Microsatellite Analysis of Genetic Diversity in Sweetpotato Varieties from Latin America

D.P. Zhang, D. Carbajulca, L. Ojeda, G. Rossel, S. Milla, C. Herrera, and M. Ghislain[1]

Sweetpotato (*Ipomoea batatas* (L.) Lam.) was originally domesticated in tropical America. The sweetpotato gene bank maintained at CIP now contains 5526 cultivated accessions from 57 countries, of which 2589 are from Latin America. Understanding the diversity in the distribution of this germplasm is essential for its rational management and use. In the present study, we analyzed a group of sweetpotato varieties from Latin America using simple-sequence repeat (SSR) markers. Twelve SSR primer pairs with dinucleotide tandem repeats were screened. Six primer pairs that generated scorable allelic information were used to type the 113 varieties. The six SSR loci revealed a total of 70 alleles, with allele size ranging from 102 base pairs to 173. Both the richness and the evenness of the alleles show a significant geographical pattern in the Latin American sweetpotato gene pool. Mesoamerica ranks the highest in terms of total number of alleles, number of region-specific alleles, and actual heterozygosity, whereas the region of Peru-Ecuador ranks the lowest on all three counts. This, together with our earlier findings based on AFLP analysis, strongly supports the hypothesis that Mesoamerica is the primary center of diversity and most likely the center of origin of the sweetpotato. Peru-Ecuador should be considered as a secondary center of sweetpotato diversity. The tetrasomic inheritance of these SSRs also supports the hypothesis that sweetpotato is an alloautohexaploid with two non-homologous genomes.

The successful genetic conservation of any given gene pool is largely dependent on understanding the diversity of its distribution in a region. This information and the organism's history of domestication are crucial for constructing core collections. A core collection is a limited subset of accessions, derived from a larger germplasm collection, chosen to represent the genetic spectrum of the whole collection (Brown, 1989). When the genetic diversity is known in advance, the preferred approach is to devise a strategy where the stratification and sampling are in proportion to the range of genetic diversity. A combination of passport data and information on genetic diversity from molecular markers would therefore enhance the formation of core collections.

It is commonly accepted that the sweetpotato is of American origin. The region between the Yucatán Peninsula of Mexico and the Orinoco River in Venezu-

[1] CIP, Lima, Peru.

ela has been postulated as the center of diversity because there is rich morphological variation in cultivated sweetpotatoes, and four major American taxa of the batata group are distributed in this region (Austin, 1988).

The sweetpotato genebank held at CIP maintains a total of 5526 cultivated accessions from 57 countries, of which 2589 are from Latin America (Huamán and Zhang, 1997). Most of these Latin American accessions are landraces or farmer's varieties. This collection is now undergoing molecular characterization. Dominant molecular markers, such as RAPD and AFLP, are being routinely used at CIP for eliminating redundancy (Zhang et al., 2001) and assessing genetic diversity (Zhang et al., 1998; 2000).

Although these dominant markers have been proved genetically informative for sweetpotato, they do not properly contribute to our understanding of the allelic diversity in this collection. In addition, the lack of sequence specificity in randomly amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP) markers limits their cross-lab application for variety identification. Therefore, a polymerase chain reaction (PCR)-based, sequence-tagged, co-dominant marker system is needed to play a complementary role in gene bank management.

Simple-sequence repeats (SSRs), also called microsatellites, are becoming the most important molecular markers in both animals and plants. SSRs are stretches of 1 to 6 nucleotide units repeated in tandem and randomly spread in eucaryotic genomes. They are highly polymorphic because of the high mutation rate affecting the number of repeated units. Polymorphisms of this length can be easily detected on high-resolution gels (e.g., sequencing gels) by running PCR-amplified fragments obtained using a unique pair of primers flanking the repeat (Weber and May, 1989).

A group of SSRs have already been developed in sweetpotato (Jarret and Bowen, 1994). These have been further screened and applied in paternity analysis in sweetpotato and its wild relative species (Buteler et al., 1997; 1999). In this paper, we report the application of these SSRs to the assessment of genetic identity and genetic diversity in a group of sweetpotato varieties from Latin America. We show the significant variation in the distribution of allelic diversity in the Latin American sweetpotato gene pool and discuss the importance of SSR markers in future germplasm management.

Materials and Methods

All plant materials were obtained from the sweetpotato germplasm collection at CIP. Four to 10 sweetpotato landraces were randomly selected from each country included in this study, resulting in a total of 113 accessions with a geographic coverage ranging from Mexico to Peru (Table 1). Healthy young leaves were collected from accessions maintained in a screen house and in vitro. DNA was extracted following the method of Doyle and Doyle (1990).

