Investigation and Development of Potato Parentage Analysis Methods Using Multiplexed SSR Fingerprinting

**M. Spanoghe · T. Marique · J. Rivière ·
D. Lanterbecq · M. Gadenne**

Abstract Parentage assessment in the potato (*Solanum tuberosum* L.), based on a reliable characterization of DNA varieties, can be a useful tool in validating pedigree information as well as for assigning potential parents to orphan varieties. The approach suggested in this study was to use the genetic profiles obtained on 577 cultivars using a set of 17 polymorphic simple sequence repeats (SSR) markers in conjunction with the pedigree information provided by either the literature or the breeders to set up reliable parentage methods. Kinship testing based on the *simple exclusion* method was developed with cultivated potatoes for which the genetic data of both parents was available in our database and proved to be hopeful in validating pedigree information. To determine parental assignment, two inferential methods, the *maximum likelihood* (LOD) and the *genetic distance ranking* (GDR), were set up and compared in order to highlight which one offered the most likely candidate parent to a chosen variety. Inferential analysis results revealed that the "alleged parent" had a higher probability of being ranked first using the LOD method. Finally, we discussed the relevance of this new original approach, particularly in relation to instances where the pedigree is either in doubt or lacking information for the cultivars.

Keywords Kinship testing method · LOD score · Pedigree analysis · Simple exclusion method · Simple sequence repeats · *Solanum tuberosum*

M. Spanoghe (✉) · T. Marique · J. Rivière · D. Lanterbecq · M. Gadenne
Laboratory of Biotechnology—Applied Genetics Unit, Haute Ecole Provinciale de Hainaut-CONDORCET, 17 Chemin du Champ de Mars, 7000 Mons, Wallonia, Belgium
e-mail: martin.spanoghe@condorcet.be

T. Marique
Laboratory of Food Technology, Centre pour l'Agronomie et l'Agro-Industrie de la Province de Hainaut (CARAH), 11 rue Paul Pastur, 7800, Ath, Wallonia, Belgium

Introduction

Pedigree information can provide relevant data for breeders to make pertinent decisions as it allows a maintenance of high levels of variability in the progeny, an increase of the heterosis effect, and a reduction of the levels of inbreeding (Tarn et al. 1992). Luo et al. (2000) emphasized that information about parental genotypes is fundamental in distinguishing recombinant and parental genotypic classes for linkage analysis. For instance, pedigree information is used for multiple quantitative traits loci (QTL) mapping (Bink et al. 2002), association mapping (Simko et al. 2004; Malosetti et al. 2007; D'hoop et al. 2010; Baldwin et al. 2011) as well as for studies on extreme resistance inheritance (Song et al. 2005). Pedigree information is also highly relevant for genetic relationship estimations (Demeke et al. 1996; Sun et al. 2003; Braun and Wenzel 2004), some of which show that full siblings and half siblings were closely clustered, indicating that estimated relationships were connected to known pedigrees (Hosaka et al. 1994; Demeke et al. 1996; Isenegger et al. 2001).

In general, pedigree information is collected either by researchers, breeders, or breeding companies themselves. Today, a large amount of pedigree data on potato cultivars is available on a Web interface (Berloo et al. 2007). Although the data is continually updated, the database does not always provide the pedigree for some cultivars (i.e., orphans, old varieties, or unavailable information). The authors also stated that different sources of pedigree data may sometimes cause conflicting parental compositions or refer to certain cultivars under different names. Involuntary errors in pollination, seed harvest, or labeling may also generate errors. Berloo et al. (2007) thereby elucidated the need for reliable genetic tools in confirming existing pedigree data.

Since the 1980s, parentage analysis of plants has been widely expanded not only from the development of methods for assessing the parentage of individual progeny, but also from the increase in the diversity and availability of molecular markers. While allozymes (Ellstrand 1984; Smouse and Meagher 1994; Krauss 1994) as well as dominant molecular markers (RFLP, RAPD) (Lewis and Snow 1992; Milligan and McMurry 1993) and co-dominant molecular markers (Ashley and Dow 1994; Dow and Ashley 1996; Double et al. 1997; Buteler et al. 1997) have widely been applied to parentage studies in plants, it seems that simple sequence repeats (SSR) or microsatellites constitute the preferred markers to perform such parentage assessment research (Ashley 2010). This is particularly due to their high discriminatory power, co-dominant transmission, and reproducible properties (Tautz and Renz 1984; Tautz 1989; Li et al. 2002). Numerous original papers reported that these hyper-variable DNA sequences are an excellent source of genetic markers for potato cultivar fingerprinting (Veilleux et al. 1995; Kawchuk et al. 1996; Provan et al. 1996; Ghislain et al. 2004; Moisan-Thiery et al. 2005; Reid et al. 2009; Reid et al. 2011) as they provide the highest DI (diversity index) over dominant markers such as RAPD, AFLP, and RFLP markers (Milbourne et al. 1997; McGregor et al. 2000).

The earliest and conceptually simplest technique of parentage analysis is the *simple exclusion* method introduced by Ellstrand (1984). Based on Mendelian rules of inheritance, this technique uses genotypic incompatibilities between tested parents and their offspring in order to reject particular parent-offspring hypotheses.

More precisely, the analysis is based on a strict examination of genotypes and the exclusion is declared when an obligatory parental-inherited allele observed in the offspring is not found in the alleged parent. This method has been extensively used in many various plant studies such as gene flow estimation analyses (Hamrick and Schnabel 1985; Ellstrand et al. 1989), including extension to both relative male and female fertility estimations and mating system examinations (Broyles and Wyatt 1991; Snow and Lewis 1993), genealogy reconstruction (Lewis and Snow 1992; Milligan and McMurry 1993), and paternity testing (Grattapaglia et al. 2004). Today, this technique is mainly used in kinship testing in humans, in particular for paternity testing (Chakraborty and Stivers 1996). However, it is recognized that the simple exclusion method presents a certain number of limitations. Firstly, this method requires the acquisition of the genotype of one "confirmed" parent (for example the mother) in order to test the other so-called alleged parent (Devlin et al. 1988). Secondly, this analysis is usually hindered by genotyping errors, null alleles, or mutations appearing between parents and children which contribute to false exclusions for a given locus as observed in humans (Chakraborty et al. 1996; Gunn et al. 1997; Junge et al. 2006) and plants (Grattapaglia et al. 2004). Finally, this method is rapidly rendered impractical when the pool of candidate parents becomes too large due to the high number of loci needed to yield a single non-excluded parent or parent-pair assignment to an offspring (Jones and Ardren 2003). Under the strict exclusion rule, a single mismatch is enough to exclude a candidate parent. However, many exclusion programs allow the user to specify the number of mismatches that are necessary to validate the exclusion, leading to a less stringent method considering the difficulties imposed by mutations or scoring errors (Jones and Ardren 2003). For instance, the standard practice in human paternity testing requires the exclusion of a minimum of three loci for up to 12 SSR markers before a test can be declared negative (Junge et al. 2006). To our knowledge, no study has achieved parentage testing using the simple exclusion method, such as kinship testing, with the potato. This method appears to be promising for the development of a pedigree validation program and can easily be implemented despite the known limitations of the method. Indeed, due to polysomic inheritance of the potato and following the possibilities given by the segregation of the alleles from a parent to the offspring (Luo et al. 2000; Hackett and Luo 2003), the application of such a method is theoretically possible. With regard to the phenomenon of double reduction, which is one of the most important features of auto-tetrasomic inheritance (Mather 1936), it was first assumed that it occurs sporadically in the cultivated tetraploid potato (Haynes and Douches 1993). However, recent studies have shown variable results for the evaluation of double reduction proportion which may range to 10% (Slater et al. 2014). Therefore, the phenomenon of double reduction in the cultivated tetraploid potato may have a potential impact on parentage analysis.

Another approach introduced by Meagher (1986), the *maximum-likelihood* method, sometimes called *most likely parent* (or likelihood ratio) method, is preferably used when genotypes are available for a set of offsprings and a set of potential parents within a defined population from which the parent is most likely to be identified. The maximum-likelihood method is an inferential method which is generally chosen after applying the simple exclusion method, specifically in instances where complete

exclusion is not possible (Jones and Ardren 2003). In this parentage assessment, the ratio precisely divides the likelihood of an individual being the parent of a given offspring by the likelihood of these individuals being unrelated (Meagher and Thompson 1986). To compare the likelihoods of different relationships, a log-likelihoods score (LOD) is calculated for each possible parent as the logarithm to the base e of the likelihood ratio (Meagher and Thompson 1986), providing a null, positive, or negative likelihood parental score. After an exhaustive evaluation of all genetically compatible parents, the individual with the highest LOD is chosen as the true parent, while all other parents are discarded. Contrary to the simple exclusion method, the maximum-likelihood method allows paternity assignment for a higher number of offspring (Devlin et al. 1988; Smouse and Meagher 1994) and usually allows for some degree of transmission errors due to genotype misreading or mutation (Jones and Ardren 2003). However, some limitations are also found with the maximum-likelihood method. Firstly, the used loci must be unlinked and unambiguously inherited (Smouse and Meagher 1994; Buteler et al. 1997). Secondly, the method is only accurate when the entire set of potential parents has been sampled (Marshall et al. 1998). Thirdly, parentage remains ambiguous when multiple parent–offspring relationships obtain the highest nonzero likelihood (Jones and Ardren 2003), typically leading to an arbitrary choice of the parent with the highest LOD score as the true parent, with little consideration to the relative likelihoods of other possible parents. Nevertheless, it is now possible to mitigate this effect by setting a threshold which allows the rejection of compatible genotypes, by either using random genotypes taken from the population sharing the same allele frequency (Meagher 1986) or by using simulation procedures (Marshall et al. 1998; Gerber et al. 2000). Other studies have emphasized that a statistical bias in favor of homozygotes exists, therefore favoring a homozygous putative parent that will always result in a higher LOD score for a given locus than would that of a heterozygous individual (Devlin et al. 1988; Smouse and Meagher 1994; Buteler et al. 1997). Finally, the model provides no explanation for when LOD scores result in a tie (Devlin et al. 1988; Schnabel 1998). However, most of these limitations can be overcome when the number of potential parental candidates is low or the number of genetic markers is increased (Buteler et al. 1997; Schnabel 1998). With regard to the potato, this inferential method seems to be an interesting approach to investigate parental relationships, particularly when parental pedigree information is unavailable (e.g., orphans and old varieties) or uncertain. As with the simple exclusion method, the possibilities given by the segregation of the alleles from a parent to the offspring in autotetraploid potatoes should not disrupt the inferential methodology.

