

# Molecular Diversity in Selected Banana Clones (*Musa AAA* “Cavendish”) Adapted to the Subtropical Environment of Formosa Province (Argentina)

José Luis Ermini<sup>1</sup>, Gerardo Carlos Tenaglia<sup>2</sup>, Guillermo Raúl Pratta<sup>1,3</sup>

<sup>1</sup>IICAR, UNR, CONICET (Instituto de Investigaciones en Ciencias Agrarias de Rosario, Universidad Nacional de Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas), Facultad de Ciencias Agrarias, Campo Experimental “J. F. Villarino”, Zavalla, Argentina

<sup>2</sup>Instituto de Investigación y Desarrollo Tecnológico para la Agricultura Familiar Región Nordeste Argentino, Laguna Naineck, Provincia de Formosa (IPAF NEA), Argentina

<sup>3</sup>Cátedra de Genética, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Zavalla, Provincia de Santa Fe, Argentina

Email: joseluis.ermini@unr.edu.ar, pratta@iicar-conicet.gob.ar, tenaglia.gerardo@inta.gob.ar

**How to cite this paper:** Ermini, J.L., Tenaglia, G.C. and Pratta, G.R. (2018) Molecular Diversity in Selected Banana Clones (*Musa AAA* “Cavendish”) Adapted to the Subtropical Environment of Formosa Province (Argentina). *American Journal of Plant Sciences*, 9, 2504-2513.

<https://doi.org/10.4236/ajps.2018.912181>

**Received:** October 29, 2018

**Accepted:** November 27, 2018

**Published:** November 30, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Banana production dates from the 1920s in Argentina. Rhizomes were introduced and dispersed by immigrants from bordering countries in the Northern provinces. There is scarce information on its genetic diversity to assist in crop breeding programs; hence studies of genetic structure between populations and individuals are fundamental for future use. Molecular markers assess the diversity of the crop. This study employed Amplified Fragment Length Polymorphism to investigate the genetic variation in local banana plants from farmer's fields. Forty-four rhizomes (selected from a total of 860 plants for being stable for production) were used as plant material and 6 primer combinations selected in a previous report were used. Polymorphic fragments present in a given genotype were assigned 1 and those absent were assigned 0. The matrix generated was analyzed by univariate and multivariate analysis. A total 540 bands were scored, of which 100% were polymorphic. The number of the pattern duplicated bands was 23 and twenty-one amplicons were exclusive to banana plants collected in the same farmer's field. Hierarchical clustering and principal coordinates analyses showed differences between genotypes. The cophenetic correlation of the cluster was 0.63 while the 18% of the total molecular variation was explained by the two principal coordinates. These analyses evidence genetic diversity of the crop according to farmer's field. In agreement to this observation, analysis of molecular va-

riance revealed that 8.9% of the variation was found among farmer populations and 91.1% within farmer populations. In conclusion, introduction of banana into the subtropical zone was associated with a broader genetic variation in order to increase the genetic homeostasis necessary for adapting the crop to the suboptimal environment of Northeastern Argentinean Formosa Province.

## Keywords

Agronomy Biodiversity, Autotriploidy, Molecular Characterization, Multivariate Analysis, Plant Genetic Resources

---

## 1. Introduction

Banana (*Musa* genus) is one of the most popular fruits all over the world. It is the first fruit crop in production and the second, after the citric family, in economic value. As a polyploid species, it presents a high level of sterility, being mainly multiplied by asexual propagation. Hence genetic variability is assumed to be scarce among commercial varieties [1].

In Argentina, a country that is mainly recognized as world provider of cereals and oil crops, banana production is an important regional activity in Formosa and other northern provinces [2]. Though it is a typical tropical crop, Formosa province—located in the Argentinean subtropical northeaster region—has promoted banana production since the 60', which acquired a great local economic impact especially for familiar and small farmers [3]. Presently, approximately 1500 hectares are produced in the province of Formosa by about 460 farmers [4]. Adaption of the crop to this marginal region was achieved mainly by natural selection for abiotic and biotic stresses but also by artificial selection intuitively practiced by farmers for fruit quality and other agronomical desired traits in their productive environment. In consequence, a wide biodiversity whose molecular basis is completely unknown has been originated in Formosa. This genetic variation largely generated by spontaneous mutation is essential for present and future banana breeding, biotechnology and industry demands.

Many authors [1] [5] [6] have previously reported the diversity and relationships among *Musa* genus by different molecular markers. The studies of genetic structure between populations and individuals are fundamental for any future investigations. Molecular markers are suitable to assess the diversity of the crop [7]. AFLP analysis is an interesting approach due to the generation of a high number of selectively amplified DNA bands that cover a large proportion of the genome under study [5]. The AFLP technique has advantages over other molecular based techniques for DNA fingerprinting including Restriction Fragment Length Polymorphism (RFLP) and Random Amplified Polymorphic DNA (RAPD). For example, AFLP technique can be used for DNA samples of any origin or complexity, and small sequence variations can be detected using only

small quantities of genomic DNA (0.05 - 0.10 µg). The capacity to generate many polymorphic bands in one lane is its major advantage because a greater number of *loci* may be simultaneously revealed by AFLP than by other currently available PCR-based techniques. Therefore, polymorphism detected per reaction is much higher. AFLP is superior in terms of the number of sequences amplified per reaction and their reproducibility. The markers produced are reliable and reproducible within and between laboratories, and are relatively easy and inexpensive to generate [6].

The general objective of this work was to apply AFLP technique to assess the genetic diversity in 44 banana clones collected in Formosa from different farmers' fields. These clones are being currently evaluated in a same assay in order to compare their agronomic aptitude under the same environment and to select the most promissory ones to conform a local variety adapted to the Argentinean northeastern crop conditions. This molecular characterization is necessary to objectively know the available genetic variability in the population formed by the clones selected at Formosa farms as well as the proportion of that genetic variability remaining among the clones that will conform the local variety.

## 2. Materials and Methods

**Study sites:** Banana clones were grown in the northeastern of Argentina, a productive zone located at 24°59'S, 58°51'W in Formosa province. The rhizomes were transplanted in the experimental field station "José F. Villarino", located at 33°01'53.4"S and 60°53'24.5"W in Santa Fe Province exclusively for DNA extraction.

**Plant material:** The banana hybrid used in this study is the autotriploid *Musa* AAA (subgroup Cavendish) [2]. The plant material was collected from different fields of agricultural producers. Around 860 plants were evaluated for their stable annual production (one harvest per year) in three consecutive years. At the end, exactly 44 rhizomes were used successfully from the high-yielding group. Plant material was codified as below: BCC, clones obtained from Pablo Bondaruck's number one field; BARB, clones obtained from Pablo Bondaruck's number two field; SBFA, clones obtained from Felipe Sanchez's number one field; SBFB, clones obtained from Felipe Sanchez's number two field; OMP, clones obtained from Omar Chaparro's field; OSP, clones obtained from Oscar Chaparro's field; TGE, check variety Gran Enano; and TGAL, check variety Gal.

**DNA leaf extraction:** A piece of young leaves were collected and conserved at -80°C. Genomic DNA was extracted from 60 mg of the leaves by commercial kit (Wizard<sup>®</sup> Genomic DNA Purification Kit, Promega).

**Genomic DNA quantification:** The extracted DNA was revealed on 1% (wt/vol) agarose gel in TBE running buffer at 80 Volts for 15 - 20 min. Visualization was achieved by using SYBR Safe DNA stain (<sup>®</sup>Invitrogen SRL). The quality and quantity were evaluated in comparison to λ phage (<sup>®</sup>Biodynamics SRL).

