Molecular and morphological characterization of S2 lines of black corn (Zea mays L.) from Ecuadorian Andes

María de Lourdes Torres1,∗, Ana Lucía Bravo1, Mario Caviedes1 Venancio Arahana1

1Laboratorio de Biotecnología Vegetal, Colegio de Ciencias Biológicas y Ambientales.
Universidad San Francisco de Quito
Calle Diego de Robles s/n y Vía Interoceánica, Quito, Ecuador
E-mail: ltorres@usfq.edu.ec

Abstract

Black corn is a native variety of this crop, present in the Andean Highlands of Ecuador. Despite its importance as a functional food, due to its anthocyanin content with antioxidant and anti-carcinogenic properties, its usage has tended to be reduced because of a widespread preference in farmers and consumers for the improved yellow corn varieties. In this research, twenty four S2 lines of black corn were morphologically and molecularly evaluated to determine their diversity level and thus their potential as a source of parents for future breeding programs, especially aimed to developing varieties with high anthocyanin content. For the molecular evaluation, a set of ten SSR primer pairs, one per chromosome, was used to amplify microsatellite regions from the corn genome. For the morphological analysis, seven traits were measured in two vegetative cycles of self pollination. A total of 43 alleles were revealed for the 10 SSR loci. The number of alleles per locus ranged from 2 to 8, with an average of 4.3± 2.06. The genetic distances ranged from 0.29 to 0.95. UPGMA analysis grouped the S2 lines in six groups with one predominant group. The morphological data showed variation especially in plant height, ear insertion height and leaf length. Endogamy depression was evidenced mainly for plant height and ear insertion height. A morphological dendrogram also grouped the 24 S2 lines of black corn in six groups, although they were not the same as in the molecular dendrogram. It is concluded that a moderate variability level exists in this collection and therefore a potential for the development of new varieties.

Keywords. Zea mays L., black corn, anthocyanins, morphological characterization, SSRs, genetic diversity.

Resumen

El maíz negro es una variedad nativa de este cultivo, presente en las zonas elevadas de los Andes ecuatorianos. A pesar de su importancia como un alimento funcional, por su alto contenido de antocianinas con propiedades antioxidantes y anti cáncer, su uso ha sido reducido a causa de la preferencia generalizada por variedades de maíz amarillo mejorado por parte de agricultores y consumidores. En este estudio, 24 líneas S2 de maíz negro fueron evaluadas morfológica y molecularmente para determinar su nivel de diversidad y por lo tanto su potencial como parentales para programas de mejoramiento futuros, especialmente aquellos enfocados en el desarrollo de variedades con alto contenido de antocianinas. Para la evaluación molecular, un set de diez primeros de SSR, uno por cromosoma, fueron usados para amplificar una región de microsatélites del genoma del maíz. Para el análisis morfológico, siete características fueron medidas en dos ciclos vegetativos de autofecondación. Un total de 43 alelos fueron encontrados para los 10 loci de SSR. El número de alelos por locus varió entre 2 y 8, con un promedio de 4.3± 2.06. La distancia genética varió entre 0.29 y 0.95. El análisis UPGMA agrupó a las líneas S2 en seis grupos con uno predominante. Los datos morfológicos mostraron variación, especialmente en cuanto a la altura de la planta, altura de inserción de espiga y longitud de hoja. Se evidenció depresión endógamica principalmente para altura de planta y altura de inserción de espiga. El dendrograma de este análisis, también agrupó a las 24 líneas S2 de maíz negro en seis grupos, aunque no fueron los mismos grupos que los presentados en el análisis molecular. Se concluye que existe un nivel moderado de variabilidad en esta colección y, por lo tanto, posee potencial para el desarrollo de nuevas variedades.

Palabras Clave. Zea mays L., maíz negro, antocianinas, caracterización morfológica, SSRs, diversidad genética.
Introduction

Corn (Zea mays L.) is part of the group of the most important cereals in the world. Its importance is related with the variety of uses in which it is involved, such as human food, animal feed and as a source for industrial products [1]. There are two possible centers of origin of this cereal: the highlands of Peru, Ecuador, and Bolivia; and the southern Mexico and Central America region [2].

In Ecuador, 24 native races of corn have been identified. One of these races is the black corn, also known as “grape bunch” [3]. This black corn is characterized by its rounded kernels with red or black pericarp which are tightly grouped giving a grape bunch appearance. The cob is medium size with a conic shape; ear row number is around eight to fourteen, stem is thin and plant height is short [3]. For centuries, black corn has been used as a colorant for food and beverages in the Andes, because of its rich anthocyanin content [4]. Studies have shown that anthocyanins found in black corn have health benefits, such as antioxidant and antimutagenic properties [5, 6]. Moreover, it has been demonstrated that black corn has higher antioxidant content than fresh blueberries [7]. Therefore, black corn is considered a potential crop in the functional and nutraceutical food industry. Due to the growing demand of natural food colorants and functional foods, this type of corn represents a valuable genetic resource from the Ecuadorian Andean Highlands.

Both phenotypic and genetic diversity are necessary to identify promising individuals in order to evaluate and utilize genetic resources for the development of commercial lines [8]. Specifically in corn breeding, the assessment of genetic diversity is essential to the exploitation of heterosis [9, 10, 11, 12]. Molecular markers assess genetic diversity at DNA level, do not present interactions with the environment, and identify high polymorphism levels which can be used for characterization of corn accessions [10, 13]. In particular, SSR markers in corn have been used to construct genetic linkage maps, study populations and evolution, assist breeding programs, and characterize inbred lines [9, 14, 15, 16].

Genetic variability in Ecuadorian highland corn populations has been studied by different groups, and all have reported high genetic diversity [4, 17, 18]. Few studies have been focused on the molecular characterization of highland corn, especially in races like black corn. Among these, in 2003, INIAP (Instituto Nacional Autónomo de Investigaciones Agropecuarias) in Ecuador [19], characterized 19 accessions and 9 lines from different types of highland corns, using 22 microsatellite sequences.

In this study, we evaluated 24 black corn S2 lines with microsatellites and morphologic traits to assess their genetic diversity and to analyze the possibility of developing commercial lines of Ecuadorian “grape bunch” black corn.

Materials and Methods

Plant material

Twenty four black corn S2 lines classified in two groups were evaluated. Group A consisted of six S2 lines whose parents were yellow corn individuals derived from an inter-population cross between two CIMMYT (International Maize and Wheat Improvement Center) gene pools, pool 5 and pool 6, which segregated black kernels. Pool 5 HLWF (High Land White Flint) and Pool 6 HLWD (High Land White Dent) are populations with a genetic base from highland corn germplasm. Group B had eighteen S2 lines whose parents were collected from El Quinche, Pichincha - Ecuador. The twenty four S2 lines were generated and evaluated in the Experimental station of INIAP-USFQ, in Tumbaco, Pichincha - Ecuador. Two self pollinations were performed to fix the black kernel color and as a source of progenitors for the development of new commercial black corn lines and varieties.

DNA extraction, PCR amplification and Electrophoresis

About 0.25 grams of fresh leaf tissue from corn seedlings were used to extract genomic DNA. The DNA extraction was performed using CTAB, as previously described [20]. A set of 10 primer pairs, one pair per chromosome, were chosen based on their high PIC value described in a previous study [15]. The DNA sequence and genomic position for each primer pair are described in Table 1. A 10 µL PCR reaction mix contained 20 ng template DNA, 0.4 mM dNTPs, 0.2 mM SSR primers (forward and reverse), 1X reaction buffer, 2mM MgCl2, and 1 U Taq polymerase (Invitrogen). The amplification was performed using a Techne TC-142 Thermal Cycler with the following cycling program: initial denaturation at 94 °C for 2 min, followed by 30 cycles consisting of 30 sec at 95 °C, 1 min at X °C (depending on the primer) and 1 min at 72 °C, and a final extension at 72 °C for 5 min. The electrophoresis was conducted at 85W for 2.5 hours on a vertical electrophoresis system (BIORAD) and the products were visualized by silver staining [21].

Molecular data analysis

Data were recorded as an allelic matrix, where molecular weights of each allele (bands) were determined through a comparison with a lineal regression based on the migration distance of a 10 bp molecular weight ladder (Invitrogen). Allele frequency for each locus was calculated using the following formula:

\[
\text{Allele frequency} = \frac{2N_{xx} + N_{xy}}{2N}
\]  
(1)

where Nxx represents homozygous individuals, Nxy are heterozygous individuals, and N is the number of samples. Polymorphism Information Content was determined according to the following formula:

\[
PIC = 1 - \sum f_i^2
\]  
(2)
Figure 1: Allele frequency per locus detected in 24 S2 lines of black corn from the Ecuadorian Andes. Allele 176 of locus Phi059 is by far the most frequent allele. On the other hand, locus Phi23376 has the highest number of alleles.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Genomic position</th>
<th>Sequence</th>
<th>PIC VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>phi 032</td>
<td>9.04</td>
<td>5’-CTCCAGCAAGTGATGCTGGAC-3’</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-GACACCCGGATTATATGGACAC-3’</td>
<td></td>
</tr>
<tr>
<td>phi 059</td>
<td>10.02</td>
<td>5’-AAGCTAATATTAGGGGCTATCCC-3’</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-TCCGATGTACTCGGGCGACTC-3’</td>
<td></td>
</tr>
<tr>
<td>phi 034</td>
<td>7.02</td>
<td>5’-TAGCGACAGTTGCCCTTCTTCT-3’</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-GGGGAGCACGCCTTCTGTTCT-3’</td>
<td></td>
</tr>
<tr>
<td>phi 056</td>
<td>1.01</td>
<td>5’-ACTTTGCTTGGCTGCGATGAC-3’</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-CGCAACACATTCCCCAGAAG-3’</td>
<td></td>
</tr>
<tr>
<td>phi 078</td>
<td>6.05</td>
<td>5’-CAGCACCAAGACTGATGCTGAA-3’</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-GGGGCCGAGGTGATGAGC-3’</td>
<td></td>
</tr>
<tr>
<td>phi 087</td>
<td>5.06</td>
<td>5’-GAGAGGAAGAGTGGTTGTTGACAC-3’</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-ACAACCAGCAGATCAGCAGATGG-3’</td>
<td></td>
</tr>
<tr>
<td>phi 076</td>
<td>4.11</td>
<td>5’-TCCTTCGGGCGCTTTCAATTGCAC-3’</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-GCATCACGACCCCGAGGATCCG-3’</td>
<td></td>
</tr>
<tr>
<td>phi 23376</td>
<td>8.03</td>
<td>5’-CCGGCATGTCATTACCTCC-3’</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-CGAGACAAAGAGAACCTCA-3’</td>
<td></td>
</tr>
<tr>
<td>phi 109642</td>
<td>2.00</td>
<td>5’-CTCCTTCTCTTTCTTGCTAGTCC-3’</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-GAGGCAGGCGAGAGGATCG-3’</td>
<td></td>
</tr>
<tr>
<td>phi 046</td>
<td>3.08</td>
<td>5’-ATCTTGCGACAGTGTGCTGACCTT-3’</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-TCGATCTCTTCCCGAAAACTTGC-3’</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Sequences, genomic position and PIC values of the SSR loci primers used in this investigation.

where $f_i^2$ is the frequency of ith allele. The data were also recorded in a binary matrix which was used to generate a similarity matrix by means of the Jaccard coefficient. The similarity matrix was then used for cluster analysis using the UPGMA clustering algorithm. The genetic associations were also analyzed with a two dimensional Principal Coordinate Analysis. All of the statistical analyses were performed using the NTSYS-pc software [22].

Morphological data analysis

Seven morphologic traits were considered in this study: plant height, ear insertion height, tassel type, stem color, leaf length, leaf width and anthocyanin concentration. The anthocyanin concentration from kernels was determined using a procedure previously described [6]. Kernels were ground to a fine powder and dried in an oven at 70 °C overnight. 100 mg of homogenized powder was then mixed with 7.5 ml 1% acetic acid. Each sample was vortexed for 10 min and sonicated for 30 min at 30 °C. Following sonication, samples were centrifuged (10,000 rpm, 25 min) and the supernatant was discarded. A volume of 1.7 mL of a methanol 1% acetic acid (90/10 v/v) solution was added to the remaining sample. Samples were vortexed for 10 min and sonicated for 30 min at 30 °C. Then, they were adjusted to pH1 and pH5 and the absorbance was measured at a wavelength of 520 nm using a spectrophotometer Spectronic 20D+ (Thermo Scientific). The following formula was used to quantify anthocyanin concentration:

\[
\text{Anthocyanin concentration} = \frac{(\text{Abs pH1} - \text{Abs pH5}) \times \text{DF}}{0.775}
\]
where AbsPHX represents the absorbance measured at an X adjusted pH value, DF is the dilution factor and 0.775 is a correction factor. For morphological quantitative data, mean and standard deviation values were calculated while qualitative data were analyzed by a determination of proportions. In order to verify differences between the mean values from lines of Group A and lines of Group B, t tests were performed. For this purpose, accessions from Group B were set in groups of n=6 to be compared to Group A, also n=6. The tests were performed using Excel 2008 (Microsoft Corporation). Quantitative and qualitative data were used to build a binary matrix, on which a similarity matrix was based, using the Jaccard coefficient. The results were used to obtain a dendrogram by UPGMA cluster analysis using the program NT-SYS 2.0.

