Theor Appl Genet (2004) 108:881–890
DOI 10.1007/s00122-003-1494-7

ORIGINAL PAPER

M. Ghislain · D. M. Spooner · F. Rodríguez ·
F. Villamón · J. Núñez · C. Vásquez · R. Waugh ·
M. Bonierbale

Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato

Received: 7 April 2003 / Accepted: 10 October 2003 / Published online: 27 November 2003
© Springer-Verlag 2003

Abstract Characterization of nearly 1,000 cultivated potato accessions with simple sequence repeats (SSRs; also referred to as microsatellites) has allowed the identification of a reference set of SSR markers for accurate and efficient genotyping. In addition, 31 SSRs are reported here for a potato genetic map, including new map locations for 24 of them. A first criterion for this proposed reference set was ubiquity of the SSRs in the eight landrace cultivar groups of the potato, *Solanum tuberosum*. All SSRs tested in the present study displayed the same allele phenotypes and allele size range in the diverse germplasm set as in the advanced potato cultivar germplasm in which they were originally discovered. Thirteen of 13 SSR products from all cultivar groups are shown to cross-hybridize with the corresponding SSR product of the source cultivar to ascertain sequence homology. Other important SSR selection criteria are quality of amplification products, locus complexity, polymorphic index content, and well-dispersed location on a potato genetic map. Screening of 156 SSRs allowed the identification of a highly informative and user-friendly set comprising 18 SSR markers for use in characterization of potato genetic resources. In addition, we have identified true- and pseudo-multiplexing SSRs for even greater efficiency.

Communicated by F. Salamini

M. Ghislain (✉) · F. Rodríguez · F. Villamón · J. Núñez ·
C. Vásquez · M. Bonierbale
International Potato Center,
P.O. Box 1558, Lima 12, Peru
e-mail: m.ghislain@cgiar.org
Tel.: +51-1-3496017
Fax: +51-1-3175326

D. M. Spooner
U.S.D.A., Agricultural Research Service,
Vegetable Crops Research Unit, Department of Horticulture,
University of Wisconsin,
Madison, WI 53706–1590, USA

R. Waugh
Scottish Crop Research Institute,
Invergowrie, Dundee, DD2 5DA, UK

Present address:
F. Rodríguez, U.S.D.A., Agricultural Research Service,
Vegetable Crops Research Unit, Department of Horticulture,
University of Wisconsin,
Madison, WI 53706–1590, USA

Present address:
F. Villamón, U.S.D.A., Agricultural Research Service,
Vegetable Crops Research Unit, Department of Horticulture,
University of Wisconsin,
Madison, WI 53706–1590, USA

Introduction

The potato collection held in trust at the International Potato Center (CIP; Lima, Peru) is comprised of landrace cultivars of *Solanum tuberosum* and 128 wild potato species, represented by a total of 5,094 clonal and true seed accessions (Huamán et al. 1997). The landrace potatoes are extremely diverse, ranging from diploids to pentaploids, and were collected in a range of agroecologies ranging from 1,000 m to 4,000 m above sea level and from 10° north latitude to 42° south latitude. The use of this gene pool in breeding programs contributes to the improvement of the potato for current and future needs. Documentation of genetic identity in such ex situ collections fulfills various curatorial and utilization needs: safeguarding of original types in germplasm repositories, determining varietal distinctness and pedigree and variety registration. The cultivated potato collection at CIP has been extensively characterized for morphological descriptors, taxonomy, disease and pest resistances and isozyme patterns (Huamán et al. 2000a, 2000b; Huamán and Spooner 2002) and partially characterized using random amplified polymorphic DNA markers (RAPD) (Ghislain et al. 1999a). Our interest was to complement these data with descriptive information that is neutral in terms of environmental influences and conducive to quantitative estimates of genetic similarity and distance.

Marker systems that are rapid, reliable, informative and relatively simple are continually sought for practical applications in germplasm conservation, management and enhancement. Simple sequence repeats (SSRs), or microsatellites, are tandem repeats of 1–5 nucleotides whose repeat number reveals genetic differences among individuals. Alleles most likely arise through DNA polymerase slippage by adding and subtracting repeat units but also through point mutations (Kruglyak et al. 1998). SSRs have the advantage of being based on simple PCR assays, are highly polymorphic, co-dominant and inexpensive to use (Powell et al. 1996; Jones et al. 1997). SSRs also provide an appropriate technology for laboratories and breeding programmes in developing countries with modest funding and expertise in molecular techniques. Additional advantages of SSRs over dominant marker systems in potato are the capacity to reflect ploidy status and their high heterozygosity, which facilitate the comparison of genetic maps based on different populations (Milbourne et al. 1998; Ghislain et al. 2001b).

SSRs present two major disadvantages. First, they have high development costs associated with the identification of DNA sequences flanking SSRs and the development and testing of reliable primers. Second, microsatellites are often useful for only closely related germplasm sources, and even moderately divergent cross-species amplification can lead to false positives and provide significant distortion in genetic similarity estimates (Peakall et al. 1998; Westman and Kresovich 1998). This was demonstrated in potato when SSRs developed for modern cultivars worked very well in a cultivated species gene pool (Raker and Spooner 2002) but produced limited amplification and clearly distorted phylogenetic information in germplasm from another phylogenetic clade of tuber-bearing *Solanum* (Lara-Cabrera and Spooner 2003). However, once SSRs are identified, their high allele and genetic information content make them a highly desirable system for fingerprinting large collections of related accessions, and the system also is amenable to automation (Mitchell et al. 1997).

SSR technology has been developed for potato using DNA from modern cultivars of *S. tuberosum*. The first generation of SSRs in potato was obtained from the identification of specific repeat motifs in gene sequences (Veilleux et al. 1995; Kawchuk et al. 1996; Provan et al. 1996; Schneider and Douches 1997). The second wave of SSRs came from screening genomic libraries enriched for repeat motifs (Milbourne et al. 1998). More recently, as reported in the present publication, the search for repeat motifs within expressed sequence tags (ESTs) from potato showed that 5% of ESTs evaluated contained SSRs. It is crucial to address the applicability of these SSRs across the various cultivar groups in potato in order to convert these markers into practical genetic tools for germplasm conservation and management and genetic mapping, and in particular to exploit the cultivated Andean gene pool. In the present report, the degree of applicability across cultivar groups and the polymorphic index content (PIC) of SSRs derived from modern cultivars of *S. tuberosum* are investigated. We use these results to propose a highly informative and user-friendly set of SSRs for cultivated potato fingerprinting, diversity assessment and phylogenetic studies.