**AFLP protocol:** The genomic DNA (30  $\mu$ L, containing approximately 300 ng) was digested with 6 units of EcoRI and 6 units of MseI in a final volume of 40  $\mu$ L incubated at 37°C for 2 h. Ligation of the digested fragments to the specific universal EcoRI and MseI adapters (EcoRI adapter, final concentration 0.25  $\mu$ M and MseI adapter, final concentration 0.25  $\mu$ M) was achieved by incubation at 37°C for 2 h with 3.6 units of T4 ligase in a final volume of 10  $\mu$ L. The preamplification samples were prepared using 1  $\mu$ L of the digested-ligated DNA plus 75 ng of each primer + 1 (EcoRI + 1: 5'-GACTGCGTACCAATTCA-3' and MseI + 1: 5'-GATGAGTCCTGAGTAAC-3'), 0.2 mM dNTPs and 1 unit of Taq polymerase in a final volume of 25  $\mu$ L. The PCR conditions were 30 cycles of 30 s at 94°C, 1 min at 56°C and 1 min at 72°C each one. Once preamplified, the solution was 1:10 diluted in sterile water. The selective amplification samples were prepared using 3  $\mu$ L of this dilution and the same mix used for preamplification, though six primer + 3 combinations reported by [8] were used in this step (Table 1). Also, a touchdown PCR was programmed, starting with a cycle of 30 s at 94°C, 30 s at 65°C and 1 min at 72°C. During the next 10 cycles, the annealing temperature was reduced to 1°C per cycle, until reaching 56°C. Then, 23 cycles of 30 s at 94°C, 1 min at 56°C and 1 min at 72°C each were repeated. The amplified fragments were boiled for 5 min and loaded into a 6% poly-acrylamide denaturing gel. The molecular weight marker Cincuenta Marker (<sup>®</sup>Biodynamics SRL) was loaded in each gel, electrophoretic separation was done for 2.5 hours at 1900 Volt. Then, the gel was stained with the silver nitrate technique, scanned and analyzed with GelPro<sup>®</sup> software.

**Measurement of genetic diversity:** The polymorphism was expressed as the presence or absence of a given amplified fragment (amplicon). Polymorphic fragments present in a given genotype were assigned 1 and those absent were assigned 0 for each locus.

**Data analysis:** Two different approaches exist to extract statistical information from AFLP data. The first one, at the individual level, we analyzed the pattern of presence or absence for each band and compared all clones pattern to calculate unique amplicons for farmers field (total unique markers). The second

**Table 1.** Combinations of EcoRI and MseI primer + 3 selected by [8] and used for selective amplification in this experiment.

Combination code	EcoRI Primer Sequence (E)	MseI Primer Sequence (M)
E38/M32	5'-GACTGCGTACCAATTCAC-3'	5'-GATGAGTCCTGAGTAACAG-3'
E39/M32	5'-GACTGCGTACCAATTCAGA-3'	5'-GATGAGTCCTGAGTAACAG-3'
E39/M34	5'-GACTGCGTACCAATTCAGA-3'	5'-GATGAGTCCTGAGTAACTA-3'
E40/M45	5'-GACTGCGTACCAATTCAGC-3'	5'-GATGAGTCCTGAGTAACTG-3'
E40/M36	5'-GACTGCGTACCAATTCAGC-3'	5'-GATGAGTCCTGAGTAACGT-3'
E/45/M35	5'-GACTGCGTACCAATTCATG-3'	5'-GATGAGTCCTGAGTAACTG-3'