The SSR primers for sweetpotato were originally developed by Jarret and Bowen (1994). We used the Mappair primer sets commercially available from Research Genetics (Huntsville, AL, USA). Eight pairs of these primers were pre-screened by Buteler et al. (1997; 1999) at the Sweetpotato Breeding Program of Louisiana State University.

A non-radioactive PCR was carried out following CIP's SSR genotyping protocol. Annealing temperature was adjusted, based on the Tm of the oligo-nucleotides. For screening of primers, the amplification products were first confirmed on 3% agarose gel with eight varieties. Primer pairs amplifying at least one PCR fragment in every variety were further tested on the complete set of varieties using a 6% denaturing sequencing gel (National

Table 1. List of the 113 Latin American sweetpotato accessions used for microsatellite analysis.

Code	Country of origin	Collection No.	Code	Country of origin	Collection No.	Code	Country of origin	Collection No.
1	Peru	CIP ARB 386	39	Colombia	DLP 2104	77	Guatemala	GUA 948
2	Peru	DLP 1900	40	Colombia	DLP 1793	78	Guatemala	GUA STRA
3	Peru	DLP 1922	41	Colombia	DLP 1858	79	Guatemala	GUA 494
4	Peru	DLP 206	42	Colombia	DLP 1736	80	Mexico	CTX 15
5	Peru	DLP 2	43	Colombia	DLP 1771	81	Mexico	CTX 7
6	Peru	DLP 1090	44	Colombia	DLP 1870	82	Mexico	CTX 22
7	Peru	DLP 2344	45	Colombia	DLP 1011	83	Mexico	CTX 24
8	Peru	ARB 234	46	Colombia	DLP 1685	84	Mexico	CTX 31
9	Peru	DLP 5314	47	Colombia	DLP 1731	85	Mexico	A 160
10	Peru	DLP 3824	48	Colombia	DLP 2046	86	Mexico	CTX 33
11	Peru	RCB IN- 90	49	Colombia	DLP 1879	87	Mexico	IO1438
12	Peru	DLP 1921	50	Colombia	DLP 1737	88	Mexico	NIAR 221
13	Peru	ARB 355	51	Colombia	DLP 1785	89	Mexico	CATIE 9232
14	Peru	ARB 455	52	Colombia	DLP 2151	90	Mexico	CTX 29
15	Peru	DLP 2298	53	Colombia	DLP 976	91	Mexico	RCB-IF-30
16	Peru	DLP 253	54	Colombia	DLP 971	92	Mexico	CATIE 9257
17	Peru	RCB IN-199	55	Colombia	LL 87-1799	93	Mexico	CTX 31
18	Peru	DLP 909	56	Colombia	DLP 972	94	Mexico	CTX 32
19	Ecuador	DLP 1192	57	Colombia	DLP 1755	95	Mexico	CTX 12
20	Ecuador	DLP 1161	58	Venezuela	DLP 2896	96	Mexico	CTX 34
21	Ecuador	DLP 1456	59	Venezuela	DLP 869	97	Mexico	CTX 5
22	Ecuador	DLP 1475	60	Venezuela	DLP 824	98	Mexico	CTX 16
23	Ecuador	DLP 1449	61	Venezuela	DLP 806	99	Mexico	CTX 9
24	Ecuador	DLP 1447	62	Venezuela	DLP 868	100	Nicaragua	DLP 4678
25	Ecuador	DLP 1156	63	Venezuela	DLP 2868	101	Nicaragua	DLP 4686
26	Ecuador	DLP 1487	64	Venezuela	DLP 2869	102	Nicaragua	DLP 4617
27	Ecuador	DLP 1153	65	Venezuela	DLP 2902	103	Nicaragua	DLP 4675
28	Ecuador	DLP 1493	66	Venezuela	DLP 2884	104	Panama	DLP 3874
29	Ecuador	DLP 1186	67	Venezuela	DLP 2876	105	Panama	DLP 3834
30	Ecuador	DLP 1397	68	Venezuela	DLP 2838	106	El Salvador	SVG 27
31	Ecuador	DLP 1257	69	Venezuela	DLP 800	107	El Salvador	SVG 12
32	Ecuador	DLP 1149	70	Venezuela	DLP 807	108	El Salvador	SVG 24
33	Ecuador	DLP 1435	71	Venezuela	DLP 822	109	El Salvador	SVG 8
34	Ecuador	DLP 1157	72	Venezuela	DLP 792	110	Honduras	DLP 4545
35	Ecuador	DLP 1231	73	Venezuela	DLP 765	111	Honduras	DLP 4494
36	Ecuador	DLP 1498	74	Venezuela	DLP 790	112	Honduras	DLP 4521
37	Ecuador	DLP 1484	75	Venezuela	DLP 2881	113	Honduras	DLP 4558
38	Ecuador	DLP 1405	76	Venezuela	DLP 791			

Diagnostic, Atlanta, GA, USA). Visualization was accomplished by silver staining (Bassam et al., 1991).

