Mol Breeding (2009) 23:377–388
DOI 10.1007/s11032-008-9240-0

Robust and highly informative microsatellite-based genetic identity kit for potato

Marc Ghislain · Jorge Núñez · María del Rosario Herrera · José Pignataro · Frank Guzman · Merideth Bonierbale · David M. Spooner

Received: 27 March 2008 / Accepted: 17 November 2008 / Published online: 12 December 2008

Abstract The fingerprinting of 742 potato landraces with 51 simple sequence repeat (SSR, or microsatellite) markers resulted in improving a previously constructed potato genetic identity kit. All SSR marker loci were assayed with a collection of highly diverse landraces of all species of cultivated potato with ploidies ranging from diploid to pentaploid. Loci number, amplification reproducibility, and polymorphic information content were recorded. Out of 148 SSR markers of which 30 are new, we identified 58 new SSR marker locations on at least one of three potato genetic linkage maps. These results permitted the selection of a new potato genetic identity kit based on 24 SSR markers with two per chromosome separated by at least 10 cM, single locus, high polymorphic information content, and high quality of amplicons as determined by clarity and reproducibility. The comparison of a similarity matrix of 742 landraces obtained with the 24 SSR markers of the new kit and with the entire dataset of 51 SSR markers showed a high correlation ($r = 0.94$) by a Mantel test and even higher correlations ($r = 0.99$) regarding topological comparisons of major branches of a neighbor joining tree. This new potato genetic identity kit is able to discriminate 93.5% of the 742 landraces compared to 98.8% with 51 SSR markers. In addition, we made a marker-specific set of allele size standards that conveniently and unambiguously provide accurate sizing of all alleles of the 24 SSR markers across laboratories and platforms. The new potato genetic identity kit will be of particular utility to standardize the choice and allele sizing of microsatellites in potato and aid in collaborative projects by allowing cumulative analysis of independently generated data.

Keywords Microsatellite · Potato · *Solanum tuberosum* · SSR

Abbreviations

CIP	Centro International de la Papa (International Potato Center)
cM	Centimorgans
EST	Expressed sequence tags
PGI kit	Potato genetic identity kit
PIC	Polymorphic information content
SSR	Simple sequence repeat
T^a	Annealing temperature
TIGR	The Institute of Genomic Research

Electronic supplementary material The online version of this article (doi:10.1007/s11032-008-9240-0) contains supplementary material, which is available to authorized users.

M. Ghislain (✉) · J. Núñez · M. del Rosario Herrera · J. Pignataro · F. Guzman · M. Bonierbale
Applied Biotechnology Laboratory, International Potato Center, (CIP), P.O. Box 1558, Lima 12, Peru
e-mail: m.ghislain@cgiar.org

D. M. Spooner
USDA-ARS, Vegetable Crops Research Unit, Department of Horticulture, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706-1590, USA

Introduction

Simple sequence repeats (SSRs, or microsatellites) have been used to great advantage in potato for studies of diversity, genetic structure, and classification (Spooner et al. 2007); tracing germplasm migrations (Spooner et al. 2005a; Rios et al. 2007); fingerprinting (Moisan-Thiery et al. 2005; Provan et al. 1996; Schneider and Douches 1997); genetic linkage mapping (Ghislain et al. 2001; Feingold et al. 2005); establishment of core collections (Ghislain et al. 2006); and investigations of duplicate collections across genebanks (Del Rio et al. 2006). Although not yet used in potato they have potential applications in studies of linkage disequilibrium (Remington et al. 2001; Stich et al. 2005) and gene flow (Devaux et al. 2005; Fenart et al. 2007). They require considerable developmental costs and often have maximum utility within a narrow range of germplasm from which they were developed. Once developed, however, they have tremendous advantages over many other marker classes to include low operational costs, codominance, hypervariablity, high quality bands, highly reproducible bands, amenability to automation, ease of multiplexing, and use with low quality DNA (Spooner et al. 2005b).

Over 200 potato SSRs have been identified through enriched genomic libraries and database searches of expressed sequence tags (ESTs) (Milbourne et al. 1998; Ashkenazi et al. 2001; Ghislain et al. 2004; Feingold et al. 2005). However, many more are becoming available as ESTs are being identified. The latest SSR summary statistics from the former The Institute for Genomic Research (TIGR) document more than 5,800 sequences with potentially useful SSR (repeats of 2–6 nucleotides) markers for potato. These SSRs differ greatly, however, in quality (clarity and repeatability of bands), map location, and polymorphism. Some of them have been tested on potato landraces and advanced varieties mapped on various potato genetic maps. However, an extensive analysis on a large collection of potato SSRs was lacking. Ghislain et al. (2004) provided the first such analysis of 156 SSRs for quality and polymorphism, chose 22 of them by a combination of the above criteria in cultivated potato, showed how some of these could be multiplexed, and mapped them.

The purpose of the present study is to screen additional potato SSRs from all taxonomic groups of potato to refine a selection of microsatellites for maximum utility in a cultivated potato background. Such large data set is available from a previous study aiming at classifying cultivated potato (Spooner et al. 2007). Its wide genetic diversity makes this data set particularly valuable for our purpose.

Materials and methods

Plant materials and DNA extraction

Seven hundred and forty-two native (landrace) potatoes belonging to a composite genotyping set of potato at the International Potato Center (CIP) were used for this study. These landraces were selected to represent all four species and taxonomic groups of potato as described above, and represent the same used in the taxonomic study of Spooner et al. (2007) and are described in the supporting dataset 1. Genomic DNA was obtained using standard protocol derived from Doyle and Doyle (1990). DNA concentration was estimated by using a TBS-380 Fluorometer (Turner BioSystems, USA) with Pico-Green® reagent and 500 ng/ml salmon DNA as reference.