This study investigated these two approaches with the aim of identifying the need for the development of reliable programs. A pedigree validation program could be constructed by implementing kinship testing based on the simple exclusion method. Whereas for varieties whose pedigree information is either unknown or uncertain, a potential parent assignment program could be applied through the use of inferential analysis such as the LOD method. The reliability of these selected methods is examined in this paper through the use of pedigree information provided by either the literature or the breeders and the efficiency of the LOD method was evaluated by comparing it to another inferential method, the GDR method, that used the Jaccard's coefficient (Jaccard 1908) as an arithmetic measure of genetic distance between two individuals to generate a ranking of the most related varieties.

Material and Methods

For a detailed view of the experimental procedure used in this study, see the schematic view in Fig. 1.

Plant Materials and Genomic DNA Isolation

Analyses were carried out on 577 European potato cultivars, including triploid or tetraploid varieties that were provided by several breeders and growers. Plant material and genotype data were collected with the best intentions by donors and the authors, respectively. The final results are published as such. The authors are not responsible for any potential inaccuracy except for those arising from genotyping analyses. In particular, legal issues or disputes raised by owners of germplasm cannot be based on data or conclusions disclosed in this publication.

Genomic DNA was isolated from germ, tuber, or leaf material using the QuickPick Plant DNA kit and Pickpen 1-M from Bio-Nobile (Turku, Finland). The concentration of the DNA solutions was quantified using a ND-3300 NanoDrop spectrofluorimeter (Thermo Scientific; Waltham, MA, USA). A subsample of each concentrated DNA
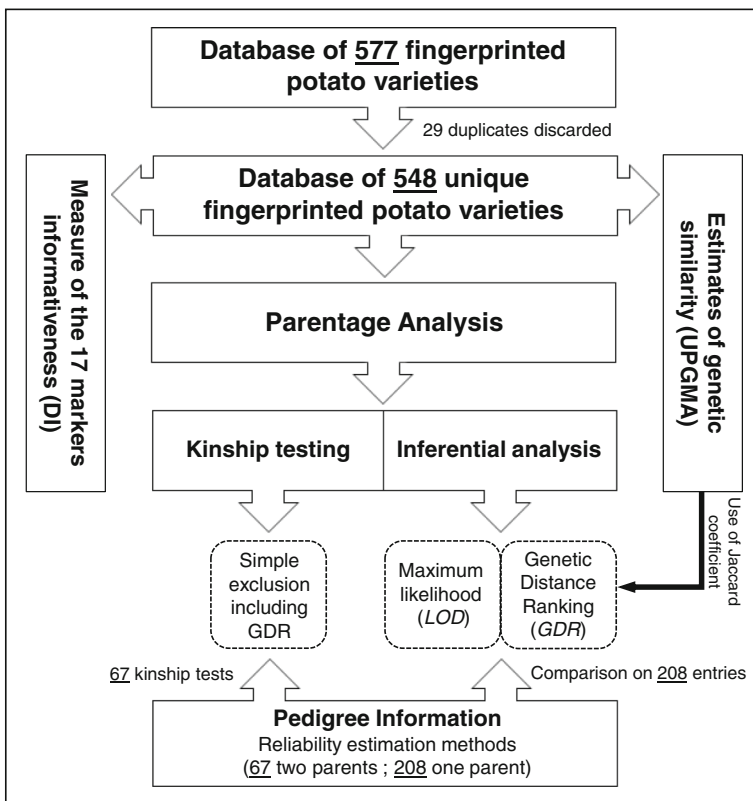

Fig. 1 Schematic view of experimental procedure

solution was diluted to 2 ng/μl, which was found to be the suitable concentration for optimizing the multiplexed-PCRs.

Microsatellite Markers

The 17 SSR primers used in the present study (Table 1) were taken from the literature (Veilleux et al. 1995; Provan et al. 1996; Milbourne et al. 1998; Ghislain et al. 2004; Feingold et al. 2005; Ghislain et al. 2009). They were chosen based on the polymorphic information content (PIC) or diversity index (DI), the length of the obtained products and also on the similarity of their optimal annealing temperatures. Twenty-six primer pairs were a priori preferentially selected based on annealing temperatures ranging from 53 to 57 °C. This range included the highest number of primers available. However, only 17 primer pairs were kept after an intensive preliminary testing phase (see "Supplementary material" for more details), which resulted in the combination of these 17 primer pairs in three sets of multiplexed-PCR.

Multiplexed-PCR Amplification

Amplification of the 17 loci was performed in two sets of six loci and one set of five loci, using a Kapa2G HotStart Multilocus Amplification Kit (Kapa Biosystems, Boston USA). The 5′ end of the forward primer of each pair was labeled with a fluorescent dye (Fam, Hex or Ned dyes; Table 1). The PCR amplification mixture in a total reaction volume of 25 μl contained 1.5× Kapa2G HotStart PCR Buffer, 0.2 mM of each dNTP, 0.1–0.2 μM of each primer (Applied Biosystems), 1 unit of KAPA2G Fast Hotstart DNA Polymerase and 20 ng of template DNA. All amplifications were carried out in the same conditions (annealing temperature of 53 °C) on a 9800 Fast Thermal Cycler (Applied Biosystems). The PCR profile was as follows: after an initial denaturing step of 2 min at 95 °C, the reaction consisted of 29 cycles of 15 s at 95 °C, 30 s at 53 °C (T_a), 12 s at 72 °C, and one cycle of 2 min at 72 °C for final extension. The completed reaction products were held at 4 °C until the electrophoresis analysis.

Analysis of Amplicons

The PCR products were diluted in water at a ratio of 1:15. A volume of 1 μl of the diluted solution was transferred into 14 μl of a denaturing mix (1% of the ROX-labeled molecular weight marker GS-400HD Rox and 99% of formamide), then denatured at 95 °C for 5 min. Each of the three sets of multiplexed-PCR was independently run on a 4-capillary ABI 3100-avant Genetic Analyzer (Applied Biosystems). The denatured samples (15 μl) were separated using a standard polyacrylamide gel (POP4; Applied Biosystems) and run following standard parameters (Injection_voltage:5.0 kV; Injection_time:8 s; Run_voltage:15.0 kV; Run_time:1250 s). Estimations of fragment sizes were enabled by including a number of tracks containing size 21 GS-400HD Rox-labeled DNA fragments (Life Technologies; Carlsbad, CA, USA), covering a size range of 50–400 bp. The lengths of the PCR products were determined using GeneMapper Software 4.0 (Applied Biosystems) and used to construct a multilocus allelic phenotype for each cultivar. A database was then compiled, regrouping the 577 varieties fingerprints.

Table 1 Multiplexed SSR sets and data analysis. Locus, core repeat, primer sequences (5'-3') with fluorochrome label, source of primer, concentration of markers for each primer mixture (conc. in mix), number of distinct peaks (NP), number of allelic phenotypes (NAP), frequency of most common allelic phenotypes (FMCAP), allele size range, and DI were estimated from 548 European potato cultivars for each multiplexed SSR set

Locus	Core repeat	Primer sequences (5'-3') forward-reverse	Source	Conc. in mix (umol/l)	NP	NAP	FMCAP	Allele size range (bp)	DI
Muliplexed SSR set 1									
STGBSS	(TCT)9	AATCGGTGATAAATGTGAATGC-Ned ATGCTTGCCATGTGATGTGT	b	0.1	11	43	0.308	119–140	0.847
STI0012	(ATT)n	GAAGCGACTTCCAAAATCAGA-6Fam AAAGGGAGGAATAGAAAACCAAAA	e	0.12	9	61	0.086	161–186	0.958
STI0036	(AC)n(TC)n	GGACTGGCTGACCATGAACT-Hex TTACAGGAAATGCAAACTTCG	e	0.1	13	108	0.091	113–144	0.970
STM1052	(AT)14GT(AT)4(GT)6	CAATTTCGTTTTTCATGTGACAC-Ned ATGGCGTAATTTGATTTAATACGTAA	c	0.2	10	101	0.099	206–252	0.961
STM1064	(TA)n (TG)n GT (TG)n	GTTCTTTGGTGGTTTTCCT-Hex TTATTTCTCTGTTGTTGCTG	c	0.12	8	28	0.347	186–193	0.789
STWAX-2	(CT)4(CT)8	CCCATAATACTGTCGATGAGCA-6Fam GAATGTAGGGAAACATGCATGA	a	0.14	17	106	0.170	205–243	0.934
Muliplexed SSR set 2									
STG0006	(AC)n	TGAAAACTGGTTTCCGCATT-6Fam TAAGCAAGCTCTCTCCAGGG	f	0.12	8	14	0.728	139–157	0.442
STG0016	(AGA)n	AGCTGCTCAGCATCAAGAGA-Ned ACCACCTCAGGCACTTCATC	f	0.12	11	43	0.141	118–156	0.928
STI0014	(TGG)n(AGG)n	AGAAACTGAGTTGTGTTTGGGA-Hex TCAACAGTCTCAGAAAACCCTCT	e	0.12	8	22	0.409	113–130	0.692
STM1049	(ATA)6	CTACCAGTTGTTGATTGTGGTG-Ned AGGGACTTTAATTTGTTGGACG	c	0.12	7	29	0.332	178–199	0.803