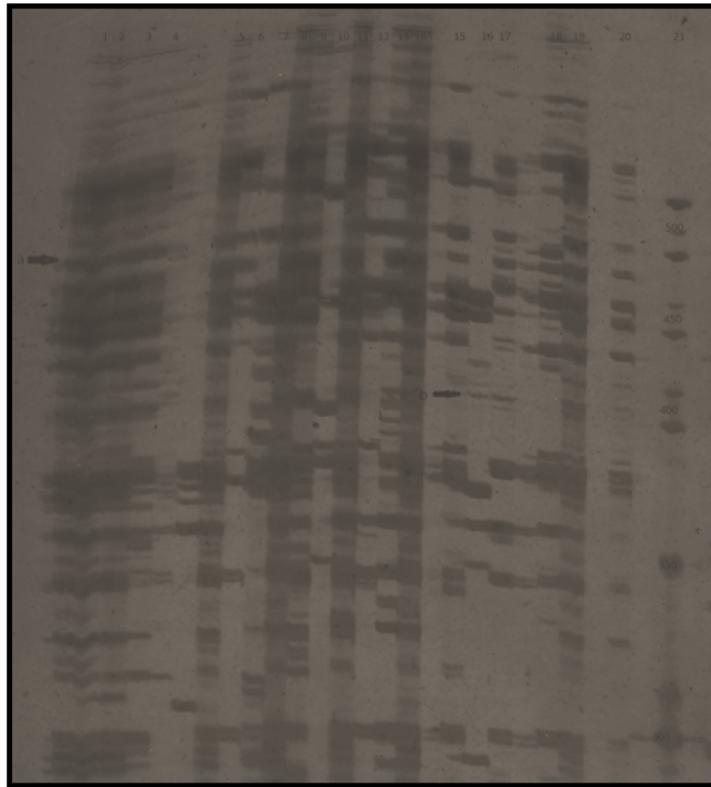
one, population orientated strategy, adopted the allele frequency-based approach, which involves estimating allele frequencies at each AFLP locus [9]. Raw data obtained from the molecular marker were analyzed using INFOGEN<sup>®</sup> software [10] to calculate: total number of amplified fragments, number of polymorphic fragments, number of equal polymorphism pattern, and polymorphism percentage were calculated. The equal polymorphism pattern shows the similarity between some AFLP profiles. In contrast, the unique amplicons show that there are singular bands. The analyses of each data results from different pattern of primer combination were achieved by multivariate analysis. The Dice coefficient was calculated for estimating genetic similarities among clones, and the Ward's method was used for clustering [11].

The Metric multidimensional scaling Principal Coordinate Analysis (PCoA) and the Analysis of Molecular Variance (AMOVA) were applied to the data to make inferences about the diversity. We applied a one-way AMOVA to compare demes (group of plants recollected at a same farmer's field) composing. In this case,  $y_{ki}$  represents the  $i_{th}$  haplotype of the  $k_{th}$  deme in population. A linear model similar to that seen in conventional analysis of variance (ANOVA):  $y_{ki} = p + a_k + w_{ki}$ , where  $p$  is the vector of the expected allelic states. The effects are here  $a$  for the deme, and  $w$  for the individual, and, as usual, the effects are assumed random, additive, and independent, with associated variance components among groups ( $\sigma_a^2$ ) and between groups ( $\sigma_b^2$ ) [12]. The components of variance partitioned within and among the farmer's field populations were estimated from a Dice distance matrix using INFOGEN<sup>®</sup> software.