Microsatellite markers and PCR conditions

Eighty-eight SSR markers were obtained from four sources: (1) 22 belong to the previously identified potato genetic identity (PGI) kit (Ghislain et al. 2004), (2) 13 from ESTs developed at the Scottish Crop Research Institute (Milbourne et al. 1998), (3) 30 identified at CIP using the potato EST database of the former The Institute for Genomic Research (http://www.tigr.org/), and (4) 23 from the University of Idaho (Feingold et al. 2005).

PCR reactions were performed in a 10 µl volume containing 100 mM Tris–HCl (Sigma), 20 mM $(NH4)_2SO_4$ (Merck), 2.5 mM $MgCl_2$ (Merck), 0.2 mM of each dNTP (Amersham), 25 pM of 700 or 800 IRDye-labeled M13 forward primer (LI-COR), 22 pM M13-tailed forward SSR primer (Invitrogen), 15 pM reverse SSR primer (Invitrogen), 1 unit of Taq polymerase, and 15 ng of genomic DNA. PCR was carried out in a PTC-100 or PTC-200 thermocycler (MJ Research Inc.) using the following cycling profiles: 4 min at 94°C; 33 cycles of 1 min at

94°C, 1 min at annealing temperature (T^a) determined experimentally for each SSR primer combination and 1 min at 72°C; with a final extension step of 4 min at 72°C. Blue Stop solution (#830-05630, LI-COR, USA) in a ratio of 1:1 was added to the PCR reaction before loading. PCR products were separated by electrophoresis on a 4300 LI-COR DNA Analyzer system. We sized alleles with the IRDye 50–350 bp fragment size ladder (LI-COR, USA). SSR alleles were detected and scored using the SAGA Generation 2 software (LI-COR, USA).

Mapping new SSR markers

Previously mapped and new SSR markers were mapped on at least one of three segregating diploid populations from which genetic maps were developed: the PD population (Ghislain et al. 2001), BCT population (Bonierbale et al. 1988) or the PCC1 population (Villamón et al. 2005). A total of 148 SSR markers were used in the present map effort and are provided in the supporting dataset 2. The segregation data of 27 SSR marker alleles located on the BCT genetic map was provided to us by the research group (Feingold et al. 2005). Marker alleles segregated as 1:1 ratio whereas skewed markers were rejected using the threshold value established for each genetic map (goodness of fit χ^2 test). Null alleles were not considered. Linkage analysis of marker alleles segregating from the respective source parent was performed using JoinMap 3.0 (Stam 1993) with a LOD score of 3.

Polymorphic information content and matrix comparison

SSR marker alleles were scored for presence or absence of the band for all 742 genotypes and treated as dominant marker. The polymorphic information content (PIC) was calculated as $\text{PIC} = 1-\sum(p_i^2)$, where p_i is the frequency of the ith allele detected in all individuals of the population (Nei 1973). In addition, the ability of a refined set of SSRs chosen here to discriminate a large dataset were compared to a neighbor-joining analysis of Spooner et al. (2007), who analyzed 742 accessions with 50 SSRs by neighbor joining in DARwin software 4.0 (http://darwin.cirad.fr/darwin/Home.php), to which we added one SSR (STM0019). For this analysis, similarity matrixes were calculated using Jaccard's coefficient and the comparison of the similarity matrixes were performed using the Mantel matrix-correspondence test in the MXCOMP option of the NTSYS 2.02h software (Sokal and Rohlf 1995). Correlations were conducted of the three main branches of this tree (the "bitter potato" (*S. ajanhuiri*, *S. curtilobum*, and *S. juzepczukii*) cluster, the diploid cluster, and the polyploid cluster), using correlation statistics in Microsoft Excel: $\rho_{X,Y} = \text{cov}\,(X,Y)/\sigma_X\sigma_Y$, where $\rho_{X,Y}$ is the correlation coefficient (r), cov (X,Y) is the co-variance of X and Y, σ_X is the standard deviation of X, and σ_y is the standard deviation of Y.

Construction of new potato SSR fragment size ladders

We initially used a pUC18 sequencing reaction or a IRDye-50–350 size standard in our LI-COR DNA Analyzer System as a fragment size ladder. To make our new kit easily applicable to the cultivated potato germplasm base across all platforms, we constructed new size ladders for each of the 24 primer pairs. We examined allele sizes from our database and selected genotypes displaying a range of sizes based on the following three criteria: (1) good separation among the alleles (>3 bp), (2) choice of allele/genotype combinations highlighting the high-frequency alleles encountered in our screening studies, and (3) the presence of the minimum and maximum size of the range of alleles, when possible. Genomic DNA was obtained from the DNA bank of CIP. Amplification products were obtained using standard protocol for SSR markers from CIP (Ghislain et al. 2004). Optimization of PCR conditions were conducted through temperature gradient PCR experiment for optimal annealing temperature, and appropriate number of amplification cycles to obtain good gel resolution of the bands. Electrophoreses to separate amplified products were performed using denaturant 6% poly-acrylamide gels and a silver stain protocol to reveal the bands (Ghislain et al. 2004).

Results

Genetic mapping

The 30 new candidate SSR markers from the TIGR database were surveyed for polymorphism in the PD,

BCT and PPC1 mapping populations (details are included in the supporting dataset 2). Two markers were monomorphic in all populations and three markers displayed a skewed segregation from the expected 1:1 ratio. The remaining 25 SSR markers could be mapped in one or two of the three populations using a LOD score of 3. The use of three segregating populations allowed us to identify 33 new map locations of 29 SSR markers not previously mapped. An integrated map was built with the three maps using a map integration function based on mean recombination frequencies and combined LOD scores of the selected sets of loci from each chromosome (Fig. 1). Four markers (STI0012, STM0019, STM0037, STM1053) of the PGI kit were monomorphic in all three segregating population tested and hence were included graphically based on published maps. Only six out of 157 map locations (STG0023, STG0027a, STM0038, STM2022, STM3009, STM51145) produced conflicting map locations on the integrated map and hence were not included. This map represents the most complete SSR potato map developed to date with 138 mapped potato SSR markers at 147 map locations

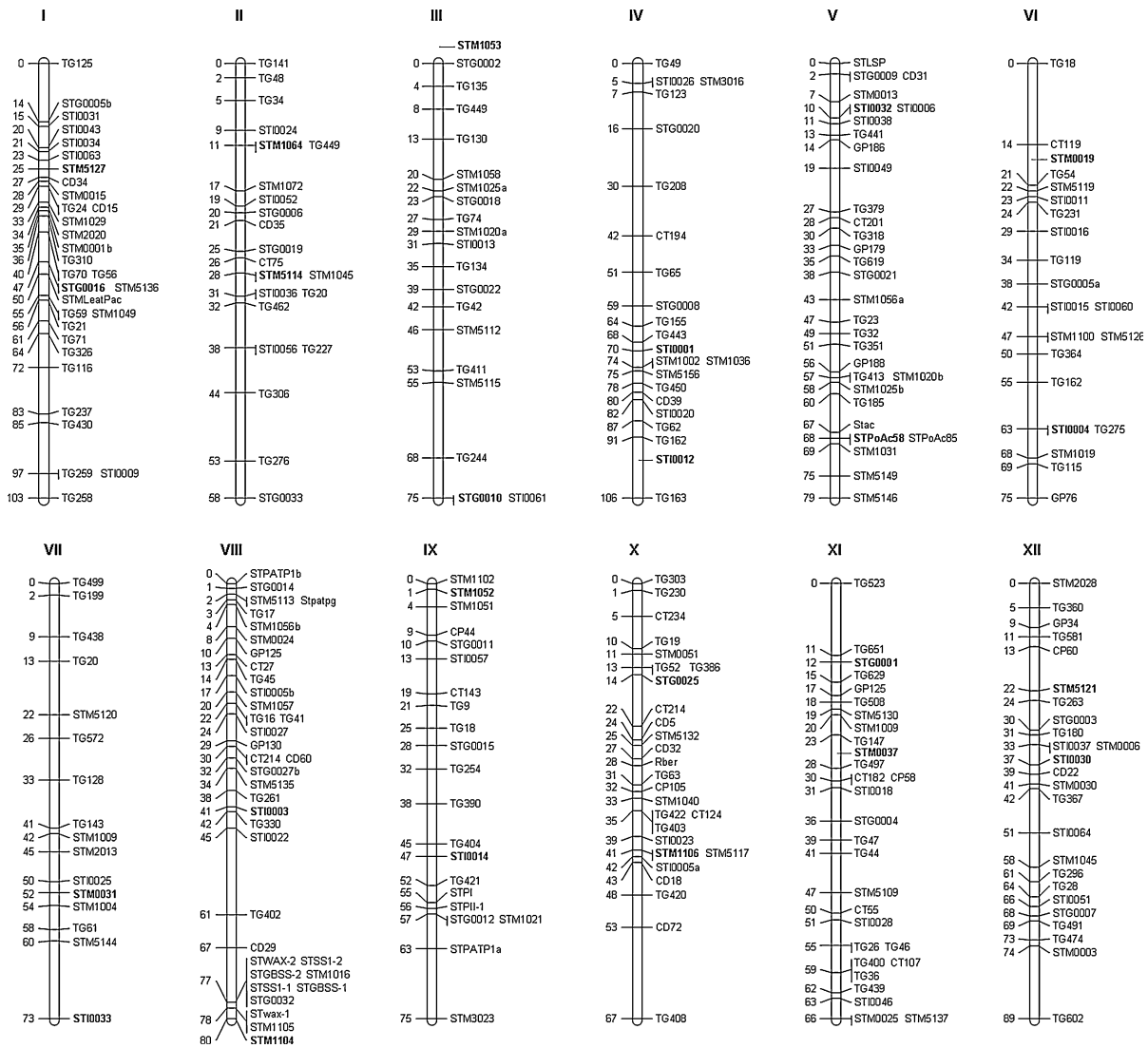

Fig. 1 Potato SSR genetic map including the 24 SSR markers of the new PGI kit (*bold*) on an integrated potato genetic map developed using framework RFLP and SSR genetic maps

Table 1 Descriptions of the 51 SSR markers and the selected 24 of the new PGI kit by their respective name, source, GenBank accession number, repeat motifs, forward and reverse primer sequences, annealing temperature, map location, allele size and number, and polymorphic information content (PIC) in 742 landraces (Spooner et al. 2007)

Name	Source	Repeat motif	GenBank accession # [SCRI unpublished #]	Primer sequences	*T*°a	Map location	Size (bp)	# Alleles	PIC	PGI kit
STG0001	g	(CT)*n*	BE340539	CAgCCAACATTTgTACCCCT ACCCCCACTTgCCATATTTT	58 (52)	XI g	137–163	15	0.699	Yes
STG0004	g	(GT)*n*	BI434815	TgAAAgCCAATCTCACTggA TATAATTggCTTgCgAgTgC	57 (55)	XI g	212–217	4	0.547	No
STG0006	g	(AC)*n*	BQ512261	TgAAAACTggTTTCCgCATT TAAgCAAgCTCTCTCCAggg	55 (57)	II g	148–180	9	0.644	No
STG0010	g	(TG)*n*	BM407152	CgATCTCTgCTTTgCAggTA gTTCATCACTACCgCCgACT	60 (55)	III g	175–192	11	0.685	Yes
STG0014	g	(AAT)*n*	BQ121319	gAATgTTTATCAgggCAATgg ggAAgATAACAggCACCCAA	54 (53)	VIII g	137–157	7	0.332	No
STG0016	g	(AGA)*n*	BI178934	AgCTgCTCAgCATCAAgAgA ACCACCTCAggCACTTCATC	55 (53)	I g	137–174	15	0.773	Yes
STG0017	g	(AGT)*n*	BQ113661	TTAATTggACAgCgTggACA AgCggAACAATgTgTAAggg	55 (55)	–	144–156	5	0.533	no
STG0021	g	(AAGA)n	BQ506520	TgCCTACTgCCCAAAACATT ACTggCTgggAAgCATACAC	57 (55)	V g	128–166	12	0.79	No
STG0025	g	(AAAC)*n*	BQ506618	TggAATCCgAATTACgCTCT AggTTTTACCACTCgggCTT	56 (55)	X g	208–223	7	0.531	Yes
STG0026	g	(CTCC)*n*	BQ119932	ACTgCCgCAAAAAgTgAAAA gCCgCTAggTggAgTAgATg	55 (55)	–	288–308	7	0.48	No
STG0032	g	(GCA)*n*	BQ121606	TTCATTTTCCgTTCAAAgCC TCgAgATCCATCATCAACgA	57 (53)	VIII g	127–136	6	0.373	No
STG0033	g	(CGG)*n*	BI935647	gCTCATTTgACTgCTAAACCC gAAAgAATTgTgCCgTCgAT	56 (55)	II g	148–162	6	0.661	No
STI0001	f	(AAT)*n*	CK860917	CAgCAAAATCAgAACCCgAT ggATCATCAAATTCACCgCT	60 (55)	IV fg	194–215	8	0.688	Yes
STI0003	f	(ACC)*n*	AW096896	ACCATCCACCATgTCAATgC CTCATggATggTgTCATTgg	60 (55)	VIII fg	137–188	17	0.746	Yes
STI0004	f	(AAG)*n*	BQ118939	GCTgCTAAACACTCAAgCAgAA CAACTACAAgATTCCATCCACAg	60 (55)	VI fg	83–126	13	0.688	Yes
STI0012	f	(ATT)*n*	U69633	gAAgCgACTTCCAAAATCAgA AAAgggAggAATAgAAACCAAAA	56 (55)	IV f	183–234	15	0.791	Yes
STI0014	f	(TGG)*n* (AGG)*n*	BQ115461	AgAAACTgAgTTgTgTTTgggA TCAACAgTCTCAgAAAACCCTCT	54 (55)	IX fg	127–157	10	0.708	Yes

Table 1 continued

Name	Source	Repeat motif	GenBank accession # [SCRI unpublished #]	Primer sequences	*T*°a	Map location	Size (bp)	# Alleles	PIC	PGI kit
STI0019	f	(ATCT)*n*	EG013494	TCCCTgTTgCCTTgAACAAT TgggAAAAggTACAAAgACgA	61 (60)	VII f	135–149	11	0.709	No
STI0022	f	(ACCCG)*n*	BQ511964	TCTCCAATTACTTgATggACCC CAATgCCATACACgTggCTA	61 (60)	VIII fg	126–156	10	0.713	No
STI0023	f	(CAG)*n*	BI920780	gCgAATgACAggACAAgAgg TgCCACTgCTACCATAACCA	61 (60)	X fg	172–245	21	0.813	No
STI0025	f	(CTCC)*n*	BQ119932	CTgCCgCAAAAAgTgAAAAC TgAATgTAggCCAAATTTTgAA	56 (60)	VII f	115–135	6	0.454	No
STI0030	f	(ATT)*n*	BF188393	TTgACCCTCCAACTATAgATTCTTC TgACAACTTTAAAgCATATgTCAgC	58 (60)	XII fg	94–137	17	0.811	Yes
STI0032	f	(GGA)*n*	BQ120452	TgggAAgAATCCTgAAATgg TgCTCTACCAATTAACggCA	61 (60)	V fg	127–148	8	0.745	Yes
STI0033	f	(AGG)*n*	BG886969	TgAgggTTTTCAgAAAgggA CATCCTTgCAACAACCTCCT	61 (60)	VII fg	131–155	9	0.512	Yes
STI0035	f	(TTC)*n*	BE919921	ACCTTTgAggAATTgCAggA CATTgAAggAgTTCCAgTCC	60 (57)	–	108–124	6	0.568	No
STI0036	f	(AC)*n* (TC)imp	BI920053	ggACTggCTgACCATgAACT TTACAggAAATgCAAACTTCg	55 (55)	II g	129–164	17	0.806	No
STI0038	f	(CTG)*n*	BQ509486	CCAAATgAggCTAAgggTgA ggCCAAgAAAATCAAgAACg	60 (58)	V g	112–127	7	0.546	No
STM0019a,b	d	(AT)*n* (GT)*n* (AT)*n* (GT)*n* (GC)*n* (GT)*n*	[PAC33]	AATAggTgTACTgACTCTCAATg TTgAAgTAAAAgTCCTAgTATgTg	– (47)	VI dg	(99–206)	17, 13	0.826	Yes
STM0030	d	compound(GT/GC) (GT)*n*	[PAC05]	AgAgATCgATgTAAAACACgT gTggCATTTTgATggATT	58 (53)	XII deg	122–168	19	0.868	No
STM0031	d	(AC)*n*…(AC)*n* GCAC (AC)*n* (GCAC)*n*	[PAC50]	CATACgCACgCACgTACAC TTCAACCTATCATTTTgTgAgTCg	53 (57)	VII dg	185–211	10	0.721	Yes
STM0037	d	(TC)*n* (AC)*n* AA (AC)*n* (AT)*n*	[PAC62]	AATTTAACTTAgAAgATTAgTCTC ATTTggTTgggTATgATA	52 (53)	XI dg	87–133	17	0.778	Yes
STM1016	d	(TCT)*n*	EU548082	TTCTgATTTCATgCATgTTTCC ATgCTTgCCATgTgATgTgT	56 (56)	VIII eg	243–275	17	0.84	No
STM1017	d	(ATT)*n*	L13771	gACACgTTCACCATAAAA AgAAgAATAgCAAAgCAA	– (53)	IX d	(132–136)	2	0.25	No
STM1031	d	(AT)*n*	X55748	TgTgTTTgTTTTTCTgTAT AATTCTATCCTCATCTCTA	– (55)	V dg	(236–301)	19	0.499	No
STM1049	d	(ATA)*n*	X13497	CTACCAgTTTgTTgATTgTggTg AgggACTTTAATTTgTTggACg	54 (57)	I dg	197–219	9	0.543	No

Table 1 continued

Name	Source	Repeat motif	GenBank accession # [SCRI unpublished #]	Primer sequences	*T*°a	Map location	Size (bp)	# Alleles	PIC	PGI kit
STM1052	d	(AT)*n* GT (AT)*n* (GT)*n*	AJ133765	CAATTTCgTTTTTTCATgTgACAC ATggCgTAATTTgATTTAATACgTAA	50 (52)	IX eg	214–263	17	0.832	Yes
STM1053	d	(TA)*n* (ATC)*n*	AB022690	TCTCCCCATCTTAATgTTTC CAACACAgCATACAgATCATC	53 (53)	III dg	170–196	10	0.688	Yes
STM1058	d	(ATT)*n*	Z13992	ACAATTTAATTCAAgAAgCTAgg CCAAATTTgTATACTTCAATATgA	55 (55)	III dg	130–139	4	0.285	No
STM1064	d	(TA)*n* (TG)*n* GT (TG)*n*	AC215425	gTTCTTTTggTggTTTTCCT TTATTTCTCTgTTgTTgCTg	55 (55)	II deg	201–213	8	0.566	Yes
STM1104	d	(TCT)*n*	EU548082	TgATTCTCTTgCCTACTgTAATCg CAAAgTggTgTgAAgCTgTgA	53 (57)	VIII deg	178–199	14	0.884	Yes
STM1106	d	(ATT)*n*	X95821	TCCAgCTgATTggTTAggTTg ATgCgAATCTACTCgTCATgg	51 (55)	X dg	145–211	17	0.821	Yes
STM2022	d	(CAA)*n*…(CAA)*n*	[C112]	gCgTCAgCgATTTCAgTACTA TTCAgTCAACTCCTgTTgCg	58 (53)	II deg	173–243	7	0.621	No
STM2030	d	(CA)*n* (TA)*n*	[C216]	TCTTCCCAAATCTAgAATACATgC AAAgTTAgCATggACAgCATTTC	– (55)	I d	(180–209)	4	0.274	No
STM3012	d	(CT)*n* (CT)*n*	[61D9]	CAACTCAAACCAgAAggCAAA gAgAAATgggCACAAAAAACA	56 (57)	IX d	180–225	9	0.65	No
STM5114	h	(ACC)*n*	[p102B19]	AATggCTCTCTCTgTATgCT gCTgTCCCAACTATCTTTgA	60 (57)	II eg	297–322	11	0.693	Yes
STM5121	h	(TGT)*n*	[p46L17]	CACCggAATAAgCggATCT TCTTCCCTTCCATTTgTCA	48 (50)	XII g	297–309	8	0.733	Yes
STM5127	h	(TCT)*n*	[p23e7]	TTCAAgAATAggCAAAACCA CTTTTTCTgACTgAgTTgCCTC	55 (60)	I eg	248–291	17	0.853	Yes
STM5140	h	(AAT)*n*	[p10B1]	gCTATTgTTgCAgATAATACg gCCATgCACTAATCTTTgATTC	60 (57)	IV e	180–219	12	0.674	No
STPoAc58	e	(TA)*n*	X55749	TTgATgAAAggAATgCAgCTTgTg ACgTTAAAgAAgTgAgAgTACgAC	– (57)	V eg	(243–263)	13	0.754	Yes
STGBSS	c	(TCT)*n*	EU548082	AATCggTgATAAATgTgAATgC ATgCTTgCCATgTgATgTgT	53 (53)	VIII eg	121–150	16	0.844	No
STWAX-2	a	(ACTC)*n*	EU548082	CCCATAATACTgTCgATgAgCA gAATgTAgggAAACATgCATgA	53 (53)	VIII eg	228–254	15	0.784	No

Data are provided for bands produced with M13-tailed PCR primers using the 4300 LI-COR DNA Analyzer System except those with parenthesis which follow Ghislain et al. (2004). Only one marker, STM0019 amplifies two loci. The source and map location of the SSR markers are indicated with a, for Veilleux et al. (1995); b, for Kawchuk et al. (1996); c, for Provan et al. (1996); d, for Milbourne et al. (1998); e, for Ghislain et al. (2004); f, for Feingold et al. (2005); g, for this article; and h, for unpublished from SCRI. Last column indicates whether the SSR marker pertains to the new PGI kit

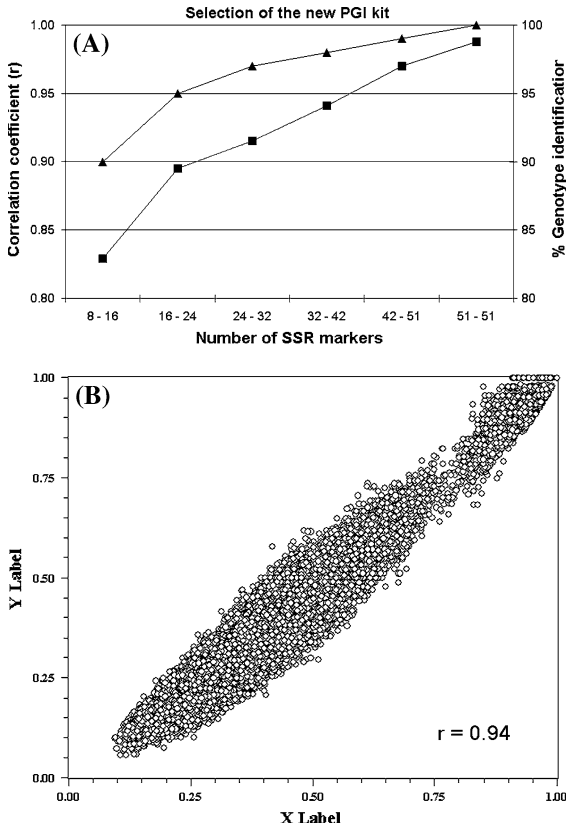

Fig. 2 Validation of the 24 SSR markers selected by discrimination analysis using genotyping data of 742 landraces: **a** Comparison of similarity matrixes generated with 8, 16, 24, 32, 42 and 51 SSR markers by *r*-values (▲), and by percentage of discrimination (■). **b** Representation of the comparison of similarity matrixes generated by 24 SSR [Y label] markers of the new PGI kit and the 51 SSR markers [X label] and the corresponding correlation coefficient *r*

(excluding the four placed graphically and the six conflicting).

Polymorphic information content and matrix comparisons

We analyzed 742 potato landraces of all four cultivated potato species: *S. tuberosum* Group Andigenum and Group Chilotanum, *S. ajanhuiri*, *S. curtilobum*, and *S. juzepczukii* with 56 SSR markers; 22 from the prior PGI kit (Ghislain et al. 2004) and the 34 most useful of the remaining 66 SSR markers based on marker quality as observed visually on gels. Out of these, five SSR markers appeared to be multi loci based on exceeding allele number considering the ploidy of the plant sample and these were not considered further.

Data obtained with the remaining 51 SSR markers on the 742 potato landraces were analyzed for polymorphic information content (PIC). Considering the cultivated potato a single gene pool (Spooner et al. 2007), markers were scored across all cultivar groups with different ploidies. PIC values per SSR marker ranged from 0.250 to 0.884 while the number of alleles per locus ranged from 2 to 21 (Table 1).

The discriminatory capacity of the SSR markers was analyzed by comparing a similarity matrix generated on 742 genotypes analyzed with 8, 16, 24, 32, 42 and 51 SSR markers ranked by their PIC value (Fig. 2). The results indicated that the 24 SSR markers with the highest PIC values provided nearly identical similarity matrices with one generated with the 32 highest PIC values ($r = 0.97$). In total, 93.5% of the genotypes can be discriminated using the selected 24 SSR markers (Table 1). However, comparison of placement of accessions within the three main branches of the 742 accession neighbor-joining tree of Spooner et al. (2007) were correlated at $r = 0.99$, suggesting that the 6.5% of the accessions not absolutely discriminated using the 24 markers are all so similar as to have little effect on major groups discovered in phenetic or phylogenetic analyses. These results lead us to propose 24 (Table 1) as an appropriate number of SSR markers for a new PGI kit.

Selection of the new PGI kit

The most informative 24 SSR markers for genotyping potato landraces were selected based on quality criteria, genome coverage, and locus-specific information content. We selected two SSR markers per chromosome with a linkage distance at least 10 cM except for chromosome VII, where markers STM0031 and STI0033 were separated by only 3 cM due to lack of alternative markers with a high PIC value (Fig. 1).

New potato SSR fragment size ladders

To construct a fragment size ladder for each SSR marker, we chose the alleles that were: (1) high-frequency, (2) covered the range of allele sizes, (3) well-separated alleles while avoiding those giving overlapping bands due to stuttering, and (4) displayed

Table 2 SSR fragment size standard for each SSR marker of the PGI kit

Locus	T°a	Accession	DNA (ng/µl)	Alleles	Size standards alleles
STG0001	58	705710	3	154 150 146 142	136 142 146 150 154 163
		706722		136	
		705029		163 146	
STG0010	58	703494	6	192 187	177 182 187 192
		705056	3	192 182 177	
STG0016	55	705191	9	160 154 148 143	137 143 148 154 160 172
		704414		154 137	
		702870		172 143	
STG0025	56	704434	3	219 215 208	208 215 219 223
		705903		223 215	
STl0001	60	704148	3	212 206 194	194 200 206 212
		704236		200	
STI0003	60	703721	9	158 137	137 149 158 170 179 188
		703284		170 158 149	
		704149		188 179 170 158	
STI0004	60	704019	3	101 95 92	83 92 95 101 107 112 121
		703882		121 107 95	
		702646		95 92 83	
		704309		112 95	
STI0012	56	704469	12	234 209 204 191	183 186 189 191 201 204 207 209 234
		705234		201 189 186 183	
		703381		207 204 201 183	
		706032		189 234	
STI0014	54	704669	3	154 148 145 139	133 139 145 148 154
		705029		148 145 133	
STI0030	58	705772	3	137 118 109	104 109 118 125 132 137
		705054		104 125	
		705958		132 109	
STI0032	61	705191	3	130 136	127 130 136 142 148
		705739		127 136 142 148	
STI0033	61	705111	3	155 149 137 131	131 137 143 149 155
		706134		143	
STM0019a	–	703783	12	(168 213)	(161 168 176 184 194 201 206 213)
	(47)	704717		(161 194)	
		705116		(201 206)	
		704074		(176 184 213)	
STM0019b	–	701399	9	(93 99 119)	(93 97 103 107 119)
	(47)	702937		(93 97 103)	
		705211		(93 107)	
STM0031	56	704592	6	203 195 185	185 195 203 211
		705875		211	

Table 2 continued

Locus	T°a	Accession	DNA (ng/µl)	Alleles	Size standards alleles
STM0037	48	705211	9	87 91	87 91 101 107 133
		703296		91 101 107	
		704073		91 133	
STM1052	50	703558	12	226 235 262	214 226 235 243 250 256 262
		705874		214 226 235 243	
		705601		226 256	
		704472		235 243 250 256	
STM1053	52	705887	3	192 189 180	170 180 186 189 192 196
		703721		189 186 170	
		705811		196 186	
STM1064	52	706702	6	209 206 201	201 206 209 213
		700299		213 209 206	
STM1104	55	704835	7.5	186 190 195	178 186 190 195 199
		703493		178 190 199	
STM1106	49	704368	9	157 163	145 157 163 169 175 181 211
		703502		169 175 181 211	
		705964		145	
STM5114	60	704903	3	314 308 302	302 308 314 322
		704183		322 314	
STM5121	50	706643		297 309	297 303 309
		703759		303	
STM5127	52	706776	3	285 276 269 263 248	248 258 263 266 269 276 285 291
		705655		291 269 266 258	
STPoAc58	–	704166	6	263 249	243 249 253 263
	(57)	702824		253 249 243	

Data are provided for bands produced with M13-tailed PCR primers using the 4300 LI-COR DNA Analyzer System except those with parenthesis which follow Ghislain et al. (2004). Potato accessions are presented by their CIP genebank numbers recommended template DNA concentration and the alleles produced

by a minimum number of landraces. These provide effective size ladders for easy extrapolation of alleles not part of the size kit. Annealing temperatures had to be verified and for few cases adjusted. Final DNA concentrations of the selected genotype range between 3 and 12 ng/µl to be mixed to produce the ladder (Table 2). We succeed in identifying a maximum of four (e.g., STM0019) genotypes to have good coverage of allele sizes. The number of alleles for each of the 24 SSR fragment size standard ranged from three (STM5121) to nine (STI0012) with an average number of 5.5 (Table 2). Overall, the 24 SSR fragment size standards produce 137 reference alleles, representing 44.7% of the total of 306 alleles found in 742 accessions by the 24 SSR markers of the new PGI kit. Allele sizes included in the size standards ranged from 83 to 322 bp providing an easy and convenient tool for identification of allele sizes (Fig. 3).

Discussion

The new PGI kit is composed of 24 SSR markers from over 200 we screened. It provides high-quality, high polymorphism alleles with two markers from each of the 12 linkage groups of potato separated by at least 10 cM, except for chromosome VII with

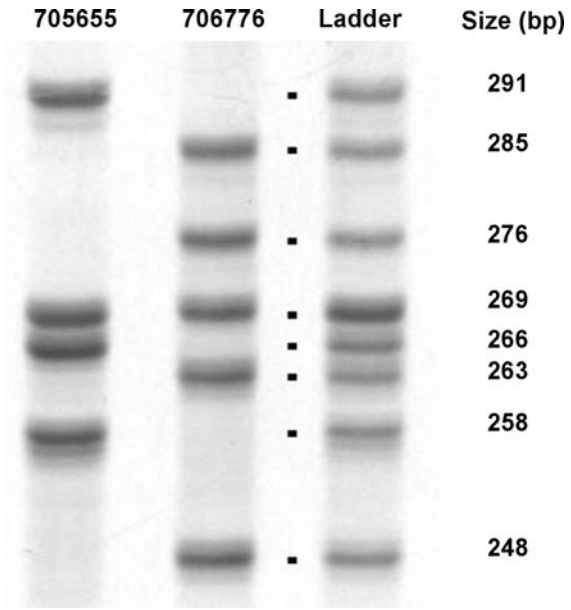

Fig. 3 SSR fragment size standard for the SSR marker STM5127 displaying eight well-defined and spaced alleles using a mix of only 2-genotypes

two SSRs separated by only 3 cM. It discriminates representative germplasm samples from all potato cultivar groups with high accuracy. Nine SSRs are from the previous PGI kit (Ghislain et al. 2004), three from ESTs developed at SCRI, four from TIGR, and eight from the University of Idaho. A composite reference DNA sample can conveniently be used to provide accurate sizing of all alleles for these SSR markers across laboratories and platforms.

The PGI kit can be used for potato germplasm characterization for a variety of purposes from identity verification (fingerprinting), to studies of genetic diversity, anchoring genetic linkage maps, establishment of core collections, and gene flow. The 24 composite reference samples of DNA for allelic size determinations will stimulate and foster collaborations worldwide on the use of SSRs for these applications. For example, we used the new PGI kit to identify potential duplicate landraces between the CIP and PROINPA Bolivian potato germplasm collections (data not shown). Some landraces belong to the same morphologically selected cluster were not grouped into the same molecular cluster especially for landraces of very diverse germplasm sets such as the *S. tuberosum* Andigenum Group. There was a total correspondence with the less diverse *S. tuberosum* Chilotanum Group. In another application, the new PGI kit has been using to genotype breeding lines and advanced cultivars of potato, and most of the breeding material groups into a well-defined cluster with landraces of the Chilotanum group (data not shown). Such grouping is expected because the germplasm of the Chilotanum group has been used extensively in potato breeding worldwide.

