African Journal of Biotechnology Vol. 11(71), pp. 13546-13559, 4 September, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB12.972
ISSN 1684–5315 ©2012 Academic Journals

Full Length Research Paper

Development of expressed sequence tags-simple sequence repeats (EST-SSRs) for *Musa* and their applicability in authentication of a *Musa* breeding population

Edwige Gaby Nkouaya Mbanjo[1,2]*, François Tchoumbougnang[2], Albert Sone Mouelle[2], Julius Enyong Oben[3], Moses Nyine[1], Carine Dochez[1,4], Morag Elizabeth Ferguson[1] and James Lorenzen[1]

[1]International Institute of Tropical Agriculture P. O. Box 30709 Nairobi, Kenya.
[2]University of Douala, Faculty of Sciences, P. O. Box 24157, Douala, Cameroon.
[3]University of Yaoundé I, Faculty of Sciences, P. O. Box 812, Yaoundé, Cameroon.
[4]Universiteitsplein 1; Universiteit Antwerpen, BE-2610 Antwerpen, Belgium.

Accepted 11 June, 2012

Maintaining identity of clones is essential in breeding programs. New EST-SSR markers have been developed for banana and used to screen a diploid population for clonal identity. A total of 410 primer pairs were designed from an EST database, validated using polyacrylamide gel electrophoresis (PAGE) and a subset was optimized for accurate genotyping on a capillary genetic analyzer. Combining PAGE and capillary electrophoresis, about 44% of the designed primers were informative in the diploid population. The majority of markers produced two alleles as expected in a diploid population. However, some showed three to four alleles, possibly indicating closely-related members of gene families. Screening of field samples using SSR markers revealed genotype identity issues in the target population. The present study demonstrates the applicability of SSRs in the establishment of parentage and relatedness between accessions. The newly-developed SSRs will be valuable tools in the understanding of *Musa* genetics, in marker-trait associations, thereby enhancing the effectiveness of breeding programs.

Key words: EST, SSRs, markers, *Musa*, breeding.

INTRODUCTION

Molecular genetic approaches in *Musa*, like other orphan crops, are often hindered by a lack of genomic resources. Large quantities of sequence data have been generated through the sequencing of expressed sequence tags (ESTs). These sequences provide a valuable resource for the discovery of SSR markers. Research into many intractable problems of banana will be aided by the use of markers leading to the improvement of this economically important crop.

Banana and plantain *(Musa* spp*.)* represent the fourth most important crop in developing countries and the eighth most important agricultural crop in the world after maize, rice, wheat, potato, cassava, soybean, and barley with annual worldwide production estimated at about 133 million tons (FAOSTAT, 2009). It constitutes a staple food and income source for millions of people, especially in Africa, with around 87% of the production remaining in domestic markets (Roux et al., 2008). Apart from their economic value, banana and plantain are also highly nutritious. They are a good source of carbohydrates (Mohapatra et al., 2010; Sharrock and Lusty, 2000);

*Corresponding author. E-mail: j.lorenzen@cgiar.org. Tel: + 255767914951.

vitamins (ß-carotene, thiamine, riboflavin, niacin, pantothenic acid, pyridoxine, folic acid, ascorbic acid (Sharrock and Lusty, 2000; Kanazawa and Sakakibara, 2000); and minerals (K, Ca, P, Fe) (Mohapatra et al., 2010).

Genotype mixtures can be a recurrent problem in banana breeding programs, as for other both clonal and seeded crops. Mix-ups could occur for various reasons such as: a) contamination with alien pollen, b) human errors, c) mislabeling, and d) lapse in sucker management. Similar experiences have been reported for other clonal crops such as cassava (Acquah et al., 2011), cocoa (Takrama et al., 2005), and strawberry (Brunings et al., 2010). Clear-cut identification of genotypes is critical to breeders and geneticists as the use of material for which parentage or identity is not clearly confirmed could jeopardize the integrity of a long-term breeding program (Khasa et al., 2003; Evans et al., 2011). The discrimination, validation, and verification of genotypes are usually performed using phenotypic descriptors. Unfortunately, visual observation may be inadequate to identify genotypes. Closely-related genotypes cannot always be easily and efficiently distinguished using phenotypic descriptors, as there may be little variability and these traits may vary with environmental factors (Al-Doss et al., 2011). DNA based identification could have a great utility in characterization and clarification of parentage-offspring relationship as well as validation of genotype identity (Fen et al., 2008), especially for large perennial crops such as banana.