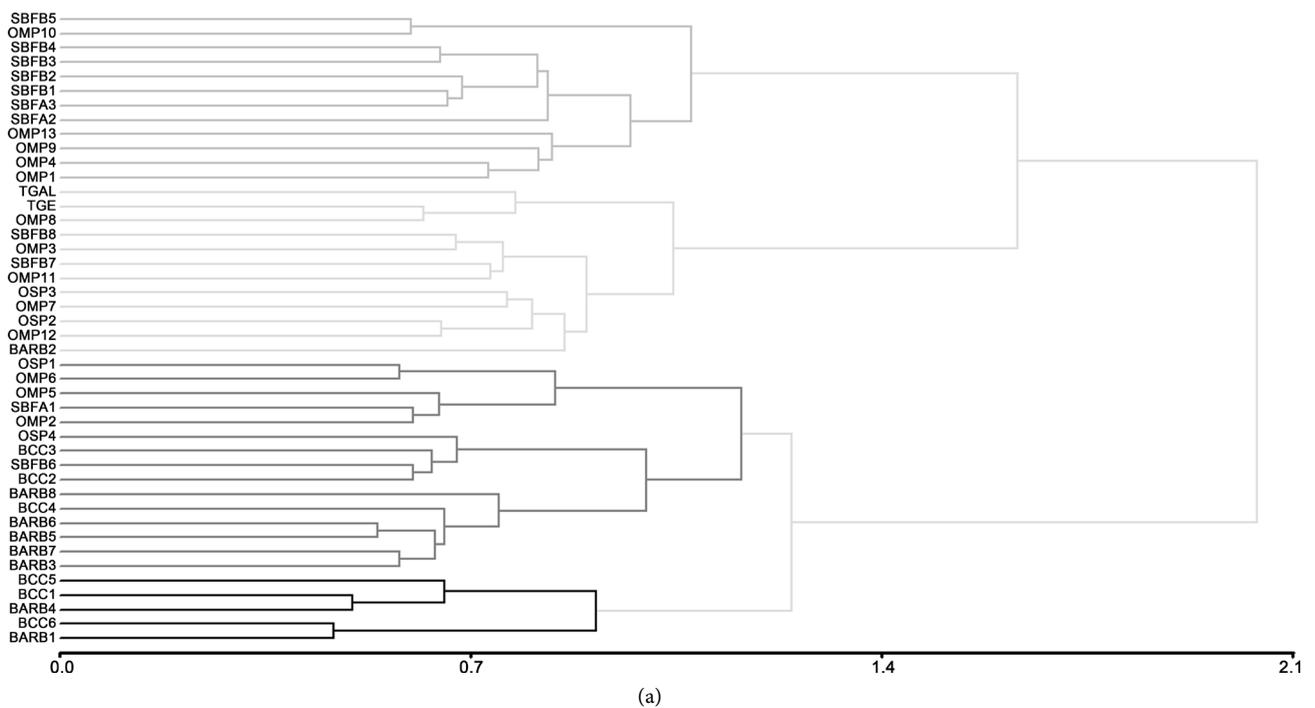
### 3. Results

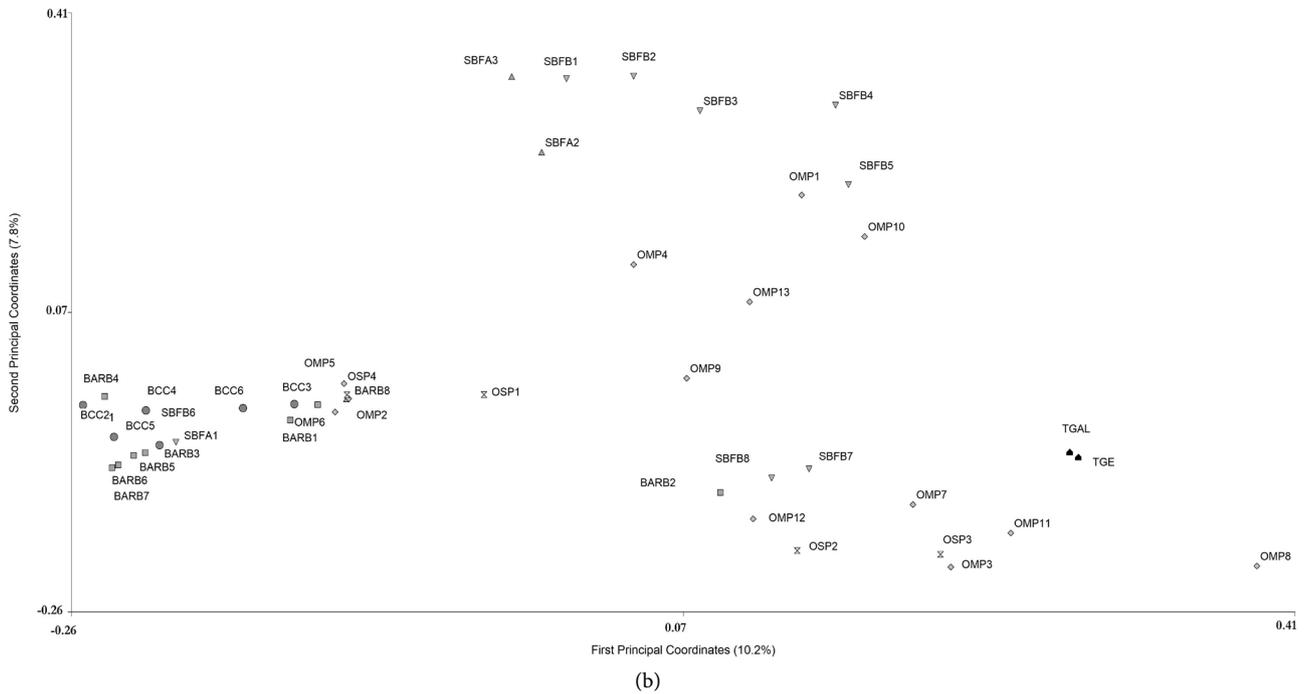
At the individual level approach, selective amplifications of banana clones with the six primer combinations revealed in total 540 reliable fragments, a 100% of total polymorphism among the set of genotypes was found (Figure 1). The number of the pattern duplicated bands were 23, and 21 amplicons were exclusive to a unique farmer's field.

Following up the population analysis, a wide range for Dice's coefficient was obtained. The resulting cluster had a cophenetic correlation of 0.63 (Figure 2(a)). Banana clones were mainly grouped in four clusters according to the farmer's field where they were collected (Figure 2(a)). In congruence, in the Principal Coordinates biplot obtained using Dice's metric and shown in Figure 2(b), the banana clones were placed accordingly to their field origin. Approximately the 18% of the total molecular variation was explained by the two principal coordinates (10.2% by the first coordinate and 7.5% by the second coordinate), which indicates that a great level of DNA polymorphism remains in this group of banana selected clones. AMOVA revealed that 8.9% of the variation was detected among farmer populations and 91.1% within farmer populations (Table 2), *i.e.*, the greatest proportion of genetic diversity was found inside each production field.



**Figure 1.** Example of a piece of the 6% poly-acrylamide denaturing gel stained with the silver nitrate technique for a primer combination E39/M32. Lines 1, 2 and 3: genotype BCC5; line 4: genotype BCC6; line 5: genotype BARB4; line 6:SBFB6; line 7: genotype OSP4; line 8: genotype BCC1; line 9: genotype BCC2; line 10: genotype BCC3; line 11: genotype BCC4; line 12: genotype BARB3; line 13: genotype BARB5; line 14: genotype BARB6; line 15: genotype BARB7; line 16: genotype BARB8; line 17: genotype SBFA1; line 18: genotype OMP2; line 19: genotype OMP5; line 20: genotype OMP6; line 21: M-50 (Molecular Marker). Arrow a: bands share with genotypes arrow b: polymorphic band.





**Figure 2.** Classification of banana genotypes according to molecular polymorphism. (a) Banana genotypes and check-testers in the cluster calculated from AFLP profiles. (b) Banana genotypes and check-testers in the biplot generated by the first (PCoA1) and the second (PCoA2) Principal Coordinates calculated from AFLP profiles.

**Table 2.** AMOVA for the plants among farmer’s field and within populations was due to individual differences within population with 400 random permutations.

Source of variation	d.f	Sum of squares	Variance components	Percentage of variation	p-value
Among farmer’s field	6	1.50	0.02	8.9	<00001
Within population	37	5.84	0.016	91.10	<00001
Total	44	7.34	0.17	100	

#### 4. Discussion

A wide range of molecular polymorphism was found among banana clones in the present research (Figure 1). Previous research in *Musa acuminata* found at least 93 % polymorphic level across all samples examined by AFLP, and the average number of bands scored per primer pair was 61, with 57 being polymorphic [5]. The highest level of polymorphism detected in this set of genotypes (100%) could be the consequence of banana adaptation to subtropical environments, *i.e.*, mutations in tropical areas may be unfavorable and eliminated by natural and artificial selection. Therefore, in the Argentinean suboptimal agroecological area for the crop such mutations would represent a source of genetic variability conferring adaptability to a more severe climate and specific soil conditions. [6] showed that the genetic diversity of *Musa balbisiana* 78.5% of polymorphism with four primer pairs. Moreover, [1] showed that the percentage of polymorphism for individual primer combinations was very high and varied

from 90.3% to 97.6%. These authors were able to identify 1012 polymorphic bands out of total 1094 bands (92.5% polymorphism) with a total of 12-primer combinations. Hence, the 6 six primer combinations polymorphism observed in the present paper (combinations were successful selected from 36 combinations from [8]) are corresponding with previous works cited.