Table 1 (continued)

Locus	Core repeat	Primer sequences (5'-3') forward-reverse	Source	Conc. in mix (umol/l)	NP	NAP	FMCAP	Allele size range (bp)	DI
STM1104	(TCT)5	TGATTCTCTTGCCTACTGTAATCG-Hex CAAAGTGGTGTGAAGCTGTGA	c	0.12	12	49	0.314	162–179	0.847
STM5127	(TCT)n	TTCAAGAATAGGCAAAACCA-Hex CTTTTTCTGACTGAGTTGCCTC	f	0.12	9	36	0.423	234–249	0.791
Muliplexed SSR set 3									
STG0001	(CT)n	CAGCCAACATTTGTACCCCT-6Fam ACCCCCACTTGCCATATTTT	f	0.12	12	80	0.109	119–144	0.954
STG0021	(AAGA)n	TGCCTACTGCCCAAAACATT-Hex ACTGGCTGGGAAGCATACAC	f	0.12	7	27	0.250	107–131	0.872
STI0030	(AAT)n	TTGACCCTCCAACTATAGATTCTTC-6Fam TGACAACTTTAAAGCATATGTCAGC	e	0.12	11	66	0.235	82–103	0.906
STM3012	(CT)4(CT)8	CAACTCAAACCAGAAGGCAAA-Hex GAGAAATGGGCACAAAAAACA	c	0.12	8	27	0.203	158–205	0.867
STPOAC58	(TA)n	TTGATGAAAGGAATGCAGCTTGTG-Ned ACGTTAAAGAAGTGAGAGTACGAC	d	0.12	10	48	0.204	224–243	0.907
Mean					10.06	52.24	0.26	82–243	0.851

The sources of the SSR markers are indicated with *a* for Veilleux et al. (1995), *b* for Provan et al. (1996), *c* for Milbourne et al. (1998), *d* for Ghislain et al. (2004), *e* for Feingold et al. (2005), and *f* for Ghislain et al. (2009)

Estimates of Genetic Similarity by the UPGMA Clustering

Genetic similarity between all potato cultivars was estimated by the UPGMA (un-weighted pair group method with arithmetic mean) clustering method using Jaccard's coefficient (Jaccard 1908). Distinct peaks were scored as either 1 or 0, for presence or absence, for each cultivar accession and for all the 17 loci. All varieties sharing identical genetic fingerprints (duplicates) as revealed by genetic similarity estimation were removed before-hand from the analysis in order to obtain a reliable measure of marker informativeness and to perform optimal treatment of parentage analysis. Out of 577 varieties, 29 duplicates were discarded from the database, and analyses were thus conducted on 548 varieties. Binary values data were then exported to MATLAB V5 software (The Math Works Inc., Natick, MA, USA) as a matrix of 548 varieties×the total number of distinct alleles (a peak was considered as a distinct allele). A similarity matrix was computed by pair-wise comparisons according to the number of shared peaks, and a frequency distribution of pairwise Jaccard similarities for all 548 varieties was generated.

Measure of Markers' Informativeness

To obtain a measure of the utility of the marker systems, a DI was calculated from fingerprint data for each marker according to the Nei's statistic (Nei 1973; Nei and Li 1979): $DI = 1 - \sum p_g^2$ where p_g is the frequency of an individual allelic phenotype, assuming that each SSR reveals one locus (Provan et al. 1996; Feingold et al. 2005).

Parentage Analysis

1. Kinship testing method
    The kinship testing method was developed based on the simple exclusion method. Using all the pedigree information disclosed either by literature or breeders, we performed kinship testing on 67 hybridized offspring whose genetic information of both parents was available in our database.
    The simple exclusion method was based on a strict examination of allelic phenotypes between the tested parents and their offspring according to the Mendelian rules of inheritance. We first examined whether all alleles present in the offspring were also found in the alleged parents and then ensured that the distribution of inherited alleles was correctly balanced in the offspring for each given locus. During this step, if one of these two conditions was not satisfied, the test accumulated one segregation mismatch (SM) for each violation (see examples in Table 2). The SM is defined as a mismatch appearing between the allelic phenotype of tested parents and a hybrid, which is supposedly due to either allelic mutation, segregation distortion, null alleles, or genotyping errors. To minimize false exclusion effects in such a method, the test was adapted by adding a genetic distance parameter in the form of the Jaccard's coefficient as defined during the UPGMA process. This allowed us to formulate a logical test as follows (Table 2): if the genetic distance between the hybrid and each alleged parent ($GD_{[HYB-PAR]}$) is less than the average of the genetic distance between the hybrid (or the alleged parents) and the varietal genetic pool ($GD_{[HYB-POOL]}$ or $GD_{[PAR-POOL]}$) (made up

Table 2 Examples of genetic profile examination and conditions of the kinship testing method for three different scenarios, where the hybrid (HYB); the parent (PAR); the allelic phenotype (peaks were assigned letter codes instead of actual size) for each individual SSR marker (M); the genetic distance based on Jaccard's coefficient (GD); the genetic distance mean of the hybrid with all other individuals in the pool ($GD_{[HYB-POOL]}$); the genetic distance between the hybrid and the alleged parent ($GD_{[HYB-PAR]}$); the genetic distance mean of the alleged parent with all other individuals in the pool ($GD_{[PAR-POOL]}$); the position value of the tested parent within the most likely parent ranking ($Rank_{PAR}$); and the number of segregation mismatches (SM) are given

		M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	M 11	M 12	M 13	M 14	M 15	M 16	M 17
1	PAR 1	CI	H	EFJ	BCFG	DG	AG	CEF	EI	BCH	AEF	CDGJ	AF	DG	CDEG	BCG	D	HP
	PAR 2	EIJ	EH	CEF	BG	DG	AEG	CEF	AIJ	CK	AEF	BDFG	AF	DG	CDE	ABF	CD	HKP
	HYB	I	H	EFJ	BFG	DG	AG	CEF	AEI	C	AEF	BCGJ	AF	DG	CEG	AB	CD	HKP
	SM no.=0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	PAR 1	CJ	EH	CDF	CG	CDG	ACEG	CF	AE	CEM	A	BCDJ	A	CDG	CDEG	AC	CD	DHKP
	PAR 2	CI	EH	CDJ	CFG	D	CF	CF	ABE	CH	E	BD	A	D	CDE	AC	CDE	P
	HYB	CI	H	CDF	CG	D	CE	CF	ABJ	C	EF	BD	AD	D	CDE	AC	CDE	KP
	SM no.=3	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0
3	PAR 1	BCIJ	H	CE	CFG	DG	CEG	CEF	AJ	AHK	BEF	BC	AD	DG	CEG	ACD	CDF	HK
	PAR 2	CDGJ	H	EF	CG	DGI	CEF	EF	EJ	AEHM	BEF	BD	AD	DGJ	CD	CDG	CJ	GHK
	HYB	CIJ	H	CEJ	CEG	CDG	EG	CEF	ABJ	AHK	BEF	ACG	AD	CDG	DEG	AC	CF	DHP
	SM no.=7	0	0	1	1	1	0	0	1	0	0	1	0	1	0	0	0	1

Conditions

	Test result		
1	If $SM = 0$		
2	Elseif $GD_{[HYB-PAR]} > GD_{[PAR-POOL]}$ and $GD_{[HYB-PAR]} > GD_{[HYB-POOL]}$ and If rank PAR> 0.647		
3	Else		

Examples

Test result	HYB	Parent 1			Parent 2			SM no.
	$GD_{[HYB-POOL]}$	$GD_{[HYB-PAR]}$	$GD_{[PAR-POOL]}$	Rank PAR	$GD_{[HYB-PAR]}$	$GD_{[PAR-POOL]}$	Rank PAR	Rank PAR
No exclusion	0.453	0.729	0.432	1	0.654	0.471	1	0
No exclusion	0.471	0.725	0.435	1	0.538	0.484	0.855	3
Exclusion	0.424	0.630	0.450	0.995	0.397	0.401	0.352	7

The first scenario presents a not excluded test where there is no SM between the alleged parental couple (Hansa/Nicola) and their progeny (Juliette). The genetic distance between each tested parent and the progeny is higher than the mean of genetic distance between the progeny (or parents) and the varietal genetic pool. The $Rank_{PAR}$ shows that both tested parents are suggested as the most likely parents (1st position of 547 for Hansa and 1st position of 546 for Nicola). The second scenario presents a not excluded test with 3 SM between the parents (Agria/Morene) and their progeny (Adriana). The genetic distance between each parent and the progeny is higher than the mean of genetic distance between the progeny (or parents) and the varietal genetic pool. The $Rank_{PAR}$ shows that one parent (Agria) is not suggested as the most likely second parent although the ranking value remains relatively high (79th position of 546) and is above the rejecting threshold (>0.647). The third scenario presents an excluded test with 7 SM between the parental couple (Anosta/Premiere) and their progeny (Amora). The genetic distance between one parent (Premiere) and its progeny is lower than the mean of genetic distance between this parent and the varietal genetic pool. The $Rank_{PAR}$ shows that one parent (Premiere) is not suggested as the most likely second parent (192th position of 546) and is under the rejecting threshold (<0.647)

of 548 varieties), the test is considered as negative (exclusion is declared). In addition, exclusion is also declared through examining the position value of both tested parents within the most likely parent ranking (Rank$_{PAR}$ as performed during GDR assessment) for each kinship test and when the resulting position value is under a threshold of 0.647. The value of the exclusion threshold should be defined as the minimum probability of being a parent and is thus established based on the position of the lowest ranked alleged second parent among all 67 kinship tests that presented no SM (for this study: Rank$_{PAR}$: 0.647=193th position of 546). We assumed that both parents could be ranked in the first two positions (1st position out of 547 for the first parent, 1st position out of 546 for the second parent) meaning that the second parent may be ranked as the second most likely parent. The analysis also focused on the correspondence between the number of SM and the result (exclusion or not) of the adapted test for each assessment.

