

# A bayesian approach to inferring the genetic population structure of sugarcane accessions from INTA (Argentina)

Mariana Inés Pocovi<sup>1\*</sup>, and Jorge Alberto Mariotti<sup>1</sup>

Understanding the population structure and genetic diversity in sugarcane (*Saccharum officinarum* L.) accessions from INTA germplasm bank (Argentina) will be of great importance for germplasm collection and breeding improvement as it will identify diverse parental combinations to create segregating progenies with maximum genetic variability for further selection. A Bayesian approach, ordination methods (PCoA, Principal Coordinate Analysis) and clustering analysis (UPGMA, Unweighted Pair Group Method with Arithmetic Mean) were applied to this purpose. Sixty three INTA sugarcane hybrids were genotyped for 107 Simple Sequence Repeat (SSR) and 136 Amplified Fragment Length Polymorphism (AFLP) loci. Given the low probability values found with AFLP for individual assignment (4.7%), microsatellites seemed to perform better (54%) for STRUCTURE analysis that revealed the germplasm to exist in five optimum groups with partly corresponding to their origin. However clusters shown high degree of admixture,  $F_{ST}$  values confirmed the existence of differences among groups. Dissimilarity coefficients ranged from 0.079 to 0.651. PCoA separated sugarcane in groups that did not agree with those identified by STRUCTURE. The clustering including all genotypes neither showed resemblance to populations find by STRUCTURE, but clustering performed considering only individuals displaying a proportional membership  $> 0.6$  in their primary population obtained with STRUCTURE showed close similarities. The Bayesian method indubitably brought more information on cultivar origins than classical PCoA and hierarchical clustering method.

**Key words:** AFLP, bayesian clustering, hierarchical clustering, principal coordinate analysis, *Saccharum officinarum*, SSR.

## INTRODUCTION

Accurate assessment of population structure and genetic diversity levels can be invaluable in crop breeding for different applications. It could provide pivotal information for resource management ensuring a diversified germplasm bank that plays a key role in both breeding and genomic research. The study of genetic diversity of germplasm collections, coupled with genetic differentiation estimates, can facilitate reliable classification of accessions, establishment of their pairwise and group genetic relationships, selection of representative samples which capture genetic diversity of the collection, detection of patterns of differentiation in the whole collection as well as in samples drawn from it and, for breeding purposes, can be particularly useful in planning crosses defining parental lines to be used in the development of new materials.

Population structure analysis has been based in principles underlying Wright's F-statistics. While this approach has been broadly used in the analysis of natural

populations, its application in the analysis of germplasm banks has been limited (Allcochete et al., 2008). In the past, determination of germplasm collections genetic structure has mainly been done using traditional multivariate statistical methods usually based on agronomic data. In recent years, the incorporation of DNA data increased the effectiveness in exploring diversity and provided accurate estimations of genetic relationships. Clustering analysis and multivariate analysis do not assume predefined structures. However, these graphical methods are only loosely connected to statistical procedures allowing the identification of homogeneous clusters of individuals. In recent years, many new methods have been developed especially for studying structure in natural and germplasm collections using molecular markers. One of these is the alternative model-based method developed by Pritchard et al. (2000) which aims at delineating clusters of individuals on the basis of their genotypes at multiple loci using a bayesian approach (Evanno et al., 2005; Odong et al., 2011).

Sugarcane is one of the most important industrial crops in tropical and subtropical regions and is cultivated in more than 90 countries around the world primarily for its ability to store high concentrations of carbohydrates for the production of sugar and biofuel. National Institute of Agricultural Technology (INTA) administrates the main Sugarcane Germplasm Bank in Argentina and, carries out a breeding program.

<sup>1</sup>Universidad Nacional de Salta, Facultad de Ciencias Naturales, Avenida Bolivia 5150, 4400 Salta, Argentina.

\*Corresponding author (fcn13161@gmail.com).

Received: 30 September 2014.

Accepted: 10 March 2015.

doi:10.4067/S0718-58392015000200003

With large number of accessions, genebank managers are faced with numerous choices on how best to conserve these resources and make them available to be utilized in breeding procedures. In the last several years, many efforts have been focused on studying sugarcane genetic diversity on the base of molecular data. Most of the works have been based on various clustering methods. Hierarchical clustering outputs binary trees organize these clusters hierarchically, with the hope that this hierarchy agrees with the intuitive organization of real-world data. These hierarchical structures are also a natural representation for data which was not generated by evolutionary processes, as in the case of materials from a collection of germplasm. There is other limitation to this traditional hierarchical clustering algorithm. The algorithm provides no guide to choosing the “correct” number of clusters or the level at which to prune the tree. Bayesian clustering algorithm overcomes these limitations using marginal likelihoods to decide which clusters to merge (Heller, 2007).