Each SSR variant was treated as an allele. The alleles were numbered sequentially starting with the shortest sequence. The co-run sequence ladder in the gel helped us to identify single base differences between any two given samples. The SSR allele composition of each variety was determined. We maintained a common approach to the scoring of the gel by placing position 1 as the lowest molecular weight species and position 2 as the next higher weight and so forth. By running a dideoxy-sequence next to several of the reactions, we were able to distinguish between bands that differed in size by a single base.

Genotypes were scored as +/- for each allele. For the purposes of this study, we defined three regions for the 10 countries:

(1) Mesoamerica (Mexico and Central America), (2) Colombia-Venezuela, and (3) Peru-Ecuador. The total number of alleles, average number of alleles per locus, and percentage of heterozygous individuals were calculated for each region (Nei, 1973). The average locus heterozygosity was determined by ALH = 100 - [S(1-{m_i/N})], where m_i is the number of accessions with mono-allelic genotypes for SSR locus i and N is the number of accessions scored.

Results and Discussion

Number of alleles

Of the 12 pairs of SSR primers, six generated a reproducible banding pattern. An SSR profile example, loci 242, is presented in Figure 1. The number of alleles varies greatly among the SSR loci. Locus IB316 only detected six alleles, whereas locus IB324 detected 30 in the 113 accessions. In total, the six SSR loci detected 70 SSR variants with an allele size ranging from 102 bp to 173 bp (data not shown). Most of the alleles in each locus can be arranged on a continuous ladder, with adjacent steps differing by two base pairs. There was no correlation between number of repeats and the level of polymorphism detected.

The total number of alleles and the number of alleles per locus are not evenly distributed among geographic regions. Mesoamerica has the highest total number of alleles (50), Peru-Ecuador has the lowest (41), and Venezuela-Colombia is in between (46). There is a total of 14 region-specific alleles (alleles found in one region but missing in the other two), of which eight are found in Mesoamerica, four in Venezuela-Colombia, and only two in Peru-Ecuador. In addition, the predominant alleles in Peru-Ecuador differed greatly from the other two regions in loci IB255 and IB286. Allele IB255/1 is predominant in Mesoamerica and in Venezuela-Colombia, shared by approximately 30% of the accessions from these two regions, but in Peru-Ecuador, fewer than 10% of the accessions have this allele (Table 2).

Actual heterozygosity

Because of the polyploid nature of sweetpotato, the dosage effects of an SSR allele (simplex, duplex, triplex, etc.) cannot be differentiated. Therefore the allele frequency could not be counted, as

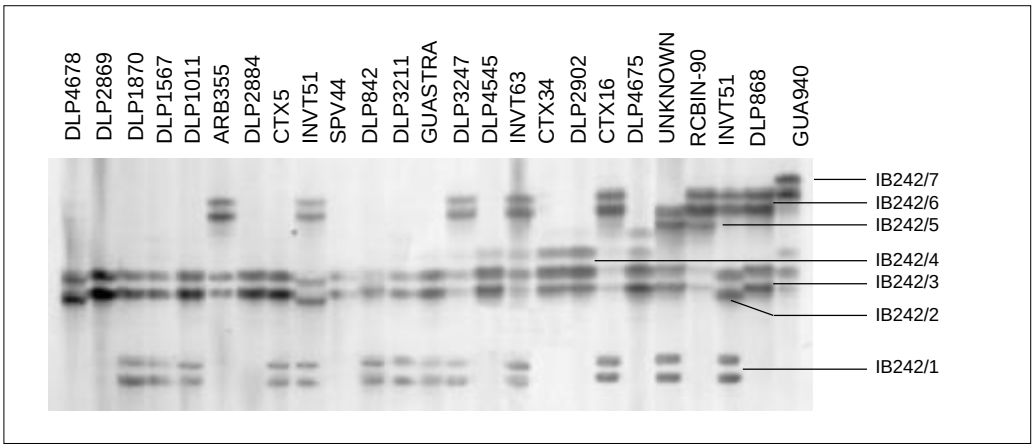

Figure 1. SSR-based multi-allelic fingerprints in 25 Latin American sweetpotato accessions. Lines and letters on the right mark the putative alleles. Alleles were numbered starting with the shortest sequence. In this example, the lowest band (allele) is #1 and the highest is #7, with alleles 2–6 lying in between these points. When the autorads were scored, genotypes were scored as 1/0 for each allele.