2. Inferential analysis

Two methods were developed to establish a parentage analysis based on a ranking of the most likely parent: the *maximum-likelihood* (LOD) method and the *genetic distance ranking* (GDR) method. For both methods, inferential analysis was performed for all of the 548 distinct varieties present in our database, each being considered as a single entry. Thus, for both methods, 548 assessments in total were conducted. For each assessment, 547 entries were inferred successively as a potential parent for the analyzed variety and then ranked as the most likely parent criterion in descending order.

The LOD method involved calculating a logarithm of the likelihood ratio (LOD) for each possible parent, which provided either a null, positive, or negative likelihood parental score (LOD score). Due to the polysomic inheritance of the potato and consequently to the constraints of the specific locus dosage during the scoring of polymorphic fragments, we used the likelihood ratio statistics, originally developed to analyze amplified fragment length polymorphism (AFLP) markers (Gerber et al. 2000) to perform this inferential analysis. The use of the LOD formulas based on dominant markers directly enabled a performance of such an analysis with an allelic phenotype system. The formulas of the likelihood ratio were then written in MATLAB V5 software (The Math Works Inc., Natick, MA, USA) (see appendix for details) to automate the process for each parent/offspring combination (($548^2 - 548)/2=149878$ combinations). For each of the 548 assessment, inference of parentage was performed on a ranking of the most probable varieties according to decreasing LOD scores.

The GDR method used the Jaccard's coefficient as an arithmetic measure of genetic distance between two individuals to generate a ranking of the most related varieties. The similarity matrix computed during the estimates of genetic similarity by the UPGMA clustering was therefore used to perform the GDR method and each of the 149878 combinations was also considered as a potential parent/offspring combination. Inference was thus performed on a ranking of the most related varieties based on a decrease in the Jaccard's coefficient.

In order to gain a clear vision on the reliability of both methods, we carried out a parental ranking comparison between both methods based on the number of varieties for which the pedigree information of one parent was known and the corresponding genetic data was available in our database. Of the 548 assessments, 208 answered to

these two criteria. The comparison study therefore focused on 208 assessments. We then evaluated the reliability to infer the "alleged parent" based on both the first ranked candidate as well as the top 30 criteria of parental ranking. Additional analyses were also conducted on the ranking distribution of first position of "alleged parent" to determine if both methods classed the "alleged parent" in the same way.

Results

Estimates of Genetic Similarity by the UPGMA Clustering

A pairwise similarity matrix based on Jaccard's coefficient was generated, which allowed the drawing of a frequency distribution of pairwise similarities for all 548 varieties (Fig. 2). The average Jaccard similarity between all 548 potato cultivars with these markers was 0.447. The 16 pairs that were the most genetically distanced shared a similarity of 0.15–0.20 whereas the 5 closest pairs shared a similarity of 0.80–0.85. As a whole, all pair values reflected a good variability in the cultivar pool with a normal distribution (extreme cases, for example those too close or too distant varieties, were not abundant).

Measure of Markers' Informativeness

A total of 17 pairs of primers were used to amplify potato microsatellites ranging from 82 to 252 bp. The average of unique allelic phenotypes for all of the 17 markers across all 548 fingerprinted potatoes averaged 52.24, ranging from 14 to 108, with the highest number for *STI0036* (Table 1). The Diversity Index (DI) varied in this study from 0.442
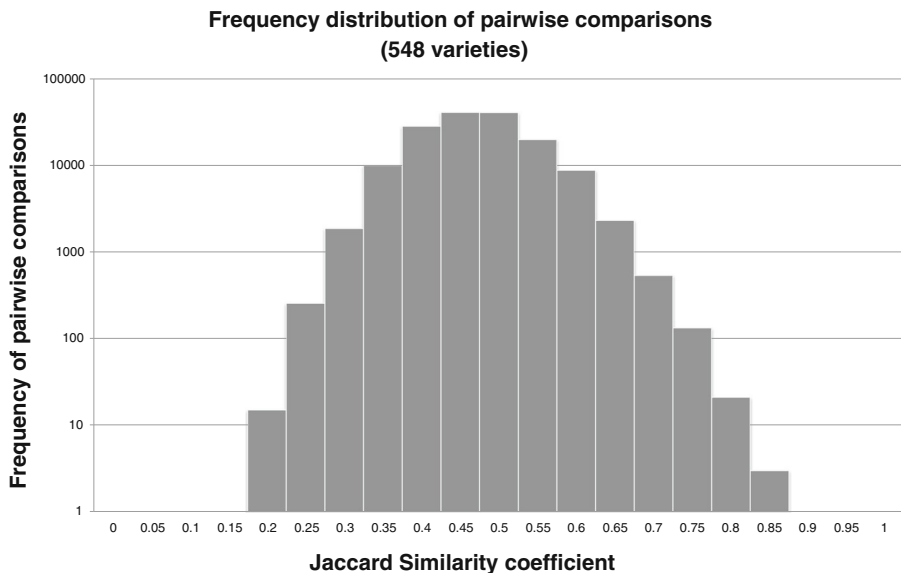

Fig. 2 Frequency distribution of pairwise comparisons for the 548 potato varieties of the database

(*STG0006*) to 0.970 (*STI0036*) across all 548 fingerprinted genotypes. Mean DI for the 17 SSR markers was 0.851. The highest values were shown for *STM1052* and *STI0036*. Nevertheless, these markers did not necessarily present the highest number of alleles (an allele corresponded to a distinct peak). Indeed, *STWAX-2* showed the highest number of alleles (17 alleles) in our study. The number of alleles across all fingerprinted genotypes averaged 10.06, ranging from 7 to 17. In total, the 17 microsatellite loci produced 171 alleles. The use of these three multiplexed-PCR sets is thus an extremely powerful tool to differentiate cultivars, with the best DI for the first set. Using the calculation based on the frequency of the most common allelic phenotypes of each marker as an upper limit, the chance of two varieties yielding identical profiles based on these 17 markers gave an infinitesimally small value. The upper limit is based on the presented frequencies of most common allelic phenotypes (see Table 1): $0.308 \times 0.086 \times \ldots \times 0.204 = 6.9 \times 10^{-12}$.

Parentage Analysis

1. Kinship testing method

   The kinship tests were conducted on 67 lines, which corresponded to the number of tests for which the genetic profile of the offspring and both of their parents (referring to the available pedigree information) were available in the genetic data base. The results showed that 61 out of 67 kinship tests were "not excluded". All of the 6 excluded tests were indicative that only one of the parents did not meet the validation criteria (see Table 3). Otherwise, no "excluded tests" involving the attributes of both parents were met in this study. The number of segregation mismatches (SM) was counted for each test. Of the 61 "not excluded" tests, 42 tests did not show any SM while the remaining 19 presented between 1 and 5 SM. The 6 "excluded" tests reported between 3 and 8 SM. This examination thus revealed that the number of SM, alone, can be no decisive for the outcome of the test, since the tests obtaining between 3 and 5 SM can either be categorized as "excluded" or "not excluded". The reliability of the material source as the cause of the exclusion can be directly rejected if a variety is excluded in a particular test while this same variety is not excluded in another test. This is for example the case of test no. 47 which excluded the variety *Charlotte* as the parent while test no. 29 confirmed that the same variety *Charlotte* is the presumed parent of the tested hybrid. Similarly, test no. 5 excluded the variety *Premiere* as the second parent while test no. 50 confirmed both parents of the hybrid *Premiere*.

2. Inferential analysis

   a. The maximum-likelihood method

      The frequency distribution of all parent/offspring combinations followed a Gaussian curve where the mean estimated LOD score was close to 0. The LOD scores ranged from a minimum of −15 to a maximum of 30. A LOD score of zero implied that the alleged parent was as likely to be a parent of an offspring with the same probability as would be a randomly selected individual. A positive LOD score implied that the alleged parent was more likely to be a parent of an offspring than a randomly selected individual would be. The maximum and minimum LOD scores were surprisingly high but had a very

Table 3 Kinship testing data where the kinship test number (no. test); the hybrid name (HYB); the parents name (PAR); the genetic distance mean of hybrid with all other individuals in the pool ($GD_{[HYB-POOL]}$); the genetic distance between hybrid and alleged parent ($GD_{[HYB-PAR]}$); the genetic distance mean of alleged parent with all other individuals in the pool ($GD_{[PAR-POOL]}$); the position value of tested parent within the most likely parent ranking ($Rank_{PAR}$); the number of segregation mismatches (SM no.); and the test result are given

no. test	HYB	GD [HYB-POOL]	PAR 1	GD [HYB-PAR]	GD [PAR-POOL]	Rank PAR	PAR 2	GD [HYB-PAR]	GD [PAR-POOL]	Rank PAR	SM no.	Test result
1	Adriana	0.471	Agria	0.538	0.483	0.855	Morene	0.725	0.435	1.000	3	Not excluded
2	Amarin	0.438	Carrera	0.604	0.471	0.991	Fresco	0.755	0.419	1.000	0	Not excluded
3	Amazone	0.418	Civa	0.612	0.435	0.998	Provita	0.660	0.417	0.998	0	Not excluded
4	Ambra	0.437	Berber	0.711	0.436	1.000	Caesar	0.571	0.451	0.978	0	Not excluded
5	Amora	0.424	Anosta	0.630	0.450	0.995	Premiere	0.397	0.401	0.352	7	Excluded
6	Anky	0.454	Amalia	0.516	0.436	0.817	Amora	0.655	0.424	0.998	2	Not excluded
7	Annabelle	0.460	Monalisa	0.547	0.475	0.932	Nicola	0.712	0.470	1.000	0	Not excluded
8	Anosta	0.450	Ostara	0.640	0.505	1.000	Provita	0.589	0.417	0.989	0	Not excluded
9	Arrow	0.453	Fresco	0.660	0.419	0.996	Solara	0.379	0.432	0.137	4	Excluded
10	Atlas	0.477	Spunta	0.690	0.507	1.000	Wilja	0.659	0.455	0.998	0	Not excluded
11	Bafana	0.480	Felsina	0.750	0.420	1.000	Victoria	0.600	0.503	0.980	0	Not excluded
12	Ballerina	0.463	Agria	0.640	0.483	0.993	Obelix	0.721	0.487	1.000	0	Not excluded
13	Bellini	0.470	Felsina	0.767	0.420	1.000	Mondial	0.667	0.489	0.991	0	Not excluded
14	Canberra	0.433	Latona	0.646	0.450	0.998	Red Scarlett	0.600	0.483	0.993	0	Not excluded
15	Cerisa	0.475	Franceline	0.739	0.472	1.000	Laura	0.600	0.438	0.969	0	Not excluded
16	Challenger	0.486	Aziza	0.627	0.420	0.985	Victoria	0.729	0.503	0.998	0	Not excluded
17	Cimega	0.463	Caesar	0.509	0.451	0.784	Mondial	0.509	0.489	0.784	5	Not excluded
18	Colomba	0.491	Agata	0.733	0.471	0.998	Carrera	0.682	0.471	0.993	0	Not excluded
19	Concurrent	0.441	Estima	0.565	0.412	0.971	Sinaeda	0.457	0.412	0.593	4	Excluded
20	Corrida	0.512	Franceline	0.694	0.472	1.000	Provento	0.615	0.470	0.962	0	Not excluded
21	Craigs Alliance	0.450	Craigs Defiance	0.617	0.480	0.995	The Alness	0.690	0.458	1.000	0	Not excluded
22	Derby	0.476	Fresco	0.627	0.419	0.993	Mondial	0.653	0.489	0.996	0	Not excluded
23	Eurostar	0.481	Innovator	0.547	0.462	0.832	Victoria	0.708	0.503	1.000	3	Not excluded
24	Excalibur	0.407	Saxon	0.642	0.371	1.000	Valor	0.433	0.457	0.679	5	Excluded
25	Fambo	0.450	Ostara	0.673	0.505	1.000	Provita	0.589	0.417	0.987	0	Not excluded
26	Felsina	0.420	Gloria	0.592	0.404	0.987	Morene	0.651	0.435	0.991	1	Not excluded
27	Forzat	0.475	Agria	0.691	0.483	0.998	Innovator	0.571	0.462	0.949	1	Not excluded
28	Furore	0.423	Alpha	0.688	0.432	0.998	Rode star	0.660	0.424	0.996	0	Not excluded
29	Gourmandine	0.451	Charlotte	0.682	0.439	0.998	Estima	0.625	0.411	0.996	0	Not excluded
30	Hanna	0.490	Morene	0.609	0.435	0.956	Spunta	0.717	0.507	0.998	4	Not excluded
31	Inova	0.445	Impala	0.585	0.465	0.974	Nicola	0.648	0.470	1.000	0	Not excluded
32	Jazzy	0.444	Cupido	0.462	0.411	*0.647*	Franceline	0.660	0.472	1.000	0	Not excluded
33	Juliette	0.453	Hansa	0.729	0.431	1.000	Nicola	0.654	0.470	1.000	0	Not excluded
34	Kardal	0.479	AM 66-0042	0.700	0.447	1.000	Astarte	0.660	0.463	0.998	0	Not excluded
35	Lacetta	0.466	Agria	0.700	0.483	0.998	Felsina	0.571	0.420	0.943	0	Not excluded
36	Lady Amarilla	0.443	Agria	0.508	0.483	0.837	Hermes	0.552	0.441	0.960	0	Not excluded
37	Laperla	0.474	Minerva	0.620	0.375	0.982	Valor	0.556	0.457	0.897	3	Not excluded
38	Lekkerlander	0.425	Dore	0.717	0.462	1.000	Dr_Mcintosh	0.675	0.468	1.000	0	Not excluded
39	Lionheart	0.470	Symfonia	0.667	0.479	0.998	Victoria	0.717	0.503	0.998	0	Not excluded
40	Luciole	0.491	Exquisa	0.623	0.495	0.974	Monalisa	0.588	0.475	0.923	0	Not excluded
41	Manitou	0.431	Laura	0.696	0.438	0.998	Maranca	0.682	0.427	0.998	0	Not excluded
42	Markies	0.452	Agria	0.647	0.483	0.998	Fianna	0.627	0.472	0.995	1	Not excluded
43	Marlen	0.431	Agria	0.589	0.483	0.995	Saturna	0.673	0.458	1.000	1	Not excluded
44	Messi	0.446	Bellini	0.717	0.470	1.000	Shepody	0.577	0.429	0.982	0	Not excluded
45	Metro	0.456	Felsina	0.652	0.420	0.995	Mondial	0.667	0.489	1.000	0	Not excluded
46	Parel	0.434	Eigenheimer	0.627	0.439	0.996	Libertas	0.667	0.483	1.000	0	Not excluded
47	94 F 2042	0.431	Charlotte	0.418	0.439	0.434	Emeraude	0.531	0.418	0.932	8	Excluded
48	Performer	0.464	Innovator	0.574	0.462	0.974	Nika	0.692	0.429	1.000	1	Not excluded
49	Picasso	0.435	Ausonia	0.646	0.422	1.000	Cara	0.771	0.478	1.000	0	Not excluded
50	Premiere	0.401	Civa	0.538	0.435	0.980	Provita	0.745	0.417	1.000	0	Not excluded
51	Primura	0.487	Majestic	0.673	0.517	0.993	Sirtema	0.635	0.464	0.982	3	Not excluded
52	Producent	0.459	AM 66-0042	0.673	0.447	1.000	Prevalent	0.621	0.474	0.996	1	Not excluded
53	Provento	0.470	Elvira	0.686	0.475	1.000	Escort	0.455	0.452	0.419	3	Excluded
54	Quincy	0.417	Astérix	0.612	0.423	1.000	Quincy	0.531	0.420	0.976	0	Not excluded
55	Romano	0.472	Desiree	0.714	0.529	0.996	Draga	0.660	0.488	0.991	0	Not excluded
56	Stayer	0.432	Kardal	0.519	0.479	0.929	Oleva	0.500	0.471	0.844	5	Not excluded
57	Stemster	0.478	Desiree	0.660	0.529	1.000	Maris Piper	0.534	0.424	0.791	5	Not excluded
58	The Alness	0.458	Abundance	0.689	0.483	0.998	Majestic	0.667	0.517	0.998	0	Not excluded
59	Timate	0.484	AM 66-0042	0.740	0.447	1.000	Elvira	0.679	0.475	0.996	0	Not excluded
60	Tresor	0.483	Corine	0.680	0.489	0.998	Fresco	0.667	0.419	0.996	0	Not excluded
61	Triplo	0.480	Agria	0.600	0.483	0.965	Fresco	0.593	0.419	0.952	2	Not excluded
62	Ukama	0.462	Marijke	0.698	0.469	0.996	Sirtema	0.589	0.464	0.969	2	Not excluded
63	Urgenta	0.429	Furore	0.673	0.423	1.000	Katahdin	0.583	0.499	0.985	0	Not excluded
64	Vitalia	0.455	Lady Christl	0.583	0.412	0.965	Rodeo	0.633	0.474	0.998	2	Not excluded
65	Volumia	0.440	Adora	0.667	0.467	1.000	Mondial	0.646	0.489	0.998	0	Not excluded
66	VR 808	0.413	Atlantic	0.545	0.448	0.982	Lady Claire	0.692	0.425	1.000	0	Not excluded
67	Yona	0.459	Caesar	0.509	0.451	0.775	Kondor	0.652	0.489	0.998	2	Not excluded

The genetic distance is calculated based on the Jaccard coefficient. The highlighted entries are the lines that did not meet the validation criteria. Value in italic is the exclusion threshold based on the $Rank_{PAR}$ ($Rank_{PAR}$ was fixed based on the last ranked candidate of the kinship test (no. 32) that presented no SM)

low frequency (0.77% for values≤−10 and≥+10) and were negligible in terms of occurrence. High LOD scores (≥ +10) were due to a combination of pairs of varieties having both a close genetic similarity and a vast genetic distance between said pairs and the other pairs in the varietal genetic pool. Regarding the ranking of the most likely parent, the 1st ranked entry across all 548 assessments averaged a LOD score of 13.99 while the last ranked entry averaged −12.87. For the 1st ranked entry, no obtained LOD scores resulted in a tie.

b. The genetic distance ranking method

Since this inferential method used the Jaccard coefficient as the genetic distance parameter, the frequency distribution of all parent/offspring combinations was the same as that generated during the UPGMA process, taking into account that each pair is considered here as a parent/offspring combination. This frequency distribution also followed a Gaussian curve where the mean estimated similarity between all the potato cultivars was 0.447. The similarity values ranged from a minimum of 0.169 to a maximum of 0.818. With regard to the ranking of the most likely parent, the 1st ranked entry across all 548 assessments averaged a Jaccard coefficient of 0.685 while the last ranked entry averaged 0.260.

c. Comparison in the ranking of the two inferential analysis methods used

The comparison for the ranking of the most likely parent between both methods was conducted on 208 entries, which corresponded to the number of varieties where one parent was known and present in our database. The positional ranking, whether in the first position or within the top 30, for the alleged parent differed according to the method (Fig. 3). Indeed, it appeared
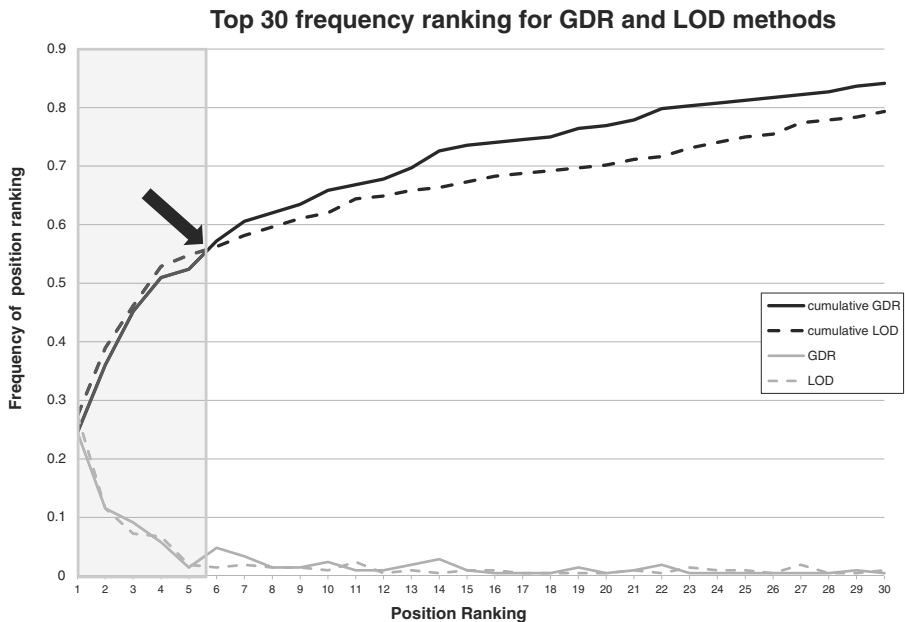

Fig. 3 Top 30 frequency ranking for GDR and LOD methods

that the "alleged parent" had a 27.40% probability (57 assessments of 208) of being ranked first using the LOD method as compared to 24.52% (51 assessments of 208) using the GDR method. Moreover, the LOD method gave a better result if we consider that the "alleged parent" has a 54.81% chance of being in the top 5 of the most likely parent ranking, as compared to 52.40% for the GDR method (see limit on Fig. 3). However, after this top 5 "virtual limit", the trend is clearly reversed in favor of the GDR method, which has the best probability to place the "alleged parent" in the top 30 (LOD: 79.33%; GDR: 84.13%). The ranking distribution of the 1st position also showed that neither method systematically classed the "alleged parent" in the same way. Indeed, among the assessments that inferred the expected parent in the 1st position (57 for LOD and 51 for GDR), both methods achieved at the same result (same inferred parent in the 1st position) only for 39 assessments.

Discussion

Since the early 1990s, the utilization of genetic markers, in particular SSR, is well documented for the potato and constitutes useful tools for cultivar identification (Kawchuk et al. 1996; Ghislain et al. 2004; Moisan-Thiery et al. 2005; Ghislain et al. 2009; Reid et al. 2009; Reid et al. 2011; Côté et al. 2013; Karaagac et al. 2014) as well as for linkage analysis (Bradshaw et al. 1998; Milbourne et al. 1998; Song et al. 2005) and genetic relationship studies (Provan et al. 1996; Schneider and Douches 1997; Ashkenazi et al. 2001; Braun and Wenzel 2004). Our study extends the scope of the polymorphic SSR markers in the potato as parentage analysis methods developed in this study have shown that SSR markers were also promising in supporting pedigree information. Pedigree confirmation finds many applications of interest in several areas. For instance, this tool can be used as an objective criterion of variety validity being that genetic heritage is an indisputable proof of variety authenticity. This may also be helpful in plant breeding, particularly in selecting parental lines. Indeed, parental choice based on a certified pedigree is essential for precise estimation of breeding values such as the maximization of the heterosis effect and the reduction in the levels of inbreeding (Tarn et al. 1992). Many studies also emphasized the importance of pedigree information in linkage analysis (Luo et al. 2000), such as quantitative trait loci (QTL) (Bradshaw et al. 1998; Bink et al. 2002), association mapping (Simko et al. 2004; Malosetti et al. 2007; D'hoop et al. 2010; Baldwin et al. 2011), and extreme resistance inheritance (Song et al. 2005). Finally, this tool may also be highly relevant for studies on genetic relationship estimations (Hosaka et al. 1994; Demeke et al. 1996; Sun et al. 2003), which have often faced problems of unreliable pedigree information (Braun and Wenzel 2004).

In our study, a large panel of microsatellite markers were tested, from which 17 loci were selected on the basis of their diversity index (DI) or polymorphic information content (PIC) and their compatibility to be used in a multiplexed-PCR. After an intensive preliminary testing phase, we set up three multiplexed SSR sets theoretically capable of discriminating all potato cultivars. However, several varieties resulted in

having the same fingerprint, and we hypothesized two different explanations for this finding. Firstly, some varieties were called by different names yet had the same allelic phenotype and therefore could be considered as either synonyms (e.g., the following pairs could be considered as synonyms: *Rosa/Platte de Florenville*; *Duke of York/Eersteling*; *Cleopatra/Margarita*; *Ratte/Nagleiner Kipfler*; *Bernadette/Dali*; *Elfe/Marabel*; *Jose/Wilja*) or somaclonal mutants (e.g., *Diamant* is mutant of *Cardinal*; *Famosa* is a mutant of *Estima*; *Margaritta* is a mutant of *Picasso*). Secondly, a human error such as sampling or handling error could also have explained such a finding. This analysis allowed us to support the information provided by the potato pedigree database (Berloo et al. 2007) and by the breeders in relation to certain similar varieties (completely identical according to our 17 SSR) bearing different names. With regard to the informativeness of these selected markers, the high DI values revealed a large variability in allelic phenotypes. In general, the mean of these values was higher than those observed in similar studies using the same markers (Provan et al. 1996; Milbourne et al. 1997; Ghislain et al. 2004; Rios et al. 2007). The high values in our study could be explained by either a larger number of fingerprinted cultivars or a wide genetic variability within the sampled pool. A similar study which was based on a microsatellite database construction of potato varieties (Reid et al. 2011) showed that markers having the highest DI value systematically presented the largest number of alleles. In our study, this was not the case: the marker *STI0036* had the highest DI mean value but presented only 13 alleles (the maximum number of alleles was found for the marker *STWAX-2* with 17 alleles). As a final point, both the compiled database which regrouped a large number of varieties and the high number of informative SSR markers used in this study appeared, in our opinion, to be pivotal in performing a reliable parentage analysis.

To our knowledge, no study has attempted parentage analysis using such molecular tools in the potato. To examine a possible first step in this direction, our initial strategy was to develop a kinship testing method, based on the simple exclusion method (Ellstrand 1984), in order to highlight errors and abnormalities present in the available pedigree information. The application of such a method was theoretically possible with polysomic inheritance of the potato and following the possibilities given by the segregations of the alleles from a parent to the offspring (Luo et al. 2000; Hackett and Luo 2003). To overcome the constraint of false exclusions at given loci, current practice requires more than three exclusion loci before a test is reported as negative (Grattapaglia et al. 2004; Junge et al. 2006). From our point of view, this approach is too restrictive because it is based on an empirical value, which does not take into account the exact number of loci used in the study nor the ploidy level. Indeed, since the majority of tested potatoes are tetraploid and our DNA fingerprinting methodology was based on a large number of markers, it was assumed that the probability of finding more allelic mutations or segregation distortion between the parent and offspring for a given locus increases. The adaptation of the simple exclusion method, with the integration of a Jaccard similarity coefficient, allowed us to circumvent the problem of false exclusion in a polyploidy parentage study. The new criteria were henceforth based on the assumption that both the genetic similarity between the hybrid and the tested parent(s) were more than likely higher than the average genetic similarities between each of them (parents and hybrid) and the tested varietal pool. In addition, the inclusion of a ranking position parameter into the logical test strengthened its result

since it appeared that relatives were closely ranked. Taking these criteria into account, the kinship testing method confirmed 61 tests (out of the 67) based on data collected in the literature. Although these validation tests confirmed the majority of the disclosed pedigree information, the exclusion cases found in our study must be considered with caution being that the results came from the development of a still-investigated method and the number of exclusions depended strictly on the availability of the three related varieties in the database (genetic data from offsprings and both parents were all required in order to perform the test). In any case, we hypothesized three exclusion categories to explain exclusion cases. The first category includes "wrong pedigree information" offspring, which can be explained by the tests where the analyzed variety is clearly not the progeny of the presumed parents (or vice versa). The second category consists in "human error", which is a situation encountered when the tested individual (parent or offspring) is not that which is expected due to genotyping errors, involuntary errors in pollination, seed harvest, or labeling as well as an accidental mixing from the breeder's collection. The third category concerns problems of "name redundancy" which arise from varieties originating from different locations yet bearing the same name (i.e., tested individual had the same name as an unrelated variety and was thus attributed to other parental couples). This is, for instance, the case for *Laura*, a variety from Austria, whose pedigree is different from the *Laura* variety from Germany. In other words, exclusion was met when the genetic fingerprint of one or more tested individuals is different from that which is expected. This could arise from either a mistake in the publication of the new variety's pedigree, an uncertain variety origin, or a human error leading to erroneous genetic data.

The second strategy undertaken in this study was to implement an inferential parentage analysis by using a large pool of fingerprinted varieties. Two different methods were applied: one based on a common genetic distance (GDR method) and the other on probabilistic tools (LOD method). Regarding the LOD method, we used the formulas based on dominant loci (Gerber et al. 2000) in order to overcome the constraints of the specific locus dosage during the scoring of polymorphic fragment (Provan et al. 1996). In our case, this approach seemed easier to implement than the one, for example, suggested by Esselink et al. (2004) since we no longer had to define any extra haplotype or extra information in the genotypic model when unambiguous evidence of null alleles was available. The formulas used for the dominant markers have been successfully applied following polyploid model (Rodzen et al. 2004) and demonstrated here suitable performance in parentage inference for the potato. For each of the two methods applied, inferential analysis was conducted on all of the 548 varieties present in our database. Each inferred a rank of 547 candidates' parents by following a decreasing order of the generated values (i.e., Jaccard's coefficient or LOD score following the used method). The reliability of both inferential methods was determined by referencing the produced ranking data of a likely parent with available pedigree information. Of the 548 varieties, the constraints of information availability allowed for only 208 evaluations. Our results revealed that, while both methods did not consistently rank the "alleged parent" in the same way, the LOD method was more likely than the GDR method to classify a true parent in the first position among the many possible candidates. In fact, this method seemed to be more appropriate since the likelihood is calculated based on segregation probabilities (Meagher and Thompson 1986) while the GDR method is based on arithmetic calculation (Jaccard 1908).

Although the LOD method gave better results in the ranking criterion, the chance of finding the alleged parent in the first position remained relatively low (27.40%), but interestingly, the probability of retrieving this alleged parent within the top five ranking was doubled (54.81%). This first investigation of inferential parentage analysis in the potato revealed that varieties with the highest scores were thus not necessarily the true parents, and consequently, all of the other 340 inferred most likely candidate parents, which were not supported by prior pedigree information, could not be assigned as a true parent with certainty. Direct examination of the ranking results for both methods indicated a significant presence of siblings in the ranking that may have been either genetically close to the analyzed variety or to the inferred parent. They were thus considered as interfering candidates. Indeed, the presence of other family members in the pool of candidate parents poses a serious challenge to parentage assessment since it may disrupt the parental ranking (Jones and Ardren 2003). Although strategies exist that account for some of the effects of family structure on the assignment of kinship, for instance by including a confidence threshold (Marshall et al. 1998; Gerber et al. 2000), additional work is needed before it is determined with certainty which techniques are most sensitive in studies involving relatives, especially for crop plants. In our potato parentage study, tested varieties were not sampled from a natural population because of the bias caused by varietal selection. Family members, other than parents such as full- or half-siblings, present in the sampled pool was thus important. Another aspect of the ranking examination revealed a less efficient ranking for many modern breeds that have a parent widely used in breeding such as *Agria*. As a consequence, these breeds tended to cluster together, also disrupting the ranking. For these reasons, the use of inferential analysis to assign a true parent is a priori to be treated with caution in such a study; but the inferential method remains a promising tool for assigning parents, especially if a genotypic examination (or known sibling examination) of the best ranked parental candidates is performed in complement. Certainly, this complementary step could increase the probability to rank the true parent in the first position by rejecting eventual interfering candidates.

In conclusion, the parentage analysis investigated in this study, based on kinship testing and inferential methods, proved to be particularly hopeful in regards to the development of pedigree information validation, or parental assignment programs. Although inferential methods and the exclusion method are based on different approaches, they both appear to be promising to support the existing potato pedigree data, either in breeder's catalogs or in online potato pedigree databases by detecting, for instance, abnormalities (e.g., typing errors, synonymous names). Moreover, the kinship testing method may be used as a hopeful complementary tool for assigning parents with inferential methods, especially when the presence of interfering candidates disrupt the parental ranking. These two approaches may also be helpful for studies of genetic diversity between individuals or varieties using genetic fingerprinting data as mentioned in specific reviews (Devlin et al. 1988; Mohammadi and Prasanna 2003; Ge et al. 2011). Indeed, our results showed that siblings were closely clustered, confirming that closely related cultivars, estimated by genetic fingerprinting, were connected to pedigrees. Such results therefore support the hypothesis that reliable information on pedigree can steer uncertain interpretation on both clustering and estimates of true measures of genetic similarity in the potato as pointed out in different studies (Hosaka et al. 1994; Braun and Wenzel 2004). Finally, the supporting pedigree information can

also be a helpful tool in plant breeding, as it allows the breeders to make informed decisions regarding the selection of germplasm for crossing schemes, in particular for the creation of new varieties.

Appendix

In this study, the *maximum likelihood* method was carried out according to the formulas used for the dominant markers described in Gerber et al. (2000), which have been successfully applied following polyploid model (Rodzen et al. 2004).

The LOD formulas were written in MATLAB V5 software (The Math Works Inc., Natick, MA, USA), as follows:

```
frequ=sum(data,1)./548;
sumlod=[];
   for i=1:548
      for j=(i+1):548
      lod=[];
         for step=1:171
            if data(i,step) == 1 && data(j,step) == 1
            lod(step)=reallog((frequ(step)+(1+frequ(step))*(1-frequ(step)))/(frequ(step)*(2-frequ(step))^2));
            elseif data(i,step) ~= data(j,step)
            lod(step)=reallog(1/(2-frequ(step)));
            else
            lod(step)=reallog(1/(1-frequ(step)));
            end
         end
      end
      sumlod(i,j)=sum(lod);
   end
end
```

References

Ashkenazi V, Chani E, Lavi U, Levy D, Hillel J, Veilleux RE (2001) Development of microsatellite markers in potato and their use in phylogenetic and fingerprinting analyses. Genome 44:50–62

Ashley MV (2010) Plant parentage, pollination, and dispersal: how DNA microsatellites have altered the landscape. Crit Rev Plant Sci 29:148–161. doi:10.1080/07352689.2010.481167

Ashley MV, Dow BD (1994) The use of microsatellite analysis in population biology: background, methods and potential applications. In: Schierwater B, Streit B, Wagner GP, DeSalle R (eds) Molecular ecology and evolution: approaches and applications. Experientia Supplementum, Birkhäuser Basel, pp 185–201. doi:10.1007/978-3-0348-7527-1_10

Baldwin SJ, Dodds KG, Auvray B, Genet RA, Macknight RC, Jacobs JME (2011) Association mapping of cold-induced sweetening in potato using historical phenotypic data. Ann Appl Biol 158:248–256. doi:10.1111/j.1744-7348.2011.00459.x

Berloo R, Hutten RCB, Eck HJ, Visser RGF (2007) An online potato pedigree database resource. Potato Res 50:45–57. doi:10.1007/s11540-007-9028-3

Bink M, Uimari P, Sillanpaa J, Janss G, Jansen C (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor Appl Genet 104:751–762. doi:10.1007/s00122-001-0796-x

Bradshaw JE, Hackett CA, Meyer RC, Milbourne D, McNicol JW, Phillips MS, Waugh R (1998) Identification of AFLP and SSR markers associated with quantitative resistance to *Globodera pallida* (Stone) in tetraploid potato (*Solanum tuberosum subsp. tuberosum*) with a view to marker-assisted selection. Theor Appl Genet 97:202–210. doi:10.1007/s001220050886

Braun A, Wenzel G (2004) Molecular analysis of genetic variation in potato (*Solanum tuberosum* L.). I. German cultivars and advanced clones. Potato Res 47:81–92. doi:10.1007/BF02731971

Broyles B, Wyatt R (1991) Effective pollen dispersal in a natural population of *Asclepias exaltata*: the influence of pollinator behavior, genetic similarity, and mating success. Am Nat 138:1239–1249. doi:10.1086/285280

Buteler MI, Labonte DR, Macchiavelli RE (1997) Determining paternity in polyploids: hexaploid simulation studies. Euphytica 96:353–361. doi:10.1023/A:1003096429238

Chakraborty R, Stivers DN (1996) Paternity exclusion by DNA markers: effects of paternal mutations. J Forensic Sci 41:671–677

Chakraborty R, Stivers DN, Zhong Y (1996) Estimation of mutation rates from parentage exclusion data: applications to STR and VNTR loci. Mutat Res 354:41–48. doi:10.1016/0027-5107(96)00014-0

Côté MJ, Leduc L, Reid A (2013) Evaluation of simple sequence repeat (SSR) markers established in Europe as a method for the identification of potato varieties grown in Canada. Am Pot J 90:340–350. doi:10.1007/s12230-013-9310-7

Demeke T, Lynch DR, Kawchuk LM, Kozub GC, Armstrong JD (1996) Genetic diversity of potato determined by random amplified polymorphic DNA analysis. Plant Cell Rep 15:662–667. doi:10.1007/BF00231920

Devlin B, Roeder K, Ellstrand NC (1988) Fractional paternity assignment: theoretical development and comparison to other methods. Theor Appl Genet 76:369–380. doi:10.1007/BF00265336

D'hoop BB, Paulo MJ, Kowitwanich K, Sengers M, Visser RGF, Van Eck HJ, van Eeuwijk FA (2010) Population structure and linkage disequilibrium unravelled in tetraploid potato. Theor Appl Genet 121:1151–1170. doi:10.1007/s00122-010-1379-5

Double MC, Cockburn A, Barry SC, Smouse PE (1997) Exclusion probabilities for single-locus paternity analysis when related males compete for matings. Mol Ecol 6:1155–1166. doi:10.1046/j.1365-294X.1997.00291.x

Dow B, Ashley M (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Mol Ecol 5:615–627. doi:10.1111/j.1365-294X.1996.tb00357.x

Ellstrand NC (1984) Multiple paternity within the fruits of the wild radish, *Raphanus sativus*. Am Nat 123:819–828. doi:10.1086/284241

Ellstrand NC, Devlin B, Marshall DL (1989) Gene flow by pollen into small populations: data from experimental and natural stands of wild radish. Proc Natl Acad Sci U S A 86:9044–9047. doi:10.1073/pnas.86.22.9044

Esselink GD, Nybom H, Vosman B (2004) Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting-peak ratios) method. Theor Appl Genet 109:402–408. doi:10.1007/s00122-004-1645-5

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

Ge J, Eisenberg A, Yan J, Chakraborty R, Budowle B (2011) Pedigree likelihood ratio for lineage markers. Int J Legal Med 125:519–525. doi:10.1007/s00414-010-0514-9

Gerber S, Mariette S, Streiff R, Bodenes C, Kremer A (2000) Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. Mol Ecol 9:1037–1048. doi:10.1046/j.1365-294x.2000.00961.x

Ghislain M, Spooner DM, Rodriguez F, Villamon F, Nunez J, Vasquez C, Waugh R, Bonierbale M (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, Nunez J, Ma RH, Pignataro J, Guzman F, Bonierbale M, Spooner D (2009) Robust and highly informative microsatellite-based genetic identity kit for potato. Mol Breed 23:377–388. doi:10.1007/s11032-008-9240-0

Grattapaglia D, Ribeiro VJ, Rezende GD (2004) Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for Eucalyptus. Theor Appl Genet 109:192–199. doi:10.1007/s00122-004-1617-9

Gunn PR, Trueman K, Stapleton P, Klarkowski DB (1997) DNA analysis in disputed parentage: the occurrence of two apparently false exclusions of paternity, both at short tandem repeat (STR) loci, in the one child. Electrophoresis 18:1650–1652. doi:10.1002/elps.1150180930

Hackett CA, Luo ZW (2003) TetraploidMap: construction of linkage map in a Autotetraploid Species. J Hered 94:358–359. doi:10.1093/jhered/esg066

Hamrick JL, Schnabel A (1985) Understanding the genetic structure of plant populations: some old problems and a new approach. In: Gregorius HR (ed) Population genetics in forestry. Lecture notes in biomathematics. Springer, Berlin, pp 50–70. doi:10.1007/978-3-642-48125-3_4

Haynes KG, Douches DS (1993) Estimation of the coefficient of double reduction in the cultivated tetraploid potato. Theor Appl Genet 85:857–862. doi:10.1007/BF00225029

Hosaka K, Mori M, Ogawa K (1994) Genetic relationships of Japanese potato cultivars assessed by RAPD analysis. Am Pot J 71:535–546. doi:10.1007/BF02851325

Isenegger DA, Taylor PWJ, Ford R, Franz P, McGregor GR, Hutchinson JF (2001) DNA fingerprinting and genetic relationships of potato cultivars (Solanum tuberosum L.) commercially grown in Australia. Aust J Agric Res 52:911–918. doi:10.1071/AR00161

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bulletin de la Société Vaudoise des Sciences Naturelles 44:223–270. doi:10.5169/seals-268384

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. Mol Ecol 12:2511–2523. doi:10.1046/j.1365-294X.2003.01928.x

Junge A, Brinkmann B, Fimmers R, Madea B (2006) Mutations or exclusion: an unusual case in paternity testing. Int J Legal Med 120:360–363. doi:10.1007/s00414-005-0045-y

Karaagac E, Yilma S, Cuesta-Marcos A, Vales MI (2014) Evaluation of simple sequence repeat (SSR) markers established in Europe as a method for the identification of potato varieties grown in Canada. Am Pot J 91:195–203. doi:10.1007/s12230-013-9338-8

Kawchuk LM, Lynch DR, Thomas J, Penner B, Sillito D, Kulcsar F (1996) Characterization of Solanum tuberosum simple sequence repeats and application to potato culiwar identification. Am Pot J 73:325–335. doi:10.1007/BF02849164

Krauss SL (1994) Restricted gene flow within the morphologically complex species Persoonia mollis (Proteaceae): contrasting evidence from the mating system and pollen dispersal. Heredity 73:142–154. doi:10.1038/hdy.1994.113

Lewis P, Snow A (1992) Deterministic paternity exclusion using RAPD markers. Mol Ecol 1:155–160. doi:10.1111/j.1365-294X.1992.tb00171.x

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11:2453–2465. doi:10.1046/j.1365-294X.2002.01643.x

Luo ZW, Hackett CA, Bradshaw JE, McNicol JW, Milbourne D (2000) Predicting parental genotypes and gene segregation for tetrasomic inheritance. Theor Appl Genet 100:1067–1073. doi:10.1007/s001220051388

Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to Phytophthora infestans in potato. Genetics 175:879–889. doi:10.1534/genetics.105.054932

Marshall TC, Slate J, Kruuk LE, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655. doi:10.1046/j.1365-294x.1998.00374.x

Mather K (1936) Segregation and linkage in autotetraploids. J Genetics 32:287–314. doi:10.1007/BF02982683

McGregor CE, Lambert CA, Greyling MM, Louw JH, Warnich L (2000) A comparative assessment of DNA fingerprinting techniques (RAPD, ISSR, AFLP and SSR) in tetraploid potato (Solanum tuberosum L.) germplasm. Euphytica 113:135–144. doi:10.1023/A:1003925620546

Meagher TR (1986) Analysis of paternity within a natural population of Chamaelirium luteum. 1. Identification of most-likely male parents. Am Nat 128:199–215. doi:10.1086/284554

Meagher T, Thompson E (1986) The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theor Popul Biol 29:87–106. doi:10.1016/0040-5809(86)90006-7

Milbourne D, Meyer R, Bradshaw J, Baird E, Bonar N, Provan J, Powell W, Waugh R (1997) Comparison of PCR-based marker systems for the analysis of genetic relationships in cultivated potato. Mol Breed 3: 127–136. doi:10.1023/A:1009633005390

Milbourne D, Meyer RC, Collins AJ, Ramsay LD, Gebhardt C, Waugh R (1998) Isolation, characterisation and mapping of simple sequence repeat loci in potato. Mol Gen Genet 259:233–245. doi:10.1007/s004380050809

Milligan B, McMurry C (1993) Dominant vs. codominant genetic markers in the estimation of male mating success. Mol Ecol 2:275–283. doi:10.1111/j.1365-294X.1993.tb00020.x

Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci 43:1235–1248. doi:10.2135/cropsci2003.1235

Moisan-Thiery M, Marhadour S, Kerlan MC, Dessenne N, Perramant M, Gokelaere T, Hingrat Y (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 U S A 70:3321–3323. doi:10.1073/pnas.70.12.3321

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76:5269–5273. doi:10.1073/pnas.76.10.5269

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

Reid A, Hof L, Esselink D, Vosman B (2009) Potato cultivar genome analysis. In: Burns R(ed) Humana Press, a part of Springer Science + Business Media, Totowa, NJ, New York, N.Y., pp 295–308. doi:10.1007/978-1-59745-062-1_23

Reid A, Hof L, Felix G, Rucker B, Tams S, Milczynska E, Esselink D, Uenk G, Vosman B, Weitz A (2011) Construction of an integrated microsatellite and key morphological characteristic database of potato varieties on the EU Common Catalogue. Euphytica 182:239–249. doi:10.1007/s10681-011-0462-6

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. doi:10.2135/cropsci2006.05.0336

Rodzen JA, Famula TR, May B (2004) Estimation of parentage and relatedness in the polyploid white sturgeon (Acipenser transmontanus) using a dominant marker approach for duplicated microsatellite loci. Aquaculture 232:165–182. doi:10.1016/S0044-8486(03)00450-2

Schnabel A (1998) Parentage analysis in plants: mating systems, gene flow, and relative fertilities. In: Carvalho GR (ed) Advances in molecular ecology. Ios Press, The Netherlands, pp 173–189

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

Simko I, Haynes KG, Jones RW (2004) Mining data from potato pedigrees: tracking the origin of susceptibility and resistance to Verticillium dahliae in North American cultivars through molecular marker analysis. Theor Appl Genet 108:225–230. doi:10.1007/s00122-003-1448-0

Slater AT, Wilson GM, Cogan NO, Forster JW, Hayes BJ (2014) Improving the analysis of low heritability complex traits for enhanced genetic gain in potato. Theor Appl Genet 127:809–820. doi:10.1007/s00122-013-2258-7

Smouse PE, Meagher TR (1994) Genetic analysis of male reproductive contributions in Chamaelirium luteum (L.) Gray (Liliaceae). Genetics 136:313–322

Snow AA, Lewis PO (1993) Reproductive traits and male fertility in plants: empirical approaches. Annu Rev Ecol Syst 24:331–351. doi:10.1146/annurev.es.24.110193.001555

Song YS, Hepting L, Schweizer G, Hartl L, Wenzel G, Schwarzfischer A (2005) Mapping of extreme resistance to PVY (Ry (sto)) on chromosome XII using anther-culture-derived primary dihaploid potato lines. Theor Appl Genet 111:879–887. doi:10.1007/s00122-005-0010-7

Sun G, Wang-Pruski G, Mayich M, Jong H (2003) RAPD and pedigree-based genetic diversity estimates in cultivated diploid potato hybrids. Theor Appl Genet 107:110–115. doi:10.1007/s00122-003-1215-2

Tarn TR, Tai GCC, De Long H, Murphy AM, Seabrook JEA (1992) Breeding potatoes for long-day, temperate climates. Plant breeding reviews. Wiley, New York, pp 217–332. doi:10.1002/9780470650363.ch8

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471. doi:10.1093/nar/17.16.6463

Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12:4127–4138. doi:10.1093/nar/12.10.4127

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. doi:10.1139/g95-153