Cultivated potato taxonomy has been especially difficult due to the likely hybrid origins, multiple origins and evolutionary dynamics of continuing hybridization. The most commonly used taxonomic system recognized seven cultivated species and seven subspecies (Hawkes 1990). Recent phenetic studies, based on morphology combined with a summary of prior crossing and molecular data, showed little or no support for these taxa and led to the reclassification of cultivated landraces (i.e., all “taxa” except modern cultivars of *S. tuberosum* subsp. *tuberosum*) as a single species, *S. tuberosum* (Huamán and Spooner 2002). Within *S. tuberosum,* eight cultivar groups were recognized: Ajanhuiri, Andigenum, Chaucha, Chilotanum (previously referred to as *S. tuberosum* subsp. *tuberosum* from Chile), Curtilobum, Juzepczukii, Phureja, and Stenotomum. These authors deferred classification of modern cultivars, traditionally classified in subsp. *tuberosum* for a later study.

Materials and methods

Germplasm and DNA extraction

A total of 931 accessions in eight taxonomic groups of cultivated potato germplasm held at CIP were characterized with SSRs (Table 1). This material had been previously cleared of redundant

Table 1 Potato accessions conserved in the gene bank at the International Potato Center and currently genotyped using SSR markers

Cultivar-groups of *Solanum tuberosum*	Ploidy	Number of accessions	Genotyped to date	Number of SSR	Total number of alleles
Stenotomum group (stn)	$2n=2x=24$	268	237	9	82
Phureja group (phu)	$2n=2x=24$	170	128	22	152
Ajanhuiri group (ajh)	$2n=2x=24$	10	4	13	49
Chaucha group (cha)	$2n=3x=36$	97	4	13	41
Juzepczukii group (juz)	$2n=3x=36$	31	4	13	36
Andigenum group (adg)	$2n=4x=48$	2,644	546	4	70
Chilotanum group (cht)	$2n=4x=48$	127	4	13	47
Curtilobum group (cur)	$2n=5x=60$	11	4	13	34
Total cultivated		3,358	931		

accessions. Plants were propagated in vitro, and total DNA was extracted using standard protocols adapted at CIP (Ghislain et al. 1999b). DNA quality and quantity were estimated by comparison with CsCl-purified λ DNA (Gibco-BRL, Gaithersburg, Md.) digested with *Pst*I on ethidium bromide-stained agarose gels.

SSR sequences

SSR sequences were identified through potato database searches (Provan et al. 1996), enriched genomic libraries (Milbourne et al. 1998) and expressed sequence tags developed at the Scottish Crop Research Institute, Invergowrie, UK. Primer sequences for an original set of 108 SSRs were obtained from Milbourne et al. (1998). Forty-eight new EST-derived primers were designed (forward and reverse) using DOPE software (http://doprimer.interactiva.de/pro/frameset.html). Criteria for design were 18- to 22-oligonucleotide length, T_m between 47 and 60°C and an amplification product above 140 bp or below 250 bp. Primers were synthesized by commercial suppliers (Gibco-BRL and Genset, La Jolla, Calif.).

SSR amplification conditions

PCR reactions were performed in a 20 µl volume containing 100 mM Tris-HCl, (Sigma-Aldrich, St. Louis, Mo.) 20 mM $(NH_4)_2SO_4$ (Merck, Lima, Peru), 2.5 mM $MgCl_2$ (Merck), 0.2 mM of each dNTP (Pharmacia, Washington, DC or Amersham Biosciences, Piscataway, NJ), 0.5 µM of each primer (forward and reverse, Genset or Operon Technologies, Alameda, Calif. and Gibco-BRL), 1 unit *Taq* polymerase (Gibco-BRL) and 10 ng of genomic DNA. PCR was carried out in a PTC-100 thermocycler (MJ Research Inc., Watertown, Mass.), set to the following program: 3 min at 94°C, 2 min at annealing temperature (T_a), 1 min 30 s at 72°C, 29 cycles of 1 min at 94°C, 2 min at T_a, and 1 min 30 s at 72°C, with a final extension step of 5 min at 72°C. In some cases (indicated as Td.60–50 in Table 2), a modified PCR program was used: 3 min at 94°C, 16 double cycles of 1 min at 94°C, 2 min at 60°C, 1.5 min at 72°C, and 1 min at 94°C, 2 min at 50°C, 1.5 min at 72°C and one final elongation cycle of 5 min at 72°C.

Amplification products were detected by polyacrylamide gel electrophoresis and silver stained using standard protocols adapted at CIP (Ghislain et al. 1999b). Allele sizes were determined by coincidence of gel-mobility between the band of the sequencing reactions (pUC18-forward primer) and either the upper band (when two are visible) or the most intense one (in the case of stutter bands) of the SSR marker.

Amplification product homology

Southern blot hybridization was conducted using the amplified SSRs from a Chilotanum group cultivar of *Solanum tuberosum* as a probe. The amplification products were isolated from agarose gel slices by centrifugation at 13,000 rpm for 20 min in an Eppendorf microcentrifuge using a SpinX tube (Costar, Cambridge, Mass.). The protocols for Southern blot and DNA detection were those provided with the ECL kit (Amersham) using stringent hybridization conditions.

Genetic localization

The ‘PD’ mapping population and corresponding molecular data set were used to localize SSR markers in the potato genome (Ghislain et al. 2001b). SSR markers were analyzed by linkage analysis of marker alleles segregating from the respective source parent following our published methodology (Ghislain et al. 2001b).

Data analysis

The PIC was calculated according to Nei’s statistic (Nei 1973): $PIC=1-\Sigma(p_i^2)$, where p_i is the frequency of the *i*th allele detected in the germplasm.

Results

SSR conservation among *Solanum tuberosum* cultivar groups

Of 156 SSR primers tested, 130 generated amplification products in a large set of cultivated potato accessions comprised of at least four genotypes from each of the eight cultivar groups of *S. tuberosum* (Table 1). SSRs produced several allele phenotypes using denaturing polyacrylamide gel electrophoresis, such as discrete bands, double or triple bands or stutter bands. Those phenotypes that represented alleles common across the groups were clearly conserved among the eight groups tested, supporting the analogous nature of the SSRs across all cultivar groups (Fig. 1). Southern blotting was performed for 13 SSRs, selected randomly, to corroborate sequence homology among amplification products (Fig. 2). Both of these analyses are concordant with the conservation of SSR loci across all cultivar groups of potato.