DNA markers are highly heritable, relatively easy to assay, and highly abundant in number (Kumar et al., 2009). Molecular markers provide a robust, rapid and effective means to differentiate even closely related genotypes and could be used by scientists to screen breeding material at various stages to verify and/or confirm identity of clones (Takrama et al., 2005; Gomez et al., 2008). They have been used in several studies including hybrid authentication and verification of successful crosses (Ali et al., 2008; Sartie and Asiedu, 2011), to verify pedigree information (Evans et al., 2011), or for identification of clones with unknown or dubious parentage (Khasa et al., 2003). DNA-based markers have also been used to differentiate between parental lines and progenies arising from self-pollination (Gomez et al., 2008; Brunings et al., 2010), for variety differentiation and monitoring adulteration (Kwon et al., 2005; Hashemi et al., 2009), as well as to assess the genetic purity of a variety (Fen et al., 2008; Dongre et al., 2011). There are various types of molecular markers: restriction fragment length polymorphism (RFLPs), random amplified polymorphic DNA (RAPD), inter-simple sequence repeats (ISSRs), sequence characterized region (SCARs), sequence tagged sites (STSs), cleaved amplification polymorphic sequences (CAPs), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), diversity array technology markers (DArT), single nucleotide polymorphism (SNP), etc which differ from each other in their principles and methodologies (Semagn et al., 2006; Kumar et al., 2009).

SSRs, sometimes known as microsatellites or short tandem repeats (STRs) are tandemly repeated motifs of 1 to 6 base pairs (bp) flanked by unique sequence (Tóth et al., 2000; Rabello et al., 2005). Polymorphisms associated with specific loci are due to the variation in length of the SSR, which in turn depends on the number of repetitions of the basic motif. Simple sequence repeats are a) highly polymorphic; b) reproducible; c) abundant and randomly dispersed within the genome; d) locus specific; e) multiallelic; f) easily and economically easy assayed by PCR and g) codominant and amenable to medium-throughput detection. These characteristics make them one of the most valuable markers over other marker systems (Rabello et al., 2005). The traditional method to generate SSRs involves isolation and sequencing of clones prior to primer design and validation. This is complex, expensive, time consuming and labor-intensive. Furthermore genomic SSRs often come from intergenic regions with no gene function. The availability of EST sequences enables the development of EST-SSRs markers through data mining. This approach is fast, inexpensive and efficient compared to the development of genomic SSRs. Moreover, EST-SSRs represent the transcribed part of the genome and are more transferable between species than genomic SSRs. (Gong et al., 2010; Mishra et al., 2011). The number of banana SSRs that have been described to date is relatively small, including studies by Crouch et al. (1997), Buhariwalla et al. (2005), Creste et al. (2006), Wang et al. (2008), Miller et al. (2010) and Wang et al. (2010, 2011).

The objective of the present work was to develop a new set of EST-based markers (EST-SSRs) for banana to be used to confirm genotype identity and detect possible genotype mixtures among the field samples. This represents the first step towards restoring an older diploid population developed 12 years ago (Dochez, 2004; Dochez et al., 2009) to map resistance to the burrowing nematode, *Radopholus similis* (Cobb) Thorne.

MATERIALS AND METHODS

Plant material

The plant material used in this study consisted of an AA diploid population with complex pedigree that included East Africa Highland Banana, Calcutta-4, a FHIA diploid (SH-3362) and Long Tavoy and their offspring (Figure 1). 92 progeny were generated from crosses between the female parent 6142-1 and the male parent 8075-7, of which 81 were evaluated to study the genetics of segregation for resistance to *Radopholus similis* (Dochez, 2004). A total of 376 plants were sampled corresponding to 81 genotypes with a number of plants per genotype varying from 2 to 8. 96 additional genotypes also derived from crossing parents 6142-1 and 8075-7 were later added to the existing genotypes. Progeny