it can in diploid species, and the expected heterozygosity (Nei, 1973) cannot be estimated. However, the observed heterozygosity can be calculated by using the percentage of multi-allelic accessions. The actual heterozygosity varies greatly across the six loci, ranging from 0.31 (IB255) to 0.92 (IB316), with a mean of 0.60. This value is reasonable, considering that sweetpotato is a hexaploid outbreeding species. The actual heterozygosity in *S. tuberosum* was reported to range from 0.477 to 0.502 (Bonierbale et al., 1993). A similar value (0.495) was found in CIP's holding of *S. andigena* using allonym analysis (Huaman et al., 2000). However, sweetpotato has a much higher percentage of tri-allelic and tetra-allelic genotypes than potato does. The present study found approximately 25% tri- and tetra-allelic accessions in sweetpotato, but there are only 3.2% in potato (Huamán et al., 2000). In other words, sweetpotato is generally much more heterozygous than potato.

Table 2. Regional distribution of allelic diversity in Latin American sweetpotato germplasm. The allelic diversity is measured by total number of alleles, region-specific alleles, and actual heterozygosity based on 6 SSR loci.

Region	Total alleles	Region-specific alleles	Sample size	Individuals (%)				
				Mono-allelic	Di-allelic	Tri-allelic	Tetra-allelic	Penta-allelic	Hexa-allelic
Peru-Ecuador	41	2	38	47.9	27.4	17.4	7.4	0	0
Colombia-Venezuela	46	4	38	29.5	45.3	20.5	4.7	0	0
Mesoamerica	50	8	37	28.6	42.7	20	8.7	0	0
Total	70	14	113	35.3	36.5	19.3	8.9	0	0

The actual heterozygosity also varies greatly among the three regions, similar to allele numbers. It is 0.714 in Mesoamerica and 0.705 in Venezuela-Colombia, but it is only 0.521 in Peru-Ecuador.

The richness and evenness of alleles in the three tropical American regions are fully compatible with our recent diversity analysis for the same three regions using AFLP (Zhang et al., 2000), which found that Central America has the highest intra-specific diversity and Peru-Ecuador the lowest. This provides further evidence that Mesoamerica is the primary center of diversity and most likely the center of origin of the sweetpotato.

Autopolyploid vs. allopolyploid

Sweetpotato is a hexaploid species; therefore, each individual genotype could contain between one and six alleles at any one locus, assuming it is an autopolyploid. However, in this study, the maximum number of alleles from any given variety was four. Varieties showing five or six distinct alleles per locus (hexa- and penta-allelic) were completely missing. The lack of hexa- and penta-allelic genotypes indicates that sweetpotato is not an autohexaploid and the SSR alleles are in a pattern of tetrasomic inheritance. This agrees with previous findings that these SSRs were in tetrasomic segregation in the sweetpotato (Buteler et al., 1999). It is also compatible with the result of our genome mapping, which demonstrated that sweetpotato has a high level of polysomic inheritance of homologous chromosomes, but with partial preferential chromosome pairing.

Both allopolyploidy (Jones, 1965) and autopolyploidy (Nishiyama, 1982) have been proposed as the nature of polyploidization in sweetpotato. Shiotani and Kawase (1989) proposed that sweetpotato has a 'tetra-disomic' genetic constitution with two different genomes (B1B1B2B2B2B2). The tetrasomic inheritance of these SSRs in the present study provides new evidence to support the hypothesis that there are two non-homolo-

gous genomes involved in sweetpotato and that the plant is most likely an alloautohexaploid.

SSR as fingerprint for variety identification

The extraordinary discriminatory capacity of microsatellite markers observed in other species has been confirmed in the present study. All 113 accessions in this study can be fully identified with as few as four SSRs. It is highly desirable to have repeatable fingerprints for germplasm management, and the presence of easily scorable, unique alleles and/or allele combinations makes microsatellite markers an ideal system for variety identification. Other PCR-based fingerprints, such as AFLP and RAPD, are also able to differentiate varieties, but often have much lower repeatability.