The objective of this paper was to analyze by means of amplified fragment length polymorphism (AFLP) and amplification simple sequence repeat (SSRs) data, the population structure and genetic diversity among sugarcane accessions selected from the INTA's Germplasm Bank (Tucumán, Argentina). In addition, we also compared the performance of hierarchical clustering and ordination techniques with model-based clustering methods.

## MATERIALS AND METHODS

Sixty three sugarcane accessions from the Instituto Nacional de Tecnología Agropecuaria INTA Germplasm Bank (Tucumán, Argentina) were included in this study (Table 1). These genotypes are of interest for breeding purposes in Argentina due to their adaptability to subtropical growing areas (short cycle and early maturity). Some of these materials are or were used as commercial

varieties in Argentina and other countries. Total genomic DNA was extracted from young leaves in these materials, according to Doyle and Doyle (1987).

### SSR amplification and AFLP procedure

Based on the consistency of band patterns obtained in a previous study, thirteen SSR primers were chosen (Table 2). Polymerase chain reactions (PCRs) were performed according to (Pocovi et al., 2013). Each of the amplifications was repeated at least twice by independent PCR to examine the reproducibility and confirm band patterns.

Amplified Fragment Length Polymorphism (AFLP) was carried out according to Pocovi et al. (2008) with 18 primers+3 combinations (M47/P32; M47/P37;M47/P39; M48/P37; M48/P43; M48/P45; M49/P37; M49/P39;M49/P43; M49/P45; M50/P39; M50/P43; M59/P37; M59/P39;M59/P43; M62/P32; M62/P43, and M62/P45).

Amplification Simple Sequence Repeat (SSR) and AFLP products were separated on 6% denaturing polyacrylamide gels for 1.5 h at 60 W using a GibcoBRL Model S2 Sequencing Gel Electrophoresis Apparatus (Life Technologies, Paisley, Scotland). The gels were stained with silver nitrate according to Creste et al. (2010). The resulting banding pattern was scored manually as presence (1) or absence (0) of amplified marker.

### Analysis of genetic structure: Bayesian approach and PCoA

The model-based program STRUCTURE (Pritchard et al., 2000) was used to infer population structure by a bayesian approach using SSR marker dataset and AFLP data. STRUCTURE 2.3.3 (Stanford University, Stanford, California, USA) was run with the admixture model, a burn-in period of 5 000 and 50 000 Markov Chain Monte Carlo. The optimal value of K was identified using the method developed by Evanno et al. (2005). Twenty independent runs were performed for each simulated

**Table 1. Sugarcane varieties included in the genetic variability analysis and country of origin (CO).**

Variety	CO	Variety	CO	Variety	CO
LCP85-384	Louisiana, USA	NA84-3471	Salta, Argentina	TUC72-16	Tucumán, Argentina
LCP86-454	Louisiana, USA	NA63-90	Salta, Argentina	TUC74-6	Tucumán, Argentina
LCP85-376	Louisiana, USA	NA76-128	Salta, Argentina	TUC71-7	Tucumán, Argentina
HoCP85-845	Louisiana, USA	NA73-2596	Salta, Argentina	TUC68-18	Tucumán, Argentina
HoCP92-648	Louisiana, USA	NA88-948	Salta, Argentina	TUC67-24	Tucumán, Argentina
HoCP92-645	Louisiana, USA	NA73-1454	Salta, Argentina	TUC79-9	Tucumán, Argentina
HoCP92-624	Louisiana, USA	CP48-103	Louisiana, USA	TUC77-42	Tucumán, Argentina
HoCP89-888	Louisiana, USA	CP68-350	Louisiana, USA	TUC78-39	Tucumán, Argentina
HoCP91-552	Louisiana, USA	CP70-1133	Louisiana, USA	TUC72-4	Tucumán, Argentina
HoCP92-631	Louisiana, USA	CP79-1380	Louisiana, USA	TUC69-2	Tucumán, Argentina
HoCP91-555	Louisiana, USA	NA84-3471	Salta, Argentina	L91-281	Louisiana, USA
HoCP88-739	Louisiana, USA	CP79-318	Louisiana, USA	RA89-686	Argentina
HoCP90-941	Louisiana, USA	CP65-350	Louisiana, USA	RA87-2	Argentina
US74-1011	USA	CP57-603	Louisiana, USA	RA91-209	Argentina
US74-1015	USA	CP57-614	Louisiana, USA	RA93-154	Argentina
US72-1289	USA	CP72-2086	Louisiana, USA	CP88-1834	Louisiana, USA
L75-33	Louisiana, USA	CP66-346	Louisiana, USA	F98-70	Tucumán, Argentina
TCP81-3067	Tucumán, Argentina	CP62-258	Louisiana, USA	F97-395	Tucumán, Argentina
TCP87-388	Tucumán, Argentina	FAM81-820	Tucumán, Argentina	F97-786	Tucumán, Argentina
NA84-3013	Salta, Argentina	FAM83-11	Tucumán, Argentina	CP65-357	Louisiana, USA
NA78-724	Salta, Argentina	TUC80-7	Tucumán, Argentina	TUC77-42(bis)	Tucumán, Argentina

**Table 2. Simple Sequence Repeat (SSR) primers used for genotyping 63 sugarcane accessions from the INTA Sugarcane Germplasm Bank (Tucumán, Argentina).**

SSR	Repeat motif	Size range (bp)	Annealing temperature (°C)	Forward Primer sequence (5' to 3')
				Reverse Primer sequence (5' to 3')
NKS26	(TG) <sub>18</sub>	194-164	54	GTT CTC GAC ATG GGC CTA CT CTG CAC TTT CGG TCC TTT TT
mSSCIR19	(GA) <sub>23</sub>	130-160	48	GGT TCC AAA ATA CAC AAA CAA TCT TAT CTA CGC ACT T
NKS38	(AG) <sub>15</sub>	92-292	55	TGA ACT CGG CAA CAG TTT TT CCC ACC AAG TCG TTC TGA AT
NKS 23	(GA) <sub>18</sub>	113-498	54	TAA ACC CCC GAA AAA GAA CC TCC GGA GGT AGA TCC ATT TG
NKS34	(GT) <sub>18</sub> (A) <sub>31</sub>	131-214	58	CGT CTT GTG GAT TGG ATT GG TGG ATT GCT CAG GTG TTT CA
mSSCIR16	(GA) <sub>18</sub>	130-300	54	TGG GGA GGG CTG ACT AGA GGC GGT ATA TAT GCT GTG
SMC703BS	(CA) <sub>12</sub>	186-229	62	GCC TTT CTC CAA ACC AAT TAG T GTT GTT TAT GGA ATG GTG AGG A
mSSCIR3	(GT) <sub>28</sub>	171-187	60	AAT GCT CCC ACA CCA AAT GC GGA CTA CTC CAC AAT GAT GC
mSSCIR18	(GA) <sub>23</sub>	170-200	52	GGG TGT TCT GTT GAG CA GAG GTA GGA GGG AGT GTT
SMC766BS	(CA) <sub>20</sub> (GA) <sub>16</sub>	170-270	60	TTA CTC GGC TGG GTT TTG TTC TAA GAA TCG TTC GCT CCA GC
SMC7CUQ	(CA) <sub>10</sub> (C) <sub>4</sub>	160-170	60	GCC AAA GCA AGG GTC ACT AGA AGC TCT ATC AGT TGA AAC CGA
mSSCIR78	(GTT) <sub>6</sub>	150-310	48	TGCCTTAAC CGT GAC ATC GAGGACGAGGAGCAGAA
mSSCIR34	(GA)	130-300	56	ATCGCCTCCACTAAATAAT TTGCTTTGCTTCCTCCTC

value of K, ranging from 1 to 10. Subsequently, a Pr(X|K) index with respect to each K was used to calculate  $\Delta K$  using the formula described by Evanno et al. (2005). The optimal K depends on the highest peak of  $\Delta K = IL''(K)/s[Pr(x|k)]$ , where  $(IL''(K))$  denotes the absolute value of the second order rate of change of Pr(X|K), and  $s[Pr(x|k)]$  the standard deviation of the Pr(X|K)). Therefore, we compared clustering at the optimum K value with those from Principal Coordinate Analysis (PCoA) and classification Unweighted Pair Group Method with Arithmetic Mean (UPGMA), and Ward to look for the best match.

Principal Coordinate Analysis was also carried out based on SSR genetic dissimilarities matrix for examining population structure. Sugarcane accessions were plotted in a bidimensional space using the INFOSTAT vs2013p (Di Rienzo et al., 2013).

### Classification

Genetic dissimilarities among all possible pairs of sugarcane accessions were calculated from SSR data as  $d = \sqrt{1 - s_{ij}}$  where  $s_{ij}$  corresponds to the Jaccard similarity coefficient. The resulting similarity matrix was subjected to cluster analysis by two different algorithms UPGMA and Ward method that employs ANOVA approach for calculating the distances between clusters. The goodness of fit of the clustering to the data matrix was determined by calculating the cophenetic correlation coefficient between the similarity matrix and the cophenetic matrix, and the reliability of the dendrogram was tested by bootstrap analysis with 100 replicates using DARwin 5.0.158 software (Centre de Coopération Internationale

en Recherche Agronomique pour le Développement (CIRAD), Paris, France) (Perrier and Jacquemoud-Collet, 2006). For each analysis, the relationships between all pairs of genotypes were visualized as dendrograms.

The results from hierarchical cluster analysis were also compared with the results from STRUCTURE with regard to cluster composition and appropriate number of clusters.

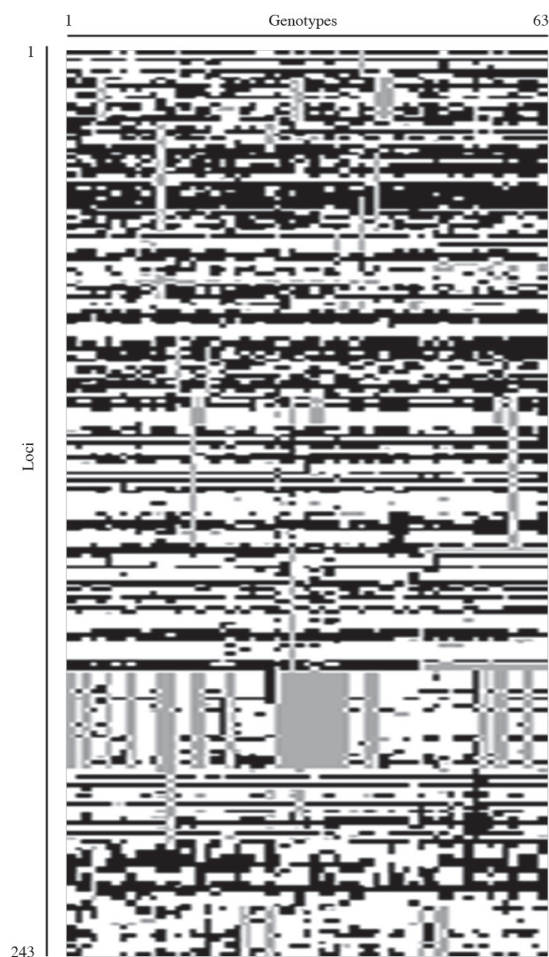
## RESULTS AND DISCUSSION

107 SSR and 136 AFLP loci were utilized to assess genetic population structure and genetic diversity among 63 sugarcane accessions. Figure 1 shows the fingerprint of the 243 loci. The illustration offers a graphic visualization of the variability estimated among the 63 sugarcane genotypes included in the study and provides a unique genetic “bar-code” for each accession.

### Sugarcane population structure inferred from SSR data: Bayesian approach

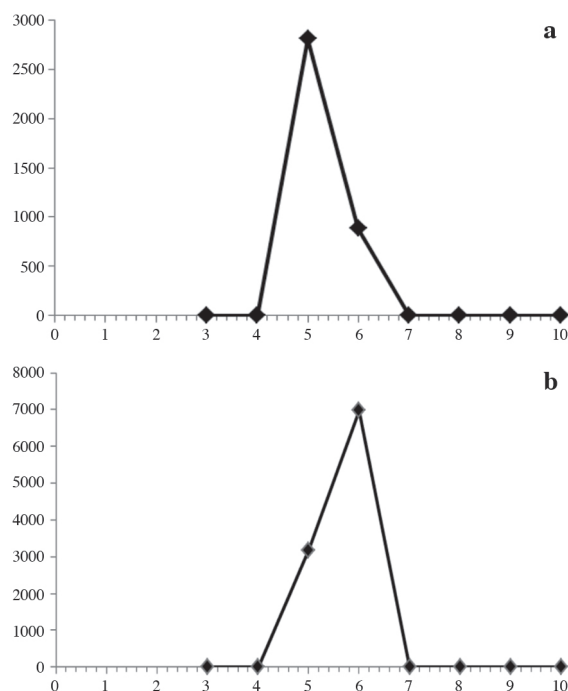
A large number of methods have been proposed to deal with the optimum number-of-clusters problem. In this paper, the modal value of the  $\Delta K$  distribution indicated that the true K value or the uppermost level of structure was five (Figure 2a).

Clearly, each of the five optimal clusters has a considerable proportion of mixed memberships sharing among clusters. The bayesian approach can be considered a quantitative clustering method as computes the proportion of the genome of an individual originating from each inferred population. From the total of 63



**Figure 1.** Simple Sequence Repeat (SSR) and Amplified Fragment Length Polymorphism (AFLP) fingerprints of sugarcane accessions from INTA (Instituto de Tecnología Agropecuaria, Argentina) Germplasm Bank. Shared absent bands (0) are shown in black; white color indicates shared bands, (1) and grey squares indicate missing data.

sugarcane accessions investigated, 34 (54%) have more than 0.60 membership in any given of the five clusters. The remaining 46% (29 accessions) share similar membership coefficients for at least two groups indicating a high degree of admixture. However, the mean value of alpha resulted 0.1772. According to Evanno et al. (2005), when alpha is close to zero, most individuals are essentially from one population or another, while  $\alpha > 1$  means that most individuals are admixed. The same concept was proposed by Ostrowski et al. (2006) who indicated that a relatively small value of the alpha parameter indicated that most accessions originated from one primary ancestor, with a few admixed individuals. In our case, few parents have been used at the beginning of the breeding program in the search of adaptability to subtropical climates (short cycle and early maturity). For this reason most sugarcane accessions share common ancestors.



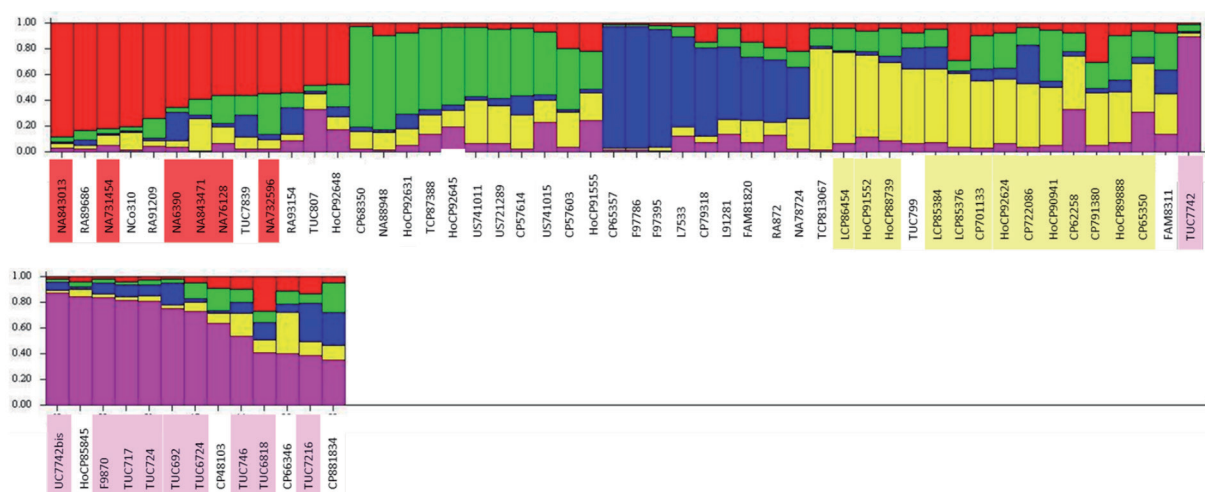
**Figure 2.** Graphical method for the detection of the true number of K groups using Single Sequence Repeat (SSR) and Amplified Fragment Length Polymorphism (AFLP) data (b).  $\Delta K$  calculated as  $\Delta K = |L'(K)|/SD$ . In a the highest peak was identified at  $K = 5$ ; in b the peak was identified in  $K = 6$ .

Cluster IV (in yellow) was the group with the largest number of sugarcane accessions (16), but having the higher percentage of accessions with less than 0.60 membership in that given cluster (25%) indicating that these accessions share similar membership coefficients for any given cluster. Clusters I (in red), II (in green), III (in blue) and V (in pink) included 13, 11, 9, and 14 accessions with 46%, 45%, 56%, and 65% with more than 0.6 membership in each given cluster respectively (Figure 3).

Mean  $F_{ST}$  values confirmed the existence of differences among clusters. Cluster 3 showed the highest  $F_{ST}$  value (0.63), followed by cluster 5 (0.38) and cluster 1 (0.29). It is generally accepted that  $F_{ST}$  values under 0.05 indicate negligible genetic differentiation while those over 0.25 indicate a great deal of genetic differentiation (Zhao et al., 2010).

Although the origin for each individual was not used in the clustering algorithm, the classification of clusters appeared moderately correlated with the origin of sugarcane accessions (Table 1, Figure 3). Twenty nine from the total of genotypes clustered according to their origins, while the remnant showed a partial or predominant membership of populations of other origins. Clusters I, IV, and V grouped 6 (46%), 13 (82%), and 10 (72%) accessions respectively who share their origins in each group (Salta, Louisiana, and Tucumán) (Figure 3).





**Figure 3. Structure of sugarcane accessions based on Bayesian assignment probabilities. Structure inferred on Single Simple Repeat (SSR) data. Each sugarcane accession is represented by a vertical line, which is partitioned into K colored segments that represent the individual's estimated membership fractions in K cluster (K = 5). Individuals have been sorted by decreasing membership in the cluster with highest assignment probability for that population.**

It can be seen that SSR were quite efficient in detecting sugarcane accessions from Louisiana and Tucumán. Relatedness among samples influences the ability of STRUCTURE to correctly detect the genetic stratification of the remaining materials. In this sense, Clusters II included mixed genotypes in relation to its origin and finally, cluster III grouped sugarcane accessions mainly from Tucumán and Salta.

### Sugarcane population structure inferred from AFLP data: Bayesian approach

The  $\Delta K$  criterion suggested gave the highest value at K = 6 for AFLP data. However, individual assignments revealed that only three accessions (4.7%) have more than 0.60 memberships in any given of the six clusters. In this regard, microsatellites seemed to perform better (54%) than AFLP data set. Differences in assignment percentages between diverse molecular data were also reported by independent studies in various germplasm materials and were attributed to different information content (Emanuelli et al., 2013).

Given the low probability values for individual assignment found with AFLP, the structure analysis was left aside in that case, and only STRUCTURE results based on SSR were compared with PCoA and clustering analyses.

### Sugarcane population structure: PCoA

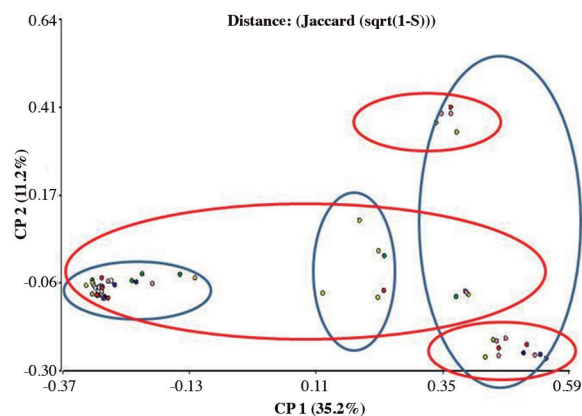
The genetic relationship among the 63 sugarcane accessions was assessed by PCoA based on genetic dissimilarities. The first two principal coordinates explained 35% and 11% of the variability respectively. There are several approaches for determining the number of components to interpret from PCoA. According to Cliff (Franco and Hidalgo, 2003), there should be considered

acceptable those coordinates whose accumulated values account for 70% or more of the total variance. In our case, the first seven coordinates accounted for 71% of the total variance. It is common that PCoA results are interpreted by simple visual inspection of plots of points along the first two or three axes. Reeves and Richards (2009) noted two main problems in this practice. First, the practice is highly subjective. Patterns can be deceiving, especially when plots are enhanced with additional visual information such as outlines, or when the point marker is varied according to *a priori* ideas of population structure. Second, the first two or three axes normally may explain only a small proportion of the total variation in a data set. More principal coordinate axes would be needed to meet the 70% threshold. Therefore, any method for interpreting ordination analyses of multilocus data should be capable of considering many or all axes simultaneously, and the visual inspection method is inadequate in this respect.

The first principal coordinate (CP1) separated individuals in three groups colored in blue (Figure 4). The second principal coordinate (CP2) separated most sugarcane accessions in other three different groups (in red). In neither case groups agreed with those identified by STRUCTURE. According to Mohammadi and Prasanna (2003), cluster analysis proved to be more sensitive for detecting pedigree relationships among genotypes than PCoA when the first two or three PCos explained < 25% of the total variation.

### Clustering inferred from SSR data

Histogram of pairwise dissimilarity from the SSR data indicates a normal distribution with a mean of 0.45 (Figure 5). The dissimilarity coefficients ranged from 0.079 to 0.651. The majority of the dissimilarity coefficients were observed between 0.5 and 0.6. The most of the SSR-based



**Figure 4.** Scatter plot of the first and second principal coordinates of the 63 sugarcane accessions. CP1 and CP2 refer to the first and second principal coordinates. The numbers in parentheses refer to the proportion of variance explained by the principal coordinates. Blue and red circles indicate different groups in CP1 and CP2 respectively. Genotype colors (yellow, red, pink and blue) correspond to those assigned by STRUCTURE.

pairwise comparisons exhibited genetic dissimilarities higher than 0.43.

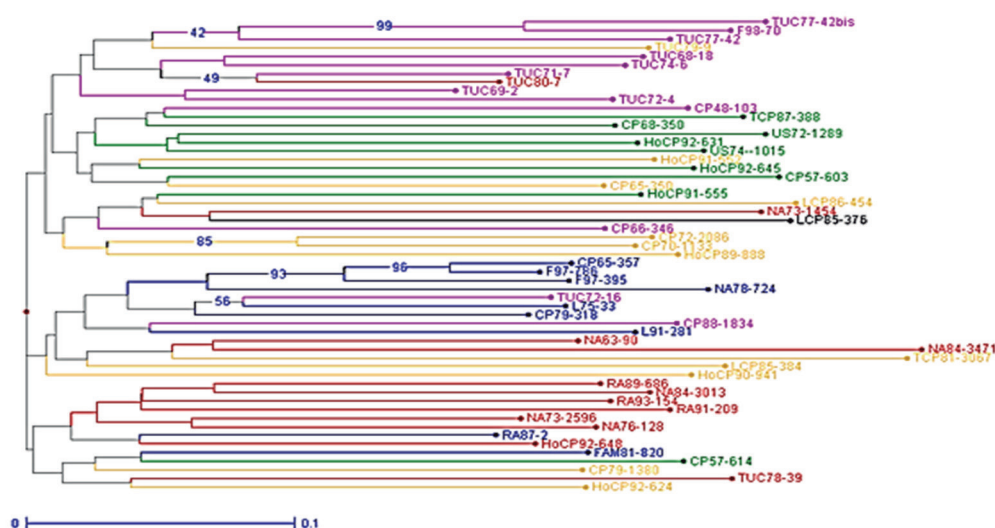
Sugarcane accessions grouped on the right tail with higher genetic distances should be considered as potential parents when planning crosses in the development of new materials. You et al. (2013) reported that the innovation of parents with higher genetic diversity has showed a positive role in sugarcane breeding programs in China broadening of the genetic basis in sugarcane hybridization.

Initially, the two most popular clustering methods used in the literature for determination of the structure of plant germplasm collections, UPGMA and Ward, were applied for constructing trees. The differences between both hierarchical clustering algorithms lie mainly in

how the distances between pairs of OTUs or clusters are defined (Odong et al., 2011). In both analyses, trees were constructed with all the genotypes with a minimal 70% of valid data (no missing data) for each unit pair (Figures 5 and 6, respectively).

Some authors agree that UPGMA produces highly unbalanced clusters, with a main cluster and many small clusters, whereas the Ward method built clusters of similar size. In our case, clear differences were evident within UPGMA and Ward clusters. Cutting of UPGMA tree resulted into three slightly unbalanced clusters, while Ward produced two major balanced clusters (Figure 7). Highest CPCC was obtained using UPGMA algorithm (0.76 vs. 0.49 for Ward), which indicated a good fit between the original pair wise distance among accessions and pair wise distances between accessions predicted constructing the dendrogram. Odong et al. (2011) showed that the cophenetic correlation coefficient is directly related to subgroup differentiation and can thus be used as an indicator of the presence of genetically distinct subgroups in germplasm collections. Given that UPGMA performed better than Ward method, the following discussion refers only to the UPGMA.

The clustering by UPGMA including all genotypes showed no resemblance to populations found by STRUCTURE. A second clustering was performed considering only individuals displaying a proportional membership > 0.6 in their primary population obtained with STRUCTURE. While this new UPGMA tree was also unable to identify the optimum number of clusters obtained by STRUCTURE (five), it showed close similarities (Figures 7 and 4a). Although analysis by STRUCTURE showed that five is the optimum number of populations, in Figure 6 it can be seen that cluster I from UPGMA contains 61% of the genotypes included



**Figure 5.** UPGMA dendrogram (Jaccard ( $\sqrt{1 - s}$ )). Tree was constructed with all the genotypes with a minimal 70% of valid data (no missing data) for each unit pair. The colors of the sugarcane accessions correspond to the color assigned according to the population to which it belongs in the STRUCTURE analysis. Numbers indicate the limit of 40% statistical support for the topology at a particular node.



