Genetic diversity of yacon (*Smallanthus sonchifolius)* in Peru

Julián Soto[1*], Cinthya Zorrilla[1**], Iván Manrique[1], William Roca[1], Tulio Medina[2], Raúl Blas[3], Juan Seminario[4], Luis Lizarraga[5], José Quispe[6], David Tay[1] , Carlos Arbizu[1]

1 Centro Internacional de la Papa (CIP). Av. La Molina 1895, La Molina, Lima, Peru
2 Instituto Nacional de Innovación Agraria (INIA). Av. La Molina 1981, Lima, Peru
3 Universidad Nacional Agraria La Molina (UNALM). Av. La Molina s/n., Lima, Peru
4 Universidad Nacional de Cajamarca (UNC). Av. Atahualpa 1050, Cajamarca, Peru
5 Universidad Nacional de San Antonio Abad del Cusco (UNSAAC). Av. de la Cultura 733, Cusco, Peru
6 Universidad Nacional de San Cristóbal de Huamanga (UNSCH). Portal Independencia 57, Ayacucho, Peru
* Corresponding author: j.v.soto@cgiar.org
** Current address: University of Wisconsin-Madison, Department of Horticulture. 1575 Linden Drive Madison, WI 53706

Abstract

Yacon -an ancient Andean root crop- is considered a truly functional food because it is an important source of health related compounds, including fructooligosaccharides and antioxidants. The plant is grown from Ecuador to northwestern Argentina, with the highest concentration of diversity between Peru and Bolivia. Genetic diversity of yacon in Peru has been studied only by means of morphological features, employing non-standardized descriptors. In this work a total of 309 polymorphic AFLP markers –obtained from 6 primer combinations- were used to study the diversity of 359 accessions of yacon maintained by six Peruvian genebanks.

Only seven accessions were identified as duplicates and 352 were unique genotypes. However, the genetic diversity index (Nei and Shannon-Weaver) and the analysis of molecular variances (AMOVA) showed low levels of diversity between genebanks, and within them. The cluster and factorial analyses identified three major groups of yacon in the collection: the first and second groups included accessions from northern and southern Peru, respectively; a third group consisting of accessions from northern, central and southern regions. These results could be useful to establish strategies for future collecting missions and germplasm conservation.

Keywords: Genetic diversity, yacón, *Smallanthus sonchinfolius,* AFLP.

Introduction

The yacon, is an Andean root crop domesticated before the pre-inca times. Its reservant roots are rich sources of fenolic compounds and fructooligosacarides (3.8% and 60% of dry weight, respectively) (Seminario et al., 2003). There are evidences that these compounds generate benefic effects in the health and play important functions in the prevention of several chronic diseases, such as diabetes, obesity, dyslipidemia, atherosclerosis and colon cancer (Valentová et al., 2004: Genta et al., 2009).

The yacon is the only species domesticate and eatable of the genus *Smallanthus*. Its natural habitat is the Andean region from the south of Colombia to the northwest of Argentina, and the most probable center of origin is a narrow slope between south of Peru and north of Bolivia (Grau and Rea, 1997). Six institutions of Peru, with a total of 359 accessions, maintain the biggest collections of yacon. These collections have been characterized morphologically in the past under different criteria and doing the evaluations in different environments and conditions. Therefore, there is no reliable information to estimate the total diversity of Peruvian yacon neither the number of morphotypes or genotypes.

At the molecular level, there are few studies reported in yacon. Milella et al. (2005) after testing five yacon clones by RAPD markers suggested that the technique could be useful for identification and differentiation of varieties. Mansilla et al. (2006) using RAPD markers in the collection of CIP reported more diversity in central Peru compared to the north and south. However, the contribution of both publications to assess the total genetic diversity of yacon was low because the numbers of samples were small and not representative. In comparison with the RAPD markers, the AFLP markers can be more useful for studies of diversity because they reveal a larger

number of polymorphisms and can differentiate duplicates genotypes. In addition, AFLP technique does not require DNA sequence information and the PCR technique is rapid and reproducible (Ferreira y Grattapaglia, 1998).

The aim of this study was to assess the genetic diversity of the six more representative *ex situ* collections of yacon of Peru, using AFLP markers.

Materials y methods

Plant material and DNA extraction

A total of 359 accessions of yacon of six institutions of Peru were tested: 48 accessions from CIP, 123 from INIA, 103 from UNC, 62 from UNSAAC, 13 from UNSCH and 10 from UNALM. DNA extraction was made from young leaves using CTAB technique (Doyle and Doyle, 1990) medium scale protocol standardized and reported by CIP (2000). The DNA concentration was calculated in comparison with known concentrations of Phage Lambda DNA digested with *Pst I* restriction enzyme and visualized in agarose gel 1%.

AFLP analysis

AFLP protocols are based on Vos et al. (1995), adapted for use with silver staining of 6% polyacrylamide gels. The protocols of DNA restriction and ligation, PCR amplifications and electrophoresis were used as reported in the manual by CIP (2000).

Data analysis

The AFLP bands pattern for each accession was recorded manually on a binary matrix. The present bands were recorded as 1 and those absent as 0. The identification of unique genotypes and redundancy was made using a cluster analysis based on Jaccard similarity index (1908) and the UPGMA algorithm (Sokal and Sneath, 1963) using the software DARwin5 (Perrier, et al., 2006). Genetic heterogeneity was estimated using the Nei's (1973) and Shannon-Weaver diversity index (1949), the genetic structure was conducted using the Principal Coordinates Factorial analysis and the changes of genetic diversity between collections was evaluated by analysis of molecular variance - AMOVA (Excoffier et al., 1992) using the software Arlequin ver. 3.11 (Excoffier et al., 2005).

Results

Selection of AFLP primers combinations (E+3/M+3)

Six primer combinations were selected from a test of 120 combinations of primers with three selective nucleotides. These primers were selected due to its high polymorphisms index and good resolution of the bands (Table 1).

Table 1. Combination of primers (E+3 / M+3) used in the molecular characterization of yacon

Cod. Lab.	Combination	Primer	Sequence (5'- 3')
E40-M36	EAGC-MACC	EAGC	GACTGCGTACCAATTC-**AGC**
		MACC	GATGAGTCCTGAGTAA-**ACC**
E38-M33	EACT-MAAG	EACT	GACTGCGTACCAATTC-**ACT**
		MAAG	GATGAGTCCTGACTAA-**AAG**
E39-M36	EAGA-MACC	EAGA	GACTGCGTACCAATTC-**AGA**
		MACC	GATGAGTCCTGAGTAA-**ACC**
E42-M35	EAGT-MACA	EAGT	GACTGCGTACCAATTC-**AGT**
		MACA	GATGAGTCCTGACTAA-**ACA**
E36-M55	EACC-MCGA	EACC	GACTGCGTACCAATTC-**ACC**
		MCGA	GATGAGTCCTGAGTAA-**CGA**
E37-M60	EACG-MCTC	EACG	GACTGCGTACCAATTC-**ACG**
		MCTC	GATGAGTCCTGAGTAA-**CTC**

Genotypes duplicates

From 359 accessions evaluated by 309 polymorphic markers generated by six AFLP primer combinations, 355 unique genotypes were found, although most of them share a close genetic similarity. In addition, three genotypes were represented by more than one accession; they were named duplicate accessions (100% of genetic similarity): two genotypes belonging to UNSAAC collection with 3 (ZS-074, ZS-024 and ZS-050) and 2 (ZS-072 and ZS-067) accessions respectively. The third genotype belongs to the CIP collection with 2 accessions (205037 and 205031).

Cluster analysis and geographical distributions pattern

Using a genetic similarity coefficient of 87% into cluster analysis, it was possible to identify 2 large molecular groups (I and II), which include 329 accessions. The 30 accessions remaining do not present a defined cluster because of their low genetic relatedness between them and the groups identified. In addition, the group I contains 4 subgroups: the subgroups A and C are formed by accession of the north, the subgroup B are formed by accessions of the south and the last subgroup and also the group II are formed by accessions from the three regions of Peru (Figure 1).

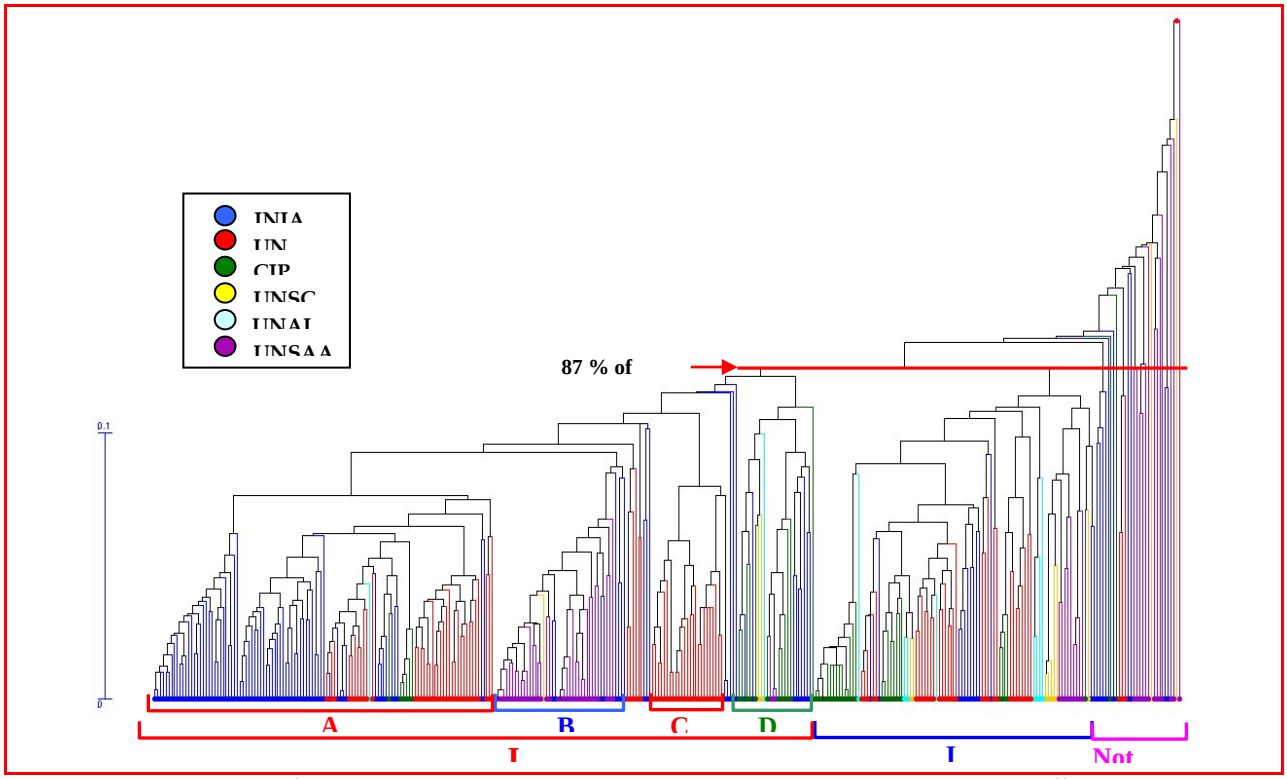

Figure 1. Dendrogram of yacon collections based on 309 AFLP markers using the Jaccard coefficient and UPGMA algorithm. subgroups A and C are formed by accessions from northern, subgroup B is formed by accessions from southern and subgroup D and Group II are formed by accessions from the three regions

The factorial analysis shows the presence of only three groups: 2 groups consisting of accessions from the north and south respectively, and a group of accessions from the three regions. Although these three groups are distinguishable in the analysis, they present a slight overlapping, indicating that the differences are not very distant (Figure 2).