Microsatellite markers have their own limitations when used on polyploid species. The most important limitation is that the real genotype of a hexaploid individual cannot be revealed because the PCR-based nature of SSRs cannot differentiate the dosage effect of a given allele (i.e., it cannot tell the difference between simplex and duplex). The occurrence of null alleles is another possible problem with the use of microsatellite markers in highly outbreeding, heterozygous species (Powell et al., 1996). Because of these limitations, the allele frequency of a given germplasm pool cannot be calculated, and classical population genetics cannot be fully applied. Nevertheless, the nature of microsatellites as being selectively neutral, co-dominant, and sequence-tagged makes them a very useful tool for germplasm management, as well as for genome mapping of sweetpotato.

New SSRs for sweetpotato are currently being developed at CIP using high quality genomic DNA library enriched for multiple types of SSR motifs. Our goal is to have a set of well-characterized SSRs available to serve the needs of sweetpotato germplasm management, both at CIP and in national programs.

Acknowledgements

We thank Robert Jarret, Don La Bonte, and Mario Buteler for providing the sequences of screened primers.

References

Austin, D.F. 1988. The taxonomy, evolution and genetic diversity of sweetpotatoes and related wild species. In: Gregory, P. (ed.). Exploration, maintenance, and utilization of sweetpotato genetic resources. International Potato Center, Lima, Peru. p. 27–60.

Bassam, B.J., G. Caetano-Anollés, and P.M. Gresshoff. 1991. Fast and sensitive silver staining of DNA in polyacrylamide gels. Annals of Biochemistry 196:80–83.

Bonierbale, M.W., R.I. Plaisted, and S.D. Tanksley. 1993. A test of the maximum heterozygosity hypothesis using molecular markers in tetraploid potatoes. Theoretical Applied Genetics 86:481–491.

Brown, A.H.D. 1989. Core collection: A practical approach to genetic resources management. Genome 31:818–824.

Buteler, M.I., R.L. Jarret, and D.R. La Bonte. 1999. Sequence characterisation of microsatellites in diploid and polyploid *Ipomoea*. Theoretical Applied Genetics 99:123–132.

Buteler, M.I., D.R. La Bonte, and R.E. Macchiavelli. 1997. Determining paternity in polyploids: Hexaploid simulation studies. Euphytica 96:353–361.

Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13–15.

Huamán, Z. and D.P. Zhang. 1997. Sweetpotato. In: Fuccillo, D., L. Sears, and P. Stapleton (eds.). Biodiversity in

trust. Cambridge Univ Press, Cambridge, UK. p. 29–38.
Huamán, Z, R. Ortiz, D.P. Zhang, and F. Rodriguez. 2000. Isozyme analysis of entire and core collections of *Solanum tuberosum* subsp. *andigena* potato varieties. Crop Science 40:273–276.
Jarret, R.L. and N. Bowen. 1994. Simple sequence repeats (SSR's) for sweet-potato germplasm characterization. Plant Genetic Resources Newsletter 100:9–11.
Jones A. 1965. Cytological observations and fertility measurements of sweetpotato [*Ipomoea batatas* (L.) Lam]. Proceedings of the American Society of Horticultural Science 86:527–537.
Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Science USA 70:3321–3323.
Nishiyama, I. 1982. Autopolyploidy evolution of the sweetpotato. In: Villareal, R.L. and T.D. Griggs (eds.). Sweet potato, proceedings of first international symposium. AVRDC Publ. No. 82–172, Tainan, Taiwan, China. p. 263–274.
Powell, W., G.B. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (micro-satellite) markers for germplasm analysis. Molecular Breeding 2:225–238.
Shiotani, I. and T. Kawase. 1989. Genomic structure of sweet potato and hexaploids in *Ipomoea trifida* (H.B.K.) Don. Japan Journal of Breeding 39:57–66.
Weber, J.L. and P.E May. 1989. Abundant class of human DNA polymorphism which can be typed using the polymerase chain reaction. American Journal of Human Genetics 44:388–386.
Zhang, D.P., J. Cervantes, Z. Huamán, and M. Ghislain. 2000. Assessing genetic diversity of sweet potato (*Ipomoea batatas* (L.) Lam.) varieties from tropical America using AFLP. Genetic Resources and Crop Evolution 47:659–665.
Zhang, D.P., M. Ghislain, Z. Huamán, A. Golmirzaie. and R.J. Hijmans. 1998. RAPD variation in sweetpotato [*Ipomoea batatas* (L.) Lam] varieties from South America and Papua New Guinea. Genetic Resources and Crop Evolution 45:271–277.
Zhang, D.P., Z. Huaman, F. Rodriguez, and M. Ghislain. 2001. Identifying duplicates in sweetpotato [*Ipomoea batatas* (L.) Lam] varieties using RAPD. Acta Horticulturae 546:535–541.