## CONCLUSIONS

The comprehensive molecular characterization of the sugarcane accessions contributes to the knowledge about levels and distribution of genetic diversity in the existing INTA sugarcane germplasm bank. Different softwares and methods used for genetic analysis were suitable for confirming a narrow genetic base of the materials; however, there were differences in the subpopulations obtained depending on methods applied. The relatively low genetic variability can partially be explained by the closeness of the subtropical materials used in the study, according to their proposed genealogies and the breeding procedures applied ("modified recurrent selection").

Comparison of the three methods showed that accessions clustered differently based on the method of analysis STRUCTURE, principal coordinates, classification. The Bayesian method used indubitably brought more information on cultivar origins than classical Principal Coordinate Analysis and hierarchical clustering method. The results provide guidance for future efficient use of sugarcane accession in breeding and germplasm bank management. An increase in the genetic diversity of parental accessions should be helpful to broaden the genetic basis of progenies subjected to selection.

## LITERATURE CITED

- Allcochete, A.A.N., P.H.N. Rangel, and M.E. Ferreira. 2008. Genetic structure of rice samples from a germplasm bank. *Scientific Research and Essay* 3(12):577-598.
- Creste, S., A.G. Klaus, L. Accoroni Pint, R. Vencovsky, M. Gimenes, X. Mauro, et al. 2010. Genetic variability among sugarcane genotypes based on polymorphisms in sucrose metabolism and drought tolerance genes. *Euphytica* 172:435-446.
- Di Rienzo, J.A., F. Casanoves, M.G. Balzarini, L. González, M. Tablada, and C.W. Robledo. 2013. InfoStat versión 2013. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. Available at <http://www.infostat.com.ar> (accessed June 2014).
- Doyle, J.J., and L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19:11-15.
- Emanuelli, F., S. Lorenzi, L. Grzeskowiak, V. Catalano, M. Stefanini, M. Troggio, et al. 2013. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biology* 13:39.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14:2611-2620.
- Franco, F.L., y R. Hidalgo (eds.) 2003. Análisis estadístico de datos de caracterización morfológica de recursos fitogenéticos. Boletín Técnico N° 8. 89 p. Instituto Internacional de Recursos Fitogenéticos (IPGRI), Cali, Colombia.
- Heller, K. 2007. Efficient bayesian methods for clustering. 129 p. PhD thesis. Gatsby Computational Neuroscience Unit, University College, London, UK.
- Mohammadi, S.A., and B.M. Prasanna 2003. Analysis of genetic diversity in crop plants. Salient statistical tools and considerations. *Crop Science* 43:1235-1248.
- Odong, T.L., J. Van Heerwaarden, J. Janse, T.J.L. Van Hintum, and F.A. Vaneeuwijk. 2011. Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data? *Theoretical and Applied Genetics* 123:195-205.
- Ostrowski, M.F., J. David, S. Santoni, H. Mckhann, X. Reboud, V. Le Corre, et al. 2006. Evidence for a large-scale population structure among accessions of *Arabidopsis thaliana*: possible causes and consequences for the distribution of linkage disequilibrium. *Molecular Ecology* 15:1507-1517.
- Perrier, X., and J.P. Jacquemoud-Collet. 2006. DARwin software. Available at <http://darwin.cirad.fr/darwin> (accessed June 2014).
- Pocovi, M., G. Collavino, F. Locatelli, G. Pacheco, D. Díaz, R. Ríos, et al. 2008. Assessing genetic variability of subtropical hybrid sugar cane (*Saccharum* spp.) materials using isozymes and AFLP. *Sugar Cane International* 26(1):3-8.
- Pocovi, M., G. Taboada, G. Collavino, A. Gutiérrez, and J. Mariotti. 2013. Optimal use of SSRs for establishing genetic relationships and variety identification in a collection of sugarcane hybrids. *Indian Journal of Plant Genetic Resources* 26:103-112.
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Reeves, P.A., and C.M. Richards. 2009. Accurate inference of subtle population structure (and other genetic discontinuities) using principal coordinates. *PLoS One* 4(1):1-11. doi:10.1371/journal.pone.0004269.
- You, Q., L. Xu, Y. Zheng, and Y. Que. 2013. Genetic diversity analysis of sugarcane parents in Chinese breeding programmes using SSR markers. *The Scientific World Journal*. doi:10.1155/2013/613062.
- Zhao, Z.H., J.Z. Xi, Q. Jia, S.F. Li, and H.Y. Huang. 2010. Analysis of genetic structure and diversity of Chai chicken breed using microsatellite markers. *Journal of Animal and Veterinary Advances* 9:1197-1200.