Results

Allelic diversity at SSR loci

A total of 43 alleles were revealed for the 10 SSR loci (Fig. 1). The number of alleles per locus ranged from 2 (loci phi059, phi109642 and phi046) to 8 (locus phi059) with an average of 4.3± 2.06. The allele frequency varied from around 2% (allele 255 of locus phi 078), to 81% (allele 176 of locus phi059). The polymorphism information content (PIC) value for the 10 loci ranged from 0.30 (phi059) to 0.83 (phi233376) with an average of 0.59 ±0.15 (Table 1). The average homozygosity level reached by the 24 S2 lines was 82% which coincides with the expected level due to the two cycles of self pollination undergone.

Genetic diversity and cluster analysis

The binary matrix was used to compute the genetic distances for all pairs of corn S2 lines studied. The genetic distances ranged from 0.29 between B1 and B3 lines to 0.95 for lines B8 and B12. Distance measures were later used to construct a hierarchical tree using the UPGMA method. At about 70% of genetic distance six groups are evident in the dendrogram constructed with the molecular data (Fig. 2). Clusters 1, 3 and 5 consist exclusively of corn lines from El Quinche (Group B), whereas the other clusters are constituted by corn lines of both El Quinche and CIMMYT (Groups A and B). Considering the level of association between clusters, cluster 1 is the most distant in the dendrogram, such that it appears as an external group, linked to the rest of clusters at about 80% distance. Clusters 2 and 3 join together at 68% distance and clusters 4 and 6 converge at 75% distance. There is also variation within the clusters, being cluster 1 the most homogeneous since the three corn lines that constitute it converge at 45% distance. The Principal Coordinate Analysis (Fig. 3) distributed the individuals in a similar way as in the dendrogram, obtaining the same major clusters.

Morphological variation

The S2 lines from Group A showed higher mean and standard deviation values compared to the S2 lines from Group B. Both groups evidenced a reduction in the magnitude of the different descriptors compared to the parental
generation. Thus, plant height and ear insertion height showed the highest depression by means of endogamy. Groups B and A showed a reduction of 36.97% and 26.82% of plant height respectively compared to the average for this descriptor in the type of corn “grape bunch”. For ear insertion height, the reduction was 27.94% for Group B, and 20.42% for Group A (data not shown).

The t test was performed to evaluate whether there was a significant difference between Groups A and B regarding morphologic descriptors. No significant differences were found between groups B1 and A for plant height, ear insertion height, and leaf length. Also, groups B2 and A had no significant differences between them in ear insertion height, leaf length and leaf width. On the other hand, Group B3 had significant differences compared to group A in leaf length, leaf width, plant height, and ear insertion height.

Regarding anthocyanin content, an ample variation was observed in the corn S2 lines studied, ranging from 2.3 g/cm3 (line A1) to 0.07 g/cm3 (lines B5 and B7), with an overall average of 0.821 g/cm3. The highest variability corresponded to Group B (from El Quinche - Ecuador), probably because of its bigger sample size. The highest value of anthocyanin content (2.3 g/cm3) however, was seen in accession A1 (from CIMMYT).

Cluster analysis was used to reveal the association between Group A and B. Genetic similarity was calculated from the morphological binary matrix by UPGMA. Six major clusters were observed in the morphological dendrogram (Figure 4). The first main cluster (1) included most of the individuals with leaf length of 46-55 cm and leaf width of 5.16 cm. The second main cluster (2) mainly grouped individuals with plant height of 106-120 cm, ear insertion height of 61-70 cm, and leaf length of 66-75 cm. The third cluster (3) grouped individuals with green stem and leaf length of 36-45 cm. The fourth cluster (4) grouped highly heterogeneous individuals that share primary-secondary-tertiary tassel type. The fifth cluster (5) contained also a heterogeneous group that shared ear insertion height of 71-80 cm. Finally, the sixth cluster (6) had individuals with green stem, primary-secondary tassel type, leaf length of 46-55 cm, and leaf width of 7.1-8 cm. Clusters 1 and 3 contained S2 lines from group B only, whereas the rest of clusters were heterogeneous for both Groups A and B.

**Discussion**

The size of the alleles (in base pairs) found in this study for the 10 SSR loci analyzed corresponded to the sizes that have been reported in the web database “MaizeGDB”
available in http://www.maize.gdb.org. The PIC value is considered as the best method to measure the distribution of alleles among a germplasm [23]. The mean PIC value of 0.59 in this study is similar to those of other studies in corn like Enoki et al [24] that reported a mean PIC value of 0.69, Smith et al [25] 0.62, and Senior et al [26]. 0.59. This PIC value showed the discriminatory power of the primers used in this study to evaluate the diversity between the 24 black corn lines analyzed. Besides, a high PIC value was expected because of the previous high values reported by Warburton et al [15] for the same set of primers. The genetic distance is the measure of the genetic relationship among individuals from a population. The genetic similarity matrix with the Jaccard coefficient had an average similarity value of 0.60±0.14. This value is similar with other studies [25, 26, 27]. In addition, corn is an allogamous specie; therefore, a high allelic diversity is likely to be shown [28]. The results also made possible the distinction between lines closely related like the minimum genetic distance of 0.29 for B1 and B3 and the maximum of 0.95 for lines B8 and B12.

The grouping of the 24 S2 lines shown in the dendrogram and the distribution shown in the PCA displayed several coincidences that suggest that this could be the real distribution between the genotypes of the accessions studied [24, 29]. These groupings would help to select the parental individuals for the following crosses between the lines in the breeding program to generate commercial black corn hybrids, and most importantly, to develop new varieties of black corn.

The dendrogram generated with the morphological traits showed clusters of individuals that share more than one characteristic. Hartings et al [30] also reported that the morphological characterization of 54 individuals, using 20 morphological traits, grouped individuals that shared more than one trait. Corn is a crop with a high diversity in phenotypic traits. However, morphological traits are not accurate to study genetic variability in a group due to the fact that they can be environmentally affected [11, 30].

More variation was observed in the group A even though it’s a small group. This was true for most characteristics, except for anthocyanin content, where group B was more variable. The reason for this might be that Group A is made of individuals from an interpopulation cross; hence it has a wider genetic base. A major genetic variability is expected and available if populations from diverse origins are intercrossed [31]. The corn germplasm in the Ecuadorian Andean region has a low tolerance to endogamy. Studies made in the South American Andean region suggested that the degree of depression by endogamy in some traits was higher in populations with floury kernels than in populations with dent and semi dent kernels [31, 32].

The results obtained in this study demonstrated that the S2 lines derived from the local population Group A and Group B had a reduction in the vigor, due to the effect of the two consecutive self pollinations. The successive self pollination decreases the frequency of heterozygotes, which results in a higher endogamy. Breeding programs that use self-pollination to develop pure
lines describe that heterozygosis is reduced by 50% in each self-pollination cycle [2]. In this study, 82% of homozygosis was found for the 10 SSR loci analyzed in the 24 accessions of black corn, which shows how the heterozygosis was reduced as a result of the two consecutive self pollinations. The main objective of self-pollination is to reduce the heterozygosis [29]. As a result, the lines had reduced the expression of different characteristics. No relationship was found between morphological and molecular distribution in the dendograms. The reason for this might be the limited number of loci and accessions analyzed in this study. Moreover, the loci examined were not necessarily related to the morphological traits. Vaz et al [11] found that traits such as plant height, ear insertion height and type of kernel were not related with the variation detected by 15 microsatellite markers. As in the present study, Vaz et al [11] also used primers based on genomic position and not on primers that targeted genes related to morphological traits [2].

Summarizing, the black corn S2 lines analyzed in this study showed an intermediate level of variability, both morphologically and molecularly, which potentially opens the possibility of finding adequate parents for the developing of new varieties with enhanced characteristics. Especially interesting was the finding of a wide range of variation in anthocyanin content, which eventually will make it possible to combine the genes contributing to this trait in one variety. Increasing the sample size in future studies will contribute to completing the spectrum of genetic variability available in black corn varieties of Ecuador.

Acknowledgements

This study was financially supported through a Small Grant by Universidad San Francisco de Quito. We would like to thank Valeria Ochoa, PhD, for her assistance in the spectrophotometry measurements for anthocyanin content in our samples, as well as Bernardo Gutiérrez for his contributions for the writing of this paper. Also, we would like to thank Viviana Jaramillo for reading this manuscript and the research group of the Plant Biotechnology Laboratory at Universidad San Francisco de Quito.

References