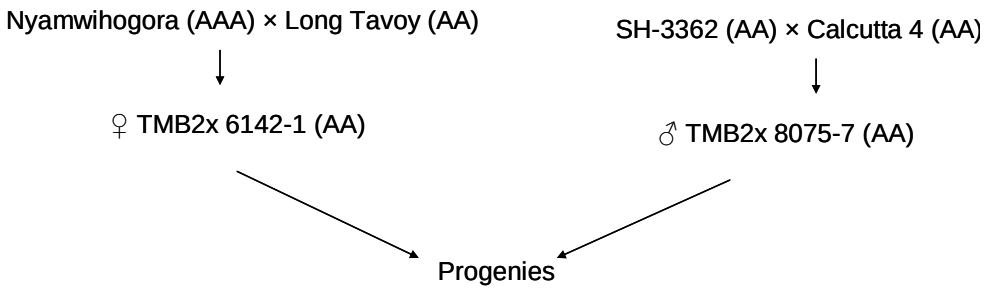

Figure 1. Pedigree of the segregating population.

genotypes were germinated *in vitro* and multiplied through tissue culture before hardening in a screenhouse and field planting (Dochez, 2004; Dochez et al., 2009).

Screening of EST database and primer design

An assembly of the *Musa* EST database available to members of the Global Musa Genomics Consortium was provided by Jun Zhuang and Chris Town of JCVI. This assembly represented 32,654 "unigenes" with 8,320 contigs and 24,329 singletons. The contigs had an average read length of 712 bp while the singletons had an average length of 472 bp. The program "Sputnik" (http://espressosoftware.com/pages/sputnik.jsp) was used to identify SSRs using a minimum quality score of 15 (min length of 17 bp). The program considered perfect as well as imperfect repeats, considering SSR motifs from di- to pentanucleotide repeats. Primer design and *in silico* validation used the program FastPCR (Kalendar et al., 2009). The primers were designed for the flanking regions of the SSR based on the following criteria: 35 to 70% GC content, a minimum melting temperature of 58 to 60°C, and absence of secondary structure. Primer lengths ranged from 19 to 30 bp and target length of amplified products from 100 to 400 bp. Unlabeled primers were synthesized from Integrated DNA Technologies (Coralville, IA, USA).

Genomic DNA extraction

Genomic DNA was extracted from fresh cigar leaves from individual genotypes as described by Mbanjo et al. (2012). The DNA concentration and DNA quality were estimated by agarose electrophoresis, and compared to serial dilution of lambda DNA phage (Promega). The gel was stained with Gel red 3X (Biotium) and visualized under UV light. The quantification and purity of obtained DNA were also evaluated at A_{260} using a Nanodrop Spectrophotometer ND-1000 V3.7 (ThermoScientific). The genomic DNA samples were diluted to a working concentration of 20 ng μl^{-1} by addition of the appropriate amount of milli-Q water.

Screening SSR primers

The newly developed EST-SSR-primer pairs were screened for amplification, polymorphism and segregation using parents of the diploid population and six randomly selected segregating individuals from the AA population. Polymerase chain reactions were conducted in a total volume of 20 µl, using 20 ng of DNA, 1X standard *Taq* reaction buffer (10 mMTris-HCl, 50 mMKCl, 1.5 mM $MgCl_2$, pH 8.3), 0.2 mM of dNTP mix, 0.12 µM of forward and reverse primers and 0.5 U of *Taq* DNA polymerase from New England Biolabs. Reactions were performed in 96-well PCR plates. The amplification profile consisted of 3 min initial denaturation step at 95°C, 35 cycles of 30 s at 95°C, 30 s annealing at 58°C, and 1 min at 72°C; concluding with a final extension of 5 min at 72°C. Primer sets which did not amplify at 58°C were later tested using the above program with an annealing temperature of 54°C.

Success of amplification was checked on 1.5% agarose gel run in 1X TBE. Band sizes were determined by comparing with 100 bp DNA ladder (Promega). Evaluation for polymorphisms and segregation was conducted for successfully amplified PCR products on 6% denaturing polyacrylamide gel (PAGE; 19:1 acrylamide:bis) in Tris-Borate-EