes were found clustered together. The genetic similarity of these genotypes is probably associated with their similarity in the genomic and amplified region. The level of observed polymorphism is very high, and the ability of the ISSR technique to effectively distinguish species in the genus Vigna was reported by Ajibade, et al. (2000). The primers with (CT)n and (GT)n motifs produced more polymorphism than any other motif. While primers with (AT)n and other motifs (Table 2.2) did not gave any amplification. Somewhat similar results were also reported by Ajibade, et al. (2000) in the genus Vigna.

As previously pointed out, during the ISSR screening we obtained good amplification products from primers based on (CT)n and (GT)n repeats while (AT)n and some other primers gave no amplification products (Table 2.2), despite the fact that (AT)n dinucleotide repeats are thought to be the most abundant motifs in plant species (Martín and Sánchez-Yélamo, 2000; Morgante and Olivieri, 1993). Similar
results were obtained in grapevine (Moreno, et al, 1998), rice (Blair, et al., 1999) and wheat (Nagaoka and Ogihara, 1997). A possible explanation of these results is that ISSR primers based on AT motifs are self-annealing, due to sequence complementarity, and would form dimers during PCR amplification (Blair, et al., 1999). Reason behind non amplification of other repeats may be their absence in the genome.

Our results indicate the presence of great genetic variability among local non-descriptive genotypes of apricot. ISSR and RAPD markers are equally useful in the assessment of apricot diversity, the detection of duplicate sample in germplasm collection, and the selection of a core collection to enhance the efficiency of germplasm management for use in apricot breeding and conservation.

ACKNOWLEDGEMENT
The author is thankful to DRDO HQ for its support and funding the project.

REFERENCES
24. Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski,