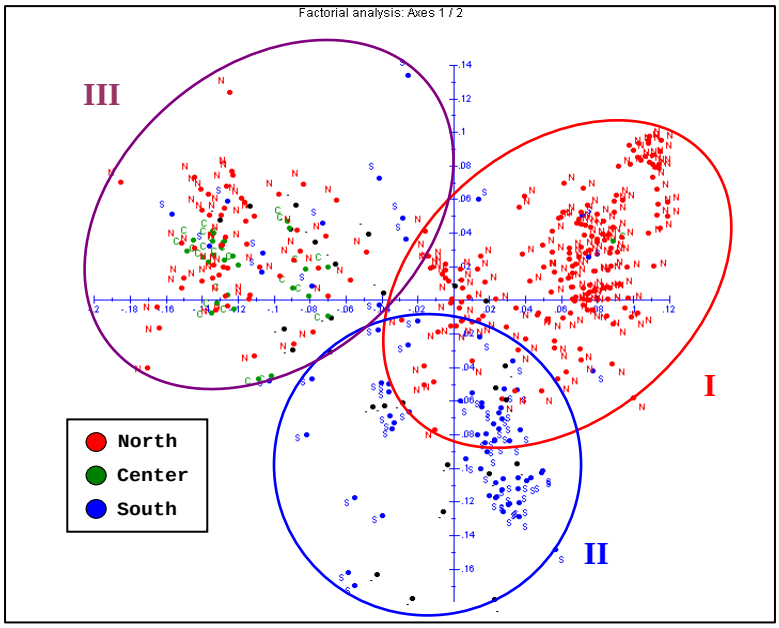

Figure 2. Factorial Analysis of yacon collections. Group I is formed by accessions from northern Peru. Group II is formed by accessions from the south and Group III is formed by accessions from the three regions

Genetic diversity

Table 2. Nei and Shannon-Weaver index

Collections	Number of accessions	Nei index	Shannon-Weaver index
CIP	48	0.14	4.51
INIA	123	0.13	4.51
UNC	103	0.13	4.27
UNSAAC	62	0.15	4.74
UNSCH	13	**0.17**	**5.17**
UNALM	10	0.14	3.92
Total	359	**0.13**	**5.90**

The genetic diversity of the entire collection of yacon is 0.13 (Nei's index) or 5.90 (Shannon-Weaver index). The diversity among collections is very similar and the highest value is presented in the UNSCH collection (Table 2).

The entire collection of yacon presents 37 exclusive bands (bands that are only present in one collection) and the UNSAAC collection has the highest number of exclusive bands. In addition, there are 192 bands (62%) that are shared among all collections (Table 3).

Table 3. Number of present, exclusive and share bands

Collections	Total number of present bands	Number of exclusive bands	Number of share bands
CIP	220	3	192
INIA	249	7	
UNC	232	5	
UNSAAC	**253**	**15**	
UNSCH	220	3	
UNALM	207	4	

Analysis of molecular variance

To perform the AMOVA analysis, we identified 3 geographical regions (northern, central and south) and also the departments that are include in each region. There are significant differences in the genetic heterogeneity of the three regions but with a low value of variation (7.28%), also the fixation index shows a moderate differentiation (Fst = 0.0728). The same situation appears between departments (Fst = 0.1248) but it is higher than regions (Table 4). Additionally, the AMOVA was performed to show the differences between the six collections. There were significant differences among the six collections with a high differentiation (Fst = 0.18 – data not shown).

Table 4. Analysis Molecular Variance - AMOVA from three regions of collection and Fixation index

Source of variation	d.f.	Sum of squares	Percentage of variation	Fixation Index
Among regions	2	525.279	7.28 (+)	0.0728
Among departments within regions	12	589.951	11.57 (+)	0.1248
Within populations	341	5697.916	81.15	0.1885
Total	355	61813.146		
Significance Level = 0.05		Significance (+)	Non significance (-)	

Discussion

The genetic diversity of yacon maintained in the six collections was low (Nei = 0.13, Shannon-Weaver = 5.90). Also, there are narrow ranges of similarity in which groups are formed in the dendrogram (~0.87-1.00 of similarity), which indicates close relation between genotypes and as a result a lack of a better clustering.

All collections represent the total genetic diversity because their diversity indexes were similar to the total index. And the AMOVA analysis shows that these collections have a moderate genetic differentiation (Fst = 0.18 – data not shown). The exclusive bands of each collection and differences in the number of accessions between collections would be the factors that make differences in diversity among collections. The highest diversity value (0.17-0.517) presented in the UNSCH collection could be overestimated due to the presence of low number of accessions.

Almost 99% of the accessions studied were unique genotypes (355), but they have a close genetic relationship (62% of shared bands). According to Grau and Rea (1997), the area from Apurimac in Peru to La Paz in Bolivia, holds the largest area of diversity, which in this work is supported by a large number of unique bands present in the collection of UNSAAC, since in this collection most of its accessions are from southern Peru.

Cluster analysis shows two main groups and subgroups A, B and C show a geographical patterns; although, this analysis also present a high number of accessions with lack of clustering. The factorial analysis shows a better association with geographical regions (north, central and southern Peru). Grau & Rea suggest that the geographical distribution is a very important marker for yacon, due to differences in tuber size that were found between the area of Ecuador, Cajamarca and Cusco in Peru. Also, they found a high variability of tuber flesh color in southern Peru and northern Bolivia compared to Ecuador and Argentina. In this analysis it was identified mainly 3 genetically different groups: the first consisting of accessions from the north, a second formed by accessions from the south and the last consisting of accessions from the 3 regions. Although these three groups can be differentiated in the factorial and AMOVA analysis, they maintained a high genetic relationship. The differences in management and selection of yacon in the north and south may be influence the genetic structure of this crop in the two regions.

The use of morphological data and other important characteristics could improve the results of this research for a better management of the genetic diversity of yacon and develop new strategies for collect and conservation of yacon in the future.

References

Doyle J.J. and Doyle J.L. 1990. *Isolation of plant DNA from fresh tissue*. Focus (Gibco BRL) 12: 13-15.

Excoffier L., Smouse P. E. and Quattro J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics, 131:479-491.

Excoffier L., Laval G., and Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online **1**:47-50.

Ferreira M. E. and Grattapaglia D. 1998. Introducción al uso de marcadores moleculares en el análisis genético. EMBRAPA, Brasil.

Genta S., Cabrera W., Habib N., Pons J., Manrique I., Grau A. and Sánchez S. 2009. Yacon syrup: Beneficial effects on obesity and insulin resistance in humans. Clinical Nutrition 28: 182-187.

Grau A. and Rea J. 1997. Yacón, *Smallanthus sonchifolius* (Poepp. & Endl) H. Robinson. Pp 199-242. In: **Hermann M. and Heller, J.** *Andean Roots and tubers: Ahipa, Arracacha, Maca and Yacon.* Promoting the conservation and use of underutilized and neglected crops. 21. Institute of Plant Genetics and Crop Plant Research, Gatersleben/ International Plant Genetic Resources Institute, Rome, Italy.

Jaccard P. 1908. Nouvelles rescherches sur la distribution florale. Bull. Soc. Vaud. Sci. Nat., 44:223-270.

International potato Center - CIP. 2000. Molecular Biology Laboratory Protocols: Plant Genotyping. Ghislain M, Zhang D, Herrera R (Eds.). Crop Improvement and Genetic Resources Department, Training manual, 3rd Ed. Lima, Peru.

Mansilla R. C., Lopez B. C., Blas S. R., Chia W.J. and Baudoin J. 2006. Análisis de la variabilidad molecular de una colección peruana de *Smallanthus sonchifolius* (Poepp. & Endl)H. Robinson 'yacon'. Ecología Aplicada (Peru) 5(1/2): 75-80.

Milella L., Salava J., Martelli G., Greco I., Cusimamani E. and Viehmannová I. 2005. Genetic Diversity between Yacon Landraces from Different Countries Based on Random Amplified Polymorphic DNAs. Czech J. Genet. Plant Breed. 41, 2005(2): 73-78

Nei M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the USA, 70 (12): 3321-3323.

Perrier X., Jacquemoud-Collet J. P. 2006. DARwin software http://darwin.cirad.fr/darwin

Seminario J., Valderrama M. and Manrique I. 2003. El Yacón, Fundamentos para el aprovechamiento de un recurso promisorio. Centro Internacional de la papa (CIP), Universidad Nacional de Cajamarca (UNC), Agencia suiza para el desarrollo y la cooperación (COSUDE). Lima, Perú.

Seminario J., Valderrama M. and Romero J. 2004. Variabilidad morfológica y distribución geográfica del yacón, *Smallanthus sonchifolius* (Poepp. & Endl)H. Robinson, en el norte peruano. Arnaldoa, 11(1): 139-160.

Sokal R. R. and Sneath P. H. 1963. Principles of Numerical Taxonomy. W. H. Freeman and Company. San Francisco – USA. 359p.

Shannon C. 1949. The mathematical theory of communication. Eds. C. E. Shannon y W. Weaver. The mathematical Theory of Comunication. University of Illinois Press. Urbana.

Valentová K., Moncion A., de Waziers I. and Ulrichová J. 2004. The effect of Smallanthus sonchinfolius leaf extracts on rat hepatic metabolism. Cell Biology and Toxicology 20 (2): 109-120.

Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Friters A., Pot J., Paleman J., Kuiper M. and Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research, 23 (21): 4407-4414.