SSR characterization

Quality of amplification

Out of 156 SSRs tested, 22 were found to be most useful for germplasm fingerprinting. The poor quality of products from the remaining sequences was due to various factors such as poorly-designed primer sequences, intron insertion, presence of stutter bands, background amplification products and overlapping loci. Comparison of

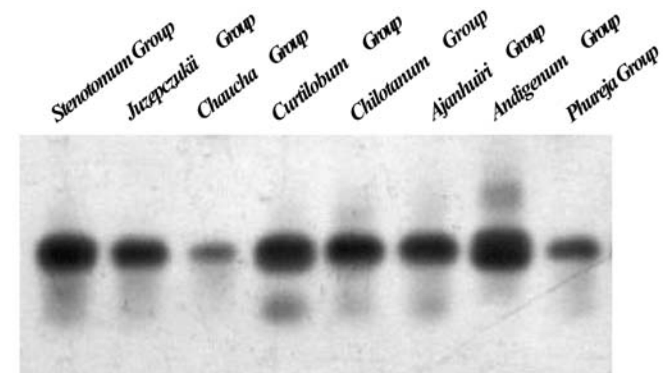


Fig. 1 Southern blot hybridization of the amplification products corresponding to microsatellite STM1049 between one accession, selected randomly, for each of the eight cultivated potato groups and the *Solanum tuberosum* product used as a probe. High stringency hybridization conditions were used

Table 2 SSR set selected by polymorphic index content per marker and location on the potato genetic map. The copy number is provided according to location on the potato genetic map and ploidy study (>1 indicates that another locus must contribute but is not distinguishable on the base of allele size or genetic mapping). In addition to the 18 SSR markers (two of which correspond to two loci), 4 additional SSR markers are provided as optional below the dashed line. *N.d.* Not determined due to lack of polymorphism in the mapping population

SCRI	Repeat	Primer sequences (5′–3′′) forward–reverse	Annealing temperature (°C)	Chrom.	Copies	Type	No. of genotypes	No. of alleles	Allele range	Multiple of repeat unit	PIC
STM1049	$(ATA)_6$	CTACCAGTTTGTTGATTGTGGTG AGGGACTTTAATTTGTTGGACG	57	I	>1	3′ UTR	315	9	184–254	Yes	0.7706
STM2022	$(CAA)_3$...$(CAA)_3$	GCGTCAGCGATTTCAGTACTA TTCAGTCAACTCCTGTTGCG	53	II	>1	Intergenic	544	13	184–244	Yes	0.7531
STM1053	$(TA)_4$ $(ATC)_5$	TCTCCCCATCTTAATGTTTC CAACACAGCATSCAGATCATC	55	III	1	3′ UTR	239	7	168–184	Yes	0.6656
STM3023a	$(GA)_9$, $(GA)_8$, $(GA)_4$	AAGCTGTTACTTGATTGCTGCA GTTCTGGCATTTCCATCTAGAGA	50	IV	1	Intergenic	279	5	169–201	Yes	0.5603
STM3023b	$(GA)_9$, $(GA)_8$, $(GA)_4$	AAGCTGTTACTTGATTGCTGCA GTTCTGGCATTTCCATCTAGAGA	50	n.d.	1	Intergenic	279	3	183–193	Yes	0.2717
STM1031	$(AT)_{13}$	TGTGTTTGTTTTTCTGTAT AATTCTATCCTCATCTCTA	55	V	1	3′ UTR	275	9	265–325	No	0.6584
STPoAc58	$(TA)_{13}$	TTGATGAAAGGAATGCAGCTTGTG ACGTTAAAGAAGTGAGAGTACGAC	57	V	1	3′ UTR	315	13	203–277	Yes	0.7033
STM0019a	$(AT)_7(GT)_{10}(AT)_4$ $(GT)_5$ $(GC)_4$ $(GT)_4$	AATAGGTGTACTGACTCTCAATG TTGAAGTAAAAGTCCTAGTATGTG	47	VI	1	Intergenic	544	27	155–241	Yes	0.8808
STM0019b	$(AT)_7(GT)_{10}(AT)_4$ $(GT)_5$ $(GC)_4$ $(GT)_4$	AATAGGTGTACTGACTCTCAATG TTGAAGTAAAAGTCCTAGTATGTG	47	n.d.	1	Intergenic	544	10	83–124	Yes	0.5919
STM0031	$(AC)_5$ ... $(AC)_3$ (GCAC) $(AC)_2$ $(GCAC)_2$	CATACGCACGCACGTACAC TTCAACCTATCATTTTGTGAGTCG	57	VII	1	Intergenic	315	11	155-205	No	0.7714
STM1052	$(AT)_{14}$ GT $(AT)_4$ $(GT)_6$	CAATTTCGTTTTTTCATGTGACAC ATGGCGTAATTTGATTTAATACGTAA	Td.60–50	VII	1	Intron	137	16	212–268	Yes	0.8320
STM2013	$(TCTA)_6$	TTCGGAATTACCCTCTGCC AAAAAAAGAACGCGCACG	55	VII	2	Intergenic	275	20	146–172	Yes	0.8728
STM1104	$(TCT)_5$	TGATTCTCTTGCCTACTGTAATCG CAAAGTGGTGTGAAGCTGTGA	57	VIII	1	3′ UTR	279	17	164–185	Yes	0.8916
STM1016	$(TCT)_9$	TTCTGATTTCATGCATGTTTCC ATGCTTGCCATGTGATGTGT	53	VIII	1	Intron	137	9	243–262	Yes	0.7757
STGBSS	$(TCT)_9$	AATCGGTGATAAATGTGAATGC ATGCTTGCCATGTGATGTGT	53	VIII	1	Intron	137	8	130–142	Yes	0.7403
STWAX-2	$(ACTC)_5$	CCCATAATACTGTCGATGAGCA GAATGTAGGGAAACATGCATGA	53	VIII	1	5′ UTR	137	8	224–243	No	0.7283
STM3012	$(CT)_4$, $(CT)_8$	CAACTCAAACCAGAAGGCAAA GAGAAATGGGCACAAAAAACA	57	IX	1	Intergenic	315	8	168–213	Yes	0.6944

Table 2 (continued)

SCRI	Repeat	Primer sequences (5′–3′) forward–reverse	Annealing temperature (°C)	Chrom.	Copies	Type	No. of genotypes	No. of alleles	Allele range	Multiple of repeat unit	PIC
STM1106	$(ATT)_{13}$	TCCAGCTGATTGGTTAGGTTG ATGCGAATCTACTCGTCATGG	55	X	1	Intron	476	15	131–197	Yes	0.8216
STM0037	$(TC)_5$ $(AC)_6$ AA $(AC)_7$ $(AT)_4$	AATTTAACTTAGAAGATTAGTCTC ATTTGGTTGGGTATGATA	53	XI	1	Intergenic	504	13	75–125	No	0.7865
STM0030	Compound (GT/GC)$(GT)_8$	AGAGATCGATGTAAAACACGT GTGGCATTTTGATGGATT	53	XII	>1	Intergenic	315	15	122–191	Yes	0.8641
STM2030	$(CA)_3$, $(TA)_5$	TCTTCCCAAATCTAGAATACATGC AAAGTTAGCATGGACAGCATTTC	55	I	1	Intergenic	275	5	180–209	No	0.2803
STM1064	$(TA)_{12}$..$(TG)_4$ GT $(TG)_5$	GTTCTTTTGGTGGTTTTCCT TTATTTCTCTGTTGTTGCTG	55	II	1	Intron	275	7	188–199	Yes	0.5189
STM1058	$(ATT)_5$	ACAATTTAATTCAAGAAGCTAGG CCAAATTTGTATACTTCAATATGA	55	III	1	3′ UTR	239	5	107–122	Yes	0.4679
STM1017	$(ATT)_5$	GACACGTTCACCATAAAA AGAAGAATAGCAAAGCAA	53	IX	1	5′ UTR	275	3	132–136		

agarose gel (3%) and polyacrylamide gel electrophoresis revealed that allele resolution could be as little as 10 bp (and no less) in agarose gels (Fig. 2). Three percent agarose gels provided enough resolution to distinguish about 60% of the potato samples tested in this study, but for genetic identification purposes, complete resolution of up to 1 bp present on polyacrylamide gels is needed.

Genome coverage

A criterion for selecting the set of 22 SSR is that it will provide genetic information for all 12 chromosomes at least once, a criterion met in our set of 22 markers (Fig. 3). At the DNA sequence level, 75 SSRs were from untranscribed regions whereas 55 were from transcribed regions.

Number of alleles

The number of alleles detected per SSR at single loci across the 935 cultivated accessions ranged from 3 to 27 (Table 2). For the 130 SSRs with clear amplification products, we analyzed polymorphism and SSR location at the gene structure level, and polymorphism was observed more frequently with SSRs found in untranscribed than in transcribed DNA (Table 3). A χ^2 test was applied to determine independence of polymorphism and location in transcribed vs. untranscribed regions. The hypothesis of independence is rejected (P=0.024). When a similar assessment was made within genes, grouping coding sequences with EST, the hypothesis of independence was again rejected (P=0.028). Therefore, the 12 SSRs that have their repeat sequence within the coding region were omitted from the group of selected SSRs because they would detect polymorphism among potato genotypes less frequently.

Allele size and repeat motif relationships

When sample size allowed us to conduct such an analysis, we found that discrete allele size differences were directly proportional to the nucleotide repeat length. This was the case for 16 SSRs out of 22 analyzed with simple motif repeats (indicated in Table 2). The rare exception alleles (not fitting with a multiple of the repeat motif) may have been generated by mechanisms other than polymerase slippage.

Polymorphic index content

About half (63 of 130) of the SSRs tested proved to be polymorphic in this large germplasm sample. PIC values were calculated for 22 microsatellites from a total number of genotypes ranging from at least 137 and up to 544

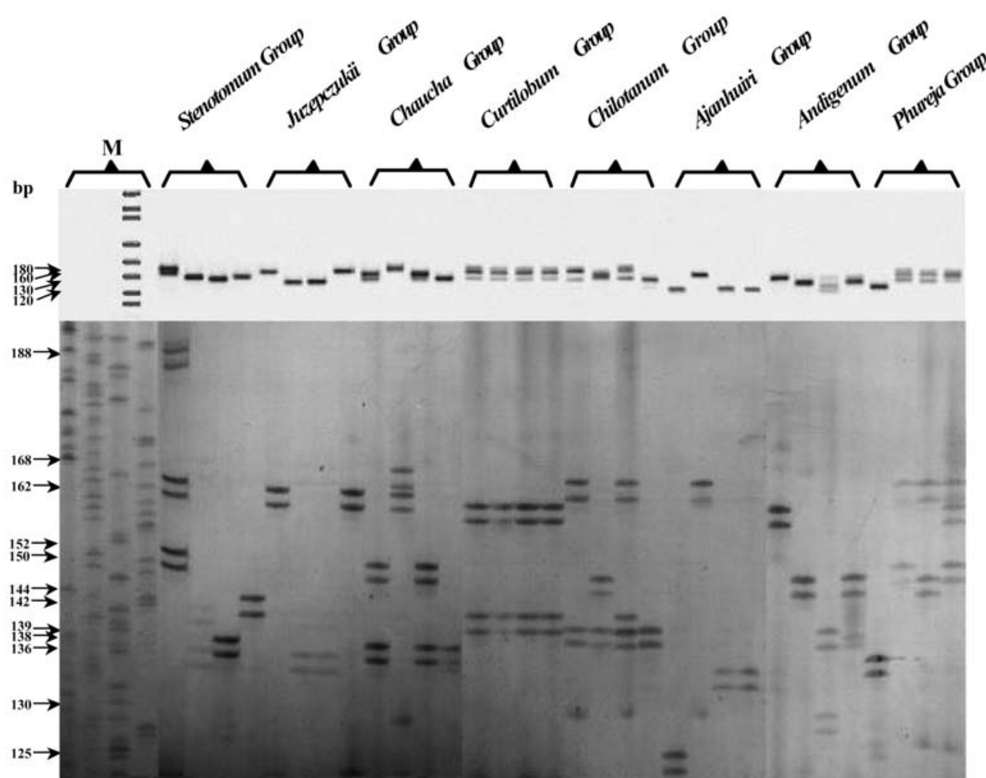


Fig. 2 Amplification products (range: 125 to 190 bp) of 36 cultivated potato accessions with microsatellite STM0030 resolved by electrophoresis on a 3% agarose gel (upper picture) and a 6% and 7 M urea polyacrylamide gel. The *M* label stands for molecular weight standards as described in the Materials and methods section

Table 3 Test of independence of polymorphism and SSRs location within or between genes. 130 SSRs with clear amplification products were considered here on the entire cultivated potato germplasm fingerprinted. *UTR* Untranslated region, *CdS* coding sequence, *EST* expressed sequence tag

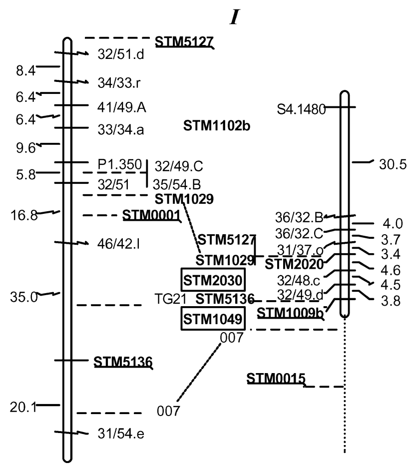
SSR location	3′ UTR[a]	5′ UTR[b]	CdS[c]	EST[d]	Intron[e]	Intergenic[f]	
Monomorphic	5	6	7	27	7	15	
Polymorphic	6	4	5	15	14	19	
χ^2 test		Monomorphic observed	Polymorphic observed	Monomorphic expected	Polymorphic expected	*P*	Chi
Ho:	Transcribed (a+b+c+d)	45	30	39	36		
	Untranscribed (e+f)	22	33	28	27	0.024	5.082
Ho:	Coding (c+d)	34	20	28	26		
	Non-coding (a+b+e+f)	33	43	39	37	0.028	4.827

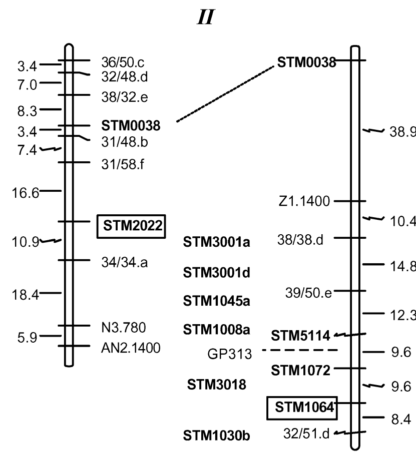
accessions from the eight cultivar groups. PIC values per SSR varied from 0.280 to 0.892 (Table 2).

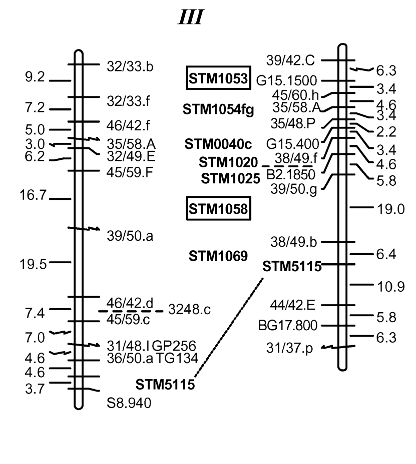
Genetic mapping of SSRs

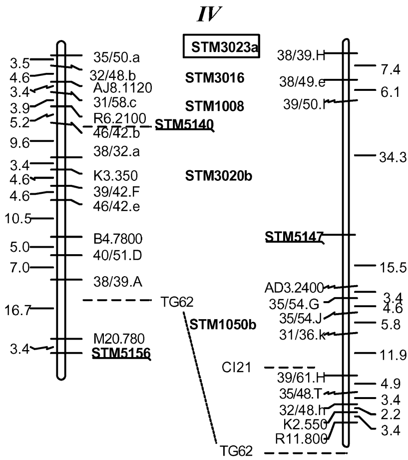
One hundred thirty SSRs were surveyed for polymorphism in the PD mapping population. Sixty-one were polymorphic, 57 of which segregated according to the 1:1 ratio expected for heterozygous loci in F_1 progeny. Thirty SSRs previously mapped in other populations by Milbourne et al. (1998) displayed identical map locations on the PD mapping population, with the identification of new, secondary loci in four cases. Twenty-seven additional SSRs were mapped in PD (Fig. 3), which includes seven already placed on the PD map (Ghislain et al. 2001b). Segregation patterns in the mapping population suggested a total of 21 null alleles for the 57 mapped markers. Null alleles were more often found from the

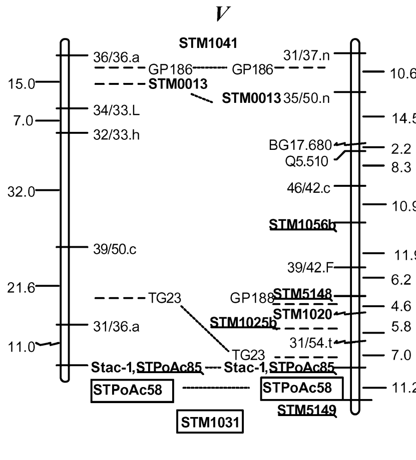
Fig. 3 Location of microsatellite markers on the PD genetic maps (Ghislain et al. 2001b). All microsatellite markers are indicated in **bold** characters. Map positions are indicated with a *plain line* for placement at LOD>2 and *dashed line* for a lower significance. Microsatellites that are *underlined* are those with a new map position. The microsatellites not mapped in PD are indicated at an approximate position between the maternal and paternal linkage groups by comparing published maps. *Boxed* microsatellites are those selected for the potato genetic identification set

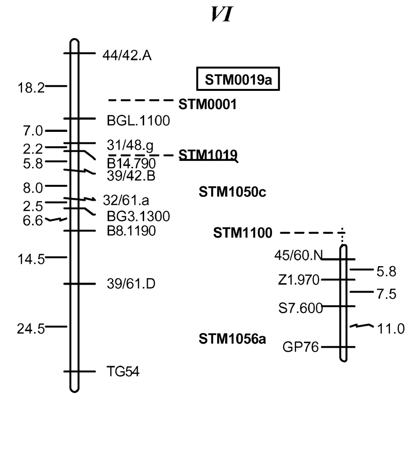

I
STM5127
32/51.d
8.4
34/33.r
6.4
41/49.A
6.4
33/34.a
STM1102b
S4.1480
9.6
P1.350
32/49.C
5.8
32/51
35/54.B
30.5
STM1029
16.8
STM0001
36/32.B
4.0
36/32.C
46/42.l
STM5127
3.7
STM1029
31/37.o
3.4
STM2020
4.6
STM2030
32/48.c
4.5
32/49.d
35.0
TG21
STM5136
3.8
STM1049
STM1009b
007
STM5136
STM0015
20.1
007
31/54.e

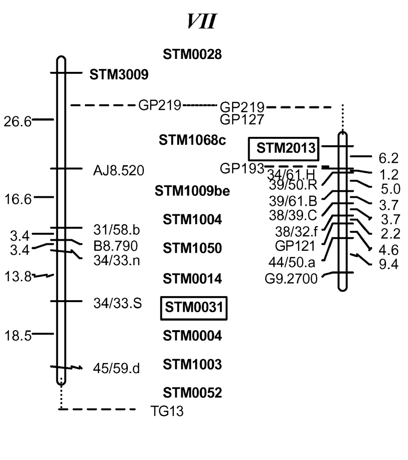

II
3.4
36/50.c
STM0038
32/48.d
7.0
38/32.e
8.3
STM0038
3.4
38.9
31/48.b
7.4
31/58.f
16.6
Z1.1400
STM2022
10.4
10.9
STM3001a
38/38.d
34/34.a
14.8
STM3001d
18.4
39/50.e
STM1045a
12.3
N3.780
STM1008a
STM5114
5.9
AN2.1400
GP313
9.6
STM1072
STM3018
9.6
STM1064
8.4
STM1030b
32/51.d

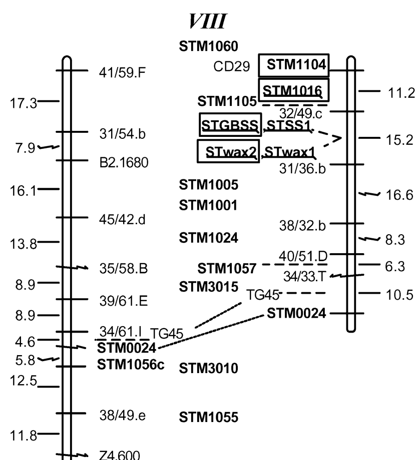

III
39/42.C
32/33.b
6.3
STM1053
9.2
G15.1500
3.4
45/60.h
7.2
32/33.f
STM1054fg
35/58.A
4.6
35/48.P
3.4
5.0
46/42.f
2.2
3.0
35/58.A
STM0040c
G15.400
3.4
6.2
32/49.E
STM1020
38/49.f
4.6
45/59.F
STM1025
B2.1850
5.8
16.7
39/50.g
STM1058
19.0
39/50.a
38/49.b
STM1069
6.4
19.5
STM5115
10.9
46/42.d
3248.c
7.4
44/42.E
45/59.c
5.8
BG17.800
7.0
6.3
31/48.l
GP256
4.6
31/37.p
36/50.a
TG134
4.6
STM5115
3.7
S8.940

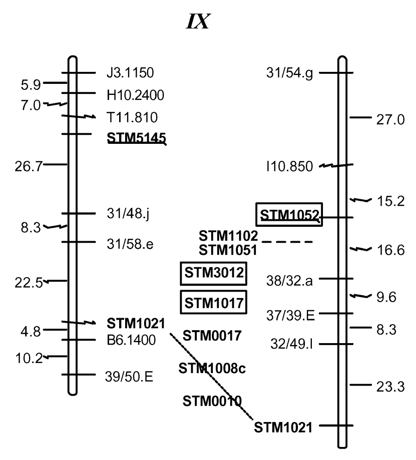

IV
STM3023a
35/50.a
3.5
38/39.H
7.4
32/48.b
4.6
STM3016
38/49.e
AJ8.1120
3.4
6.1
31/58.c
3.9
STM1008
39/50.l
R6.2100
5.2
STM5140
46/42.b
9.6
38/32.a
34.3
3.4
K3.350
4.6
STM3020b
39/42.F
4.6
46/42.e
10.5
B4.7800
STM5147
5.0
40/51.D
15.5
7.0
38/39.A
AD3.2400
3.4
TG62
35/54.G
4.6
16.7
STM1050b
35/54.J
5.8
31/36.k
M20.780
11.9
3.4
STM5156
CI21
39/61.H
4.9
35/48.T
3.4
32/48.h
2.2
K2.550
3.4
R11.800
TG62

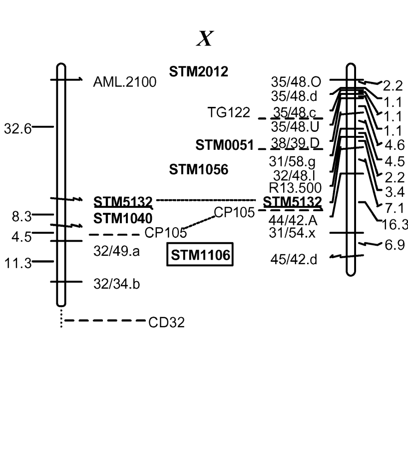

V
STM1041
36/36.a
31/37.n
GP186
GP186
10.6
15.0
STM0013
STM0013
35/50.n
34/33.L
14.5
7.0
32/33.h
BG17.680
2.2
Q5.510
8.3
46/42.c
32.0
10.9
STM1056b
39/50.c
11.9
39/42.F
6.2
21.6
TG23
GP188
STM5148
4.6
STM1020
31/36.a
STM1025b
5.8
TG23
31/54.t
11.0
7.0
Stac-1,STPoAc85
Stac-1,STPoAc85
STPoAc58
STPoAc58
11.2
STM5149
STM1031

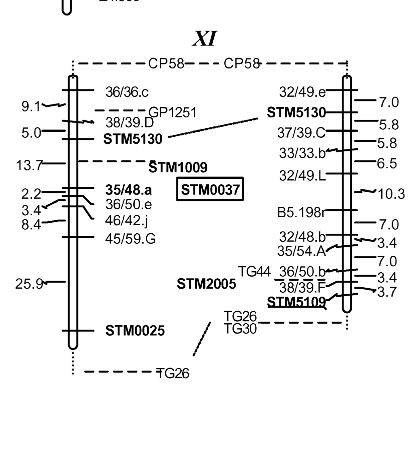

VI
44/42.A
STM0019a
18.2
STM0001
BGL.1100
7.0
2.2
31/48.g
5.8
B14.790
STM1019
39/42.B
8.0
STM1050c
32/61.a
2.5
BG3.1300
6.6
B8.1190
STM1100
45/60.N
14.5
5.8
Z1.970
39/61.D
7.5
S7.600
24.5
11.0
STM1056a
GP76
TG54

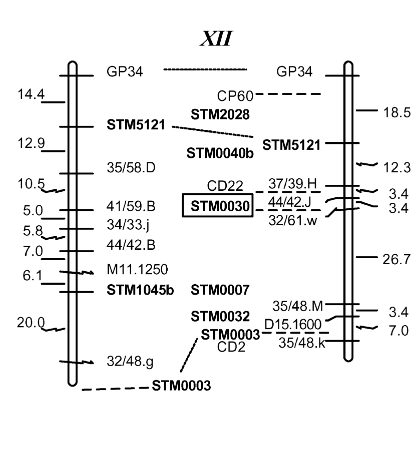

VII
STM0028
STM3009
GP219
GP219
26.6
GP127
STM1068c
STM2013
6.2
AJ8.520
GP193
34/61.H
1.2
39/50.R
5.0
STM1009be
16.6
39/61.B
3.7
38/39.C
STM1004
3.7
38/32.f
2.2
3.4
31/58.b
GP121
4.6
3.4
B8.790
STM1050
44/50.a
9.4
34/33.n
G9.2700
13.8
STM0014
34/33.S
STM0031
18.5
STM0004
45/59.d
STM1003
STM0052
TG13


VIII
STM1060
STM1104
CD29
41/59.F
STM1016
11.2
STM1105
17.3
32/49.c
STGBSS
STSS1
31/54.b
15.2
STwax2
STwax1
7.9
B2.1680
31/36.b
STM1005
16.1
16.6
STM1001
45/42.d
38/32.b
STM1024
8.3
13.8
40/51.D
STM1057
6.3
35/58.B
34/33.T
8.9
STM3015
TG45
10.5
39/61.E
STM0024
8.9
34/61.I
TG45
4.6
STM0024
5.8
STM1056c
STM3010
12.5
38/49.e
STM1055
11.8
Z4.600


IX
J3.1150
31/54.g
5.9
H10.2400
7.0
T11.810
27.0
STM5145
26.7
I10.850
15.2
STM1052
31/48.j
8.3
STM1102
31/58.e
STM1051
16.6
STM3012
38/32.a
22.5
STM1017
9.6
37/39.E
4.8
STM1021
8.3
B6.1400
STM0017
32/49.I
10.2
STM1008c
39/50.E
23.3
STM0010
STM1021


X
STM2012
AML.2100
35/48.O
2.2
35/48.d
1.1
TG122
35/48.c
1.1
35/48.U
1.1
32.6
STM0051
38/39.D
4.6
STM1056
31/58.g
4.5
32/48.I
2.2
R13.500
3.4
STM5132
STM5132
7.1
8.3
STM1040
CP105
44/42.A
16.3
4.5
CP105
31/54.x
6.9
32/49.a
STM1106
11.3
45/42.d
32/34.b
CD32


XI
CP58
CP58
36/36.c
32/49.e
7.0
9.1
GP1251
STM5130
38/39.D
5.8
5.0
STM5130
37/39.C
5.8
33/33.b
6.5
13.7
STM1009
32/49.L
2.2
35/48.a
STM0037
10.3
3.4
36/50.e
8.4
46/42.j
7.0
B5.198r
45/59.G
32/48.b
3.4
35/54.A
7.0
TG44
36/50.b
3.4
25.9
STM2005
38/39.F
3.7
STM5109
TG26
STM0025
TG30
TG26


XII
GP34
GP34
14.4
CP60
18.5
STM2028
STM5121
12.9
STM5121
35/58.D
STM0040b
12.3
10.5
CD22
37/39.H
3.4
41/59.B
STM0030
44/42.J
3.4
5.0
34/33.j
32/61.w
5.8
44/42.B
7.0
26.7
6.1
M11.1250
STM1045b
STM0007
35/48.M
3.4
STM0032
20.0
D15.1600
7.0
STM0003
35/48.k
CD2
32/48.g
STM0003

Phureja group-derived parent than from the dihaploid of an advanced potato cultivar (13 and 8, respectively).

Identification of a SSR set for potato genotyping

Out of the 22 SSRs most extensively characterized in the present study, a group of 18 can be recommended as most informative for genotyping cultivated potatoes, based on quality criteria, genome coverage, and locus-specific information content (Table 2). In order to improve the applicability of this set of 18 SSRs, we tested the multiplexing several SSRs in the same reaction mixture. Two pairs (STM3023+STM0030; STPoAc58+STM3012) can be used as a true multiplex reaction due to their compatible annealing temperature and non-overlapping alleles. Furthermore, pseudo-multiplex (separate reactions but simultaneous loading) was more successful with two triplex (STM2013+STM1052+STM0037; STM1104+STGBSS+STPoAc58) and three duplex (STM0019+STM1031; STM1106+STWAX-2; STM0030+STM1016) combinations.

Discussion

A variety of molecular markers have been successfully applied to potato genetic resources at low taxonomic levels (species or closely related species groups) to examine diversity, taxonomy, and to search for effective fingerprinting tools. Each system bears advantages and disadvantages, depending on the genetic distance of the populations examined and the nature of the question addressed. We have chosen to use SSRs for potato genetic identification because of their high genetic information content, high reproducibility, and simplicity of use. These factors are appropriate for fingerprinting large germplasm collections, investigating genome introgression through breeding, and diversity and taxonomic questions. They are appropriate, cost-effective and simple tools for laboratories in developing countries with financial constraints.

The 130 SSRs with good amplification products were from gene sequences and in genomic regions with no known function. Those found in DNA devoid of any known function were significantly more polymorphic than in motifs located in the coding sequence of genes. The SSRs encompassing non-coding sequences were considered for the subsequent selection of the set because of increased polymorphism and because of putative selective neutrality. Untranslated (5′ and 3′) and intron sequences did not show significant differences with intergenic DNA for the presence of polymorphic markers. Therefore, their inclusion was independent of whether or not they belong to a gene or not.

When these SSRs were analyzed by segregation analysis in a mapping population, we found a higher degree of null alleles from the Phureja-derived parent than for the dihaploid modern cultivar parent. This may be due to lower sequence conservation of the primer for the Phureja group-derived parent because primers were designed from modern cultivar sequences. Null alleles could not be identified in germplasm as a single allele might reflect two genetic states, heterozygous (SSR allele and a null allele) or homozygous. Therefore, a higher number of null alleles present in more distant germplasm will provide underestimates of heterozygosity. The SSRs with clear amplification products were analyzed by segregation analysis using the framework maps derived from the PD population. Twenty-seven new SSRs could be placed on the potato genetic map and four new locations were found for previously described SSRs. This increases the coverage of the potato genome with SSR markers.

In the present study, three of the eight taxonomic groups were characterized with a large sample. We found differences between groups for the particular SSRs studied. The Andigena group revealed on average a higher number of alleles per SSR and a broader range of allele size compared to the modern cultivars. However, the unequal sample size and number of SSRs assayed prohibit comparison of heterozygosity values among cultivar groups.

An SSR set for potato genetic identification has been selected using the highest PIC value, broad genome coverage, low copy number and high quality of amplification as selection criteria. This set consists of 18 SSRs covering all chromosomes. We have provided an additional optional group of four SSRs with lower PIC values that will be replaced as new optimal SSRs are discovered and characterized.

This potato genetic identification (PGI) kit will have many uses, and a wide range of genetic resources conservation issues may now be studied through this simple and inexpensive genotyping technique. First, genetic relationships within and between germplasm collections may be established. SSRs can help to estimate genetic distance of germplasm pools and genetic gaps in these collections by comparison with natural populations or farmer-maintained varieties. SSR-based studies of ex situ collections will provide genetic data useful for the identification of the smallest sub-set encompassing the entire SSR allelic diversity at these 18 loci. Such core collections will facilitate the evaluation large gene bank holdings for useful traits and potentially the detection of marker associations.

SSRs will be useful to identify redundancies between collections held at different repositories. Such a project has already been successfully approached with RAPDs, to show differences in reputed duplicate populations between the US and Russian potato gene banks (Bamberg et al. 2001). The collection genotyping will also be an excellent reference to study the dynamics of on-farm conservation. Interesting issues include the rate of selection that farmers apply when maintaining their agrobiodiversity.