As proposed in previous paragraphs, at the individual level and applying dominant markers, there is a direct identity assumed between the number of unique bands observed and the number of identifiable *loci* for the sample of individuals. When very low DNA sequences divergence between two individuals, the simple match coefficient is expected to be the most suitable measure of similarity in the case of closely related haploid individuals [9]. In this research, the Dice coefficient—which is equivalent to Nei and Li coefficient and Sørensen coefficient [13]—was calculated so that simultaneous absence of a given marker is not considered. Distances among clones widely varied and in agreement, clustering of clones did not necessarily reflect the geographical origin of clones, *i.e.* the farmer's field where they were collected (**Figure 2(a)**). Otherwise, clustering was accomplished by the Ward's method, which is the only one among the agglomerative clustering methods that is based on a classical sum-of-squares criterion so that observed groups are the result of minimizing the within-group dispersion at each binary fusion. In addition, Ward's method is interesting because it looks for clusters in multivariate Euclidean space [11]. Additionally, Ward's method minimizes the increase in total within-cluster sum squared error. Thus, increase is proportional to the square Euclidean distance between clusters centres [14]. In consequence, it could be deduced that heterogeneous genotypes are planted at each farmer's field. Furthermore, Nei and Li index is based on the assumption that differences in banding patterns arise as mutations from a common ancestor, whereas differences within populations may also reflect other factors such as founder effects populations [9]. So it follows that, as explained earlier, farmer would select mutations conferring favourable phenotypes in their own specific conditions.

At a population level, when drawn onto a PCoA ordination diagram, the Ward clustering results often delineate clusters that visually correspond to the density centres in PCoA ordination [11]. On the other hand, Ward's method can also be applied to dissimilarities other than the Euclidean distance. For these dissimilarities, ordinations can be produced by principal coordinate analysis, which is also called classical multidimensional scaling [11]. PCoA takes a symmetric matrix of distances of any type among replicates and produces corresponding Cartesian (Euclidean) coordinates for each replicate that, in the full-dimensional principal coordinate space, preserve the original distances calculated among replicates [15]. Hence, the molecular analysis recapitulates a pattern of the crop cultivation, which can be observed in **Figure 2(b)**: a soft tendency of grouping some clones collected in the same farmer field is balanced by clones from other farmers placing near such geographically determined similarities.

Finally, Analysis of variance (ANOVA) can be modified to directly incorporate this molecular information and thus to become an analysis of molecular variance (AMOVA) [12]. AMOVA tests by a permutational procedure the significance of each variance component, whether estimated from allele-frequency differences or considering molecular information also. For this purpose, the null hypothesis is that all samples come from an unstructured population, without differences among the groups (populations or races) defined within it, so that all variation is due to the random sampling of individuals or populations. In this research, the AMOVA incorporates information about molecular diversity to acquire knowledge of crop production insertion (Table 2). This effect is the result of the artificial selection that was made. The plant material resulting from this process is shown by the low diversity within the farmer's field but with germplasm adapted to the subtropical environment. The result of the AMOVA shows how the crop is adapted to the environment by farmers. Cluster and biplot group plants showing that each farmer has his own method of crop breeding.

## 5. Conclusion

The number of equal polymorphism pattern and the grouping in the cluster and Principal Component biplot, indicate the similarity of the genotypes according to the agronomical development of the crop. Natural and farmers selection on spontaneous mutation may be the cause of groups in the biplot and cluster. The unique bands for each farmer fields suggest that each farmer selection was the most suitable variants for their cultivation system. The component of variation between farmers is smaller than the component within farmers because the genetic structure is common to all them, thus reflecting a high variation among the clones within each field. Introduction of banana into the subtropical zone was associated with a broader genetic variation in order to increase the genetic homeostasis necessary for adapting the crop to the suboptimal environment of Northeastern Argentinean Formosa Province.

## Compliance

Compliance with Ethical Standards is not applicable since neither human beings nor animal were used in this research.

## Funding

This work was supported by the Centro de Validación de Tecnologías Agropecuarias (CEDEVA) de Misión Tacaaglé, Formosa, CONICET (PUE 22920160100043CO, resolución D.2555/16), and Fundación Ciencias Agrarias (NUR, PICT 2014).

## Conflicts of Interest

The authors declare that they have no conflict of interest.

## References

- [1] Opara, U.L., Jacobson, D. and Al-Saady, N.A. (2010) Analysis of Genetic Diversity in Banana Cultivars (*Musa cvs.*) from the South of Oman Using AFLP Markers and Classification by Phylogenetic, Hierarchical Clustering and Principal Component Analyses. *Journal of Zhejiang University SCIENCE B*, **11**, 332-341.  
<https://doi.org/10.1631/jzus.B0900310>
- [2] Robinson, J.C. and Galán Saúco, V. (2012) Plátanos y bananas. Paraninfo, Madrid.
- [3] Molina, N.A., Scribano, F., Tenaglia, G.C. and Rodríguez, D. (2014) Costo de Producción de Banana en Formosa. (I. ediciones, Ed.) INTA EEA Bella Vista. *Serie Técnica*, **50**, 22.
- [4] Molina, N.A., Scribano, F., Tenaglia, G.C. and Rodríguez, D. (2016) Análisis técnico y económico para la producción de banana y mango en Formosa. (I. ediciones, Ed.) Publicación de la EEA INTA Bella Vista. *Serie Técnica*, **55**, 1-10.
- [5] Wong, C., Kiew, R., Loh, J.P., Gan, L.H., Set, O., Lee, S. and Gan, Y.Y. (2001) Genetic Diversity of the Wild Banana *Musa acuminata* Colla in Malaysia as Evidenced by AFLP. *Annals of Botany*, **88**, 1017-1025.  
<https://doi.org/10.1006/anbo.2001.1542>
- [6] Wang, X.L., Chiang, T.Y., Roux, N., Hao, G. and Ge, X.J. (2007) Genetic Diversity of Wild Banana (*Musa balbisiana* Colla) in China as Revealed by AFLP Markers. *Genetic Resources and Crop Evolution*, **54**, 1125-1132.  
<https://doi.org/10.1007/s10722-006-9004-9>
- [7] Ermini, J.L., Tenaglia, G.C. and Pratta, G.R. (2016) Genetic Diversity, Ancestry Relationships and Consensus among Phenotype and Genotype in Banana (*Musa acuminata*) Clones from Formosa (Argentina) Farmers. *Plant Cell Biotechnology and Molecular Biology*, **17**, 267-278.
- [8] Ermini, J.L., Tenaglia, G. and Pratta, G.R. (2013) Marcadores de AFLP en el cultivo de banana: selección de combinación de cebadores y caracterización de la biodiversidad. *Rev. Fac. Cien. Exac. Quím. Nat. Univ. Morón*, **11**, 93-109.
- [9] Kosman, E. and Leonard, K.J. (2015) Similarity Coefficients for Molecular Markers in Studies of Genetic Relationships between Individuals for Haploid, Diploid, and Polyploid Species. *Molecular Ecology*, **14**, 415-424.  
<https://doi.org/10.1111/j.1365-294X.2005.02416.x>
- [10] Balzarini, M. and Di Rienzo, J. (2016) InfoGen Versión 2016. FCA, Universidad Nacional de Córdoba. <http://www.info-gen.com.ar>
- [11] Murtagh, F. and Legendre, P. (2014) Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, **31**, 274-295. <https://doi.org/10.1007/s00357-014-9161-z>
- [12] Balding, D.J., Bishop, M. and Cannin, C. (2008) Handbook of Statistical Genetics (Vol. 2). John Wiley Sons, Toronto.
- [13] Bonin, A., Ehrich, D. and Manel, S. (2007) Statistical Analysis of Amplified Fragment Length Polymorphism Data: A Toolbox for Molecular Ecologists and Evolutionists. *Molecular Ecology*, **16**, 3737-3758.  
<https://doi.org/10.1111/j.1365-294X.2007.03435.x>
- [14] Szekely, G.J. and Rizzo, M.L. (2005) Hierarchical Clustering via Joint Between-Within Distances: Extending Ward's Minimum Variance Method. *Journal of Classification*, **22**, 151-183. <https://doi.org/10.1007/s00357-005-0012-9>
- [15] Gower, J.C. (1966) Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. *Biometrika*, **53**, 325-338.  
<https://doi.org/10.1093/biomet/53.3-4.325>