In summary, these highly characterized new SSR markers have tremendous utility for a variety of applications and can stimulate standardization and international collaborations within the cultivated potato gene pool. The PGI kit including primers and fragment size standard are available upon request. A SSR database of the cultivated potato is available on line from the bioinformatics portal of the Generation Challenge Program web site (www.generationcp.org) and of CIP (http://research.cip.cgiar.org/confluence/display/IPD/SSR+Marker). The latter provides a full description of each SSR markers, amplification and detection conditions, and the genotyping data of all potato landraces available to date. It is expected that with increased uses the SSR database will be integrated with the germplasm database of the CGIAR centers.

Acknowledgments We thank Sergio Feingold and Jim Lorenzen for providing their data set of the map location of SSR markers. This research was supported by the Grant SP1C2-2004-5 from the Generation Challenge Program, the International Potato Center, and by U.S. National Science Foundation Division of Environmental Biology (NSF DEB) 0316614 entitled: A world-wide treatment of *Solanum* (http://www.nhm.ac.uk/research-curation/projects/solanaceaesource//). The authors are grateful to the conscious dedication of Luciano Fernandez for his technical support.

References

Ashkenazi V, Chani E, Lavi U, Levy D, Hillel J, Veilleux RE (2001) Development of microsatellites markers in potato and their use in phylogenetic and fingerprinting analysis. Genome 44:50–62. doi:10.1139/gen-44-1-50

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Del Rio A, Bamberg J, Huamán Z (2006) Genetic equivalence of putative duplicate germplasm collections held at CIP and US potato genebanks. Am J Potato Res 83:279–285

Devaux C, Lavigne C, Falentin-Guyomarc'h H, Vautrin S, Lecomte JE, Klein K (2005) High diversity of oilseed rape pollen clouds over an agro-ecosystem indicates long-distance dispersal. Mol Ecol 14:2269–2280. doi:10.1111/j.1365-294X.2005.02554.x

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. BRL Focus 12:13–15

Feingold S, Lloyd J, Norero N, Bonierbale M, Lorenzen J (2005) Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum* L.). Theor Appl Genet 111:456–466. doi:10.1007/s00122-005-2028-2

Fenart S, Austerlitz F, Cuguen J, Arnaud J-F (2007) Long distance pollen-mediated gene flow at a landscape level: the weed beet as a case study. Mol Ecol 16:3801–3813. doi:10.1111/j.1365-294X.2007.03448.x

Ghislain M, Trognitz B, del Rosario Herrera M, Solis J, Casallo G, Vásquez C et al (2001) Genetic loci associated with field resistance to late blight in offspring of *Solanum phureja* and *S. tuberosum* grown under short-day conditions. Theor Appl Genet 103:433–442. doi:10.1007/s00122-001-0545-1

Ghislain M, Spooner DM, Rodríguez F, Villamón F, Núñez J, Vásquez C et al (2004) Selection of highly informative and user-friendly microsatellites (SSRs) for genotyping of cultivated potato. Theor Appl Genet 108:881–890. doi:10.1007/s00122-003-1494-7

Ghislain M, Andrade D, Rodríguez F, Hijmans RJ, Spooner DM (2006) Genetic analysis of the cultivated potato *Solanum tuberosum* L. Phureja Group using RAPDs and nuclear SSRs. Theor Appl Genet 113:1515–1527. doi:10.1007/s00122-006-0399-7

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. doi:10.1007/BF02849164

Milbourne D, Meyer R, Collins A, Ramsay L, Gebhardt C, Waugh R (1998) Isolation, characterization and mapping of simple sequence repeat loci in potato. Mol Gen Genet 259:233–246. doi:10.1007/s004380050809

Moisan-Thiery M, Marhadour S, Kerlan MC, Dessenne N, Perramant M, Gokelaere T et al (2005) Potato cultivar identification using simple sequence repeats markers (SSR). Potato Res 48:191–200. doi:10.1007/BF02742376

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323. doi:10.1073/pnas.70.12.3321

Provan J, Powell W, Waugh R (1996) Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). Theor Appl Genet 92:1078–1084. doi:10.1007/BF00224052

Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484. doi:10.1073/pnas.201394398

Rios D, Ghislain M, Rodriguez F, Spooner DM (2007) What is the origin of the European potato? Evidence from Canary Island landraces. Crop Sci 47:1271–1280

Schneider K, Douches DS (1997) Assessment of PCR-based simple sequence repeats to fingerprint North American potato cultivars. Am Potato J 74:149–160. doi:10.1007/BF02851594

Sokal RR, Rohlf FJ (1995) Biometry, 3rd edn. Freeman, New York, 0 7167 2411 1

Spooner DM, Núñez J, Rodríguez F, Naik PS, Ghislain M (2005a) Nuclear and chloroplast DNA reassessment of the origin of Indian potato varieties and its implications for the origin of the early European potato. Theor Appl Genet 110:1020–1026. doi:10.1007/s00122-004-1917-0

Spooner DM, van Treuren RR, de Vicente MC (2005b) Molecular markers for germplasm and genebank management. Tech Bull 10. International Plant Genetic Resources Institute, Rome

Spooner DM, Núñez J, Trujillo G, del Rosario Herrera M, Guzmán F, Ghislain M (2007) Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. Proc Natl Acad Sci USA 104:19398–19403. doi:10.1073/pnas.0709796104

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744

Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC (2005) Linkage disequilibrium in European elite maize germplasm investigated with SSRs. Theor Appl Genet 111:723–730. doi:10.1007/s00122-005-2057-x

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

Villamón FG, Spooner DM, Orillo E, Mihovilovich E, Pérez W, Bonierbale MW (2005) Late blight resistance linkages in a novel cross of the wild potato species *Solanum paucissectum* (series *Piurana*). Theor Appl Genet 111:1201–1214. doi:10.1007/s00122-005-0053-9