Aside from these conservation issues, the PGI kit will be also used to help manage germplasm collections. The

most important management question is the true-to-type nature of accessions held ex situ. The PGI kit will be applied, for example, at CIP for new accession acquisition to have a reference genotype of the plant material. This reference will allow us to monitor authentic identification throughout the introduction of the plant material in the collection through the long and tedious process involving tissue culture manipulations of eliminating pathogens.

The PGI kit is being used to help fingerprint breeding lines for tracking purposes and may be useful for pedigree analysis. It may also help protect varieties as one descriptor integrated among others, such as tuber morphology. However, it is not yet clear when distinctness can be established using DNA markers, and what level of DNA differences are needed to establish whether the new variety is "essentially derived."

Last but not least, the PGI kit can be used to trace past, current, and future gene flow worldwide. SSR markers have the potential to address different hypotheses for the evolution of *S. tuberosum* group Andigenum by comparing allele frequencies among other taxonomic groups and wild species. The international movement of the cultivated potato half a millennium ago can also be followed by SSR markers. We have shown in a separate study that the long-standing varieties in India are more closely related to the Chilotanum group germplasm from the Chiloe islands than to the Andean potato germplasm (Ghislain et al. 2001a). A secondary center of diversity can be studied to determine parentage of clones that were transported long ago, and can be used to validate or recognize the indigenous role in germplasm conservation. Examples are the occurrence of potato on the Canary Islands and the New-Zealand Maori germplasm, where potato genetic resource conservation is part of the indigenous culture established during post-Colombian times. The PGI kit will also be useful to evaluate natural gene flow, an important issue in the context of the preservation of agrobiodiversity.

Acknowledgements The authors are grateful to Drs. Zosimo Huamán, William Roca, Rene Gomez, Ana Panta and Judith Toledo for providing the experimental materials. We also thank Daniel Andrade, Olenka Frías, Mercedes Ames, Giovana Perazzo and María del Rosario Herrera for polymorphism index content on subsets of the germplasm collection and help in mapping SSR markers; our collaborators at the SCRI, Dan Milbourne and Walter De Jong for providing SSR sequences and EST sequences. This work was made possible through grants from the United Nations Development Special Project Program, the Bundesministerium für Wirtschaftliche Zusammenarbeit und Entwicklung (BMZ) and the Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) (project no 96.7860.8-001.00).

References

Bamberg JB, Kiru D, Del Rio AH (2001) Comparison of reputed duplicate populations in the Russian and potato gene banks using RAPD markers. Am J Potato Res 78:365–369

Ghislain M, Zhang D, Fajardo D, Huamán Z, and Hijmans R (1999a) Marker-assisted sampling of the cultivated Andean potato *Solanum phureja* collection using RAPD markers. Genet Res Crop Evol 46:547–555

Ghislain M, Zhang D, Herrera MR (eds) (1999b) Molecular biology laboratory protocols: plant genotyping training manual, 2nd edn. International Potato Center (CIP), Lima, Peru

Ghislain M, Rodríguez F, Núñez J, Naik P, Huamán Z, and Bonierbale M (2001a) Comparison of genetic diversity of potato varieties from India and South America. Proceedings of the Plant and Animal Genome Conference IX, January 2001. Poster 287, p 132

Ghislain M, Trognitz B, Herrera MR, Solis J, Casallo G, Vasquez C, Hurtado O, Castillo R, Portal L, and Orrillo, M (2001b) Genetic loci associated with field resistance to late blight in offspring of *Solanum phureja* and *tuberosum* grown under short-day conditions. Theor Appl Genet 103:433–442

Hawkes JG (1990) The potato: evolution, biodiversity and genetic resources. Belhaven, Washington, DC

Huamán Z, Spooner DM (2002) Reclassification of landrace populations of cultivated potatoes (*Solanum* sect. *Petota*). Am J Bot 89:947–965

Huamán Z, Golmirzaie A, Amoros W (1997) The Potato. In: Fuccillo D, Sears L, Stapleton P (eds), Biodiversity in trust. Cambridge University Press, Cambridge, pp 21–28

Huamán Z, Ortiz R, Gómez R (2000a) Selecting a *Solanum tuberosum* subsp. *andigena* core collection using morphological, geographical, disease and pest descriptors. Am J Potato Res 77:183–190

Huamán Z, Ortiz R, Zhang D, Rodríguez F (2000b) Isozyme analysis of entire and core collections of *Solanum tuberosum* subsp. *andigena* potato cultivars. Crop Sci 40:273–276

Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevschi A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vasquez A, Karp A (1997) Reproducibility testing of RAPD, AFLP, and SSR markers in plants by a network of European laboratories. Mol Breed 3:381–390

Kawchuk LM, Lynch DR, Thomas J, Penner B, Sillito D, Kulcsar F (1996) Characterization of *Solanum tuberosum* simple sequence repeats and application to potato cultivar identification. Am Potato J 73:325–335

Kruglyak S, Durrett RT, Schug MD, and Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci USA 95:10774–10778

Lara-Cabrera S, Spooner DM (2003) Taxonomy of Mexican diploid wild potato (*Solanum* sect. *Petota*) species: morphological and microsatellite data. Monogr Syst Bot Moi Bot Gard (in press)

Milbourne D, Meyer RC, Collins AJ, Ramsay LD, Gebhardt C, Waugh R (1998) Isolation, characterization and mapping of simple sequence repeat loci in potato. Mol Gen Genet 259:233–245

Mitchell SE, Kresovich S, Jester CA, Hernandez CJ, Szewc-McFadden AK (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. Crop Sci 37:617–624

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Peakall R, Gilmore S, Keys W, Morgante M, Rafalski A (1998) Cross-species amplification of soybean (*Glycine max*) simple sequence repeats (SSRs) within the genus and other legume genera: Implications for the transferability of SSRs in plants. Mol Biol Evol 15:1275–1287

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Provan J, Powell W, Waugh R (1996) Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). Theor Appl Genet 92:1078–1084

Raker C, Spooner DM (2002) The Chilean tetraploid cultivated potato, *Solanum tuberosum*, is distinct from the Andean populations; microsatellite data. Crop Sci 42:1451–1458

Schneider K, Douches DS (1997) Assessment of PCR-based simple sequence repeats to fingerprint North American potato cultivars. Am Potato J 74:149–160

Veilleux RE, Shen LY, Paz MM (1995) Analysis of the genetic composition of anther-derived potato by randomly amplified polymorphic DNA and simple sequence repeats. Genome 38:1153–1162

Westman AL, Kresovich S (1998) The potential of cross-taxa simple-sequence repeat (SSR) amplification between *Arabidopsis thaliana* L. and crop brassicas. Theor Appl Genet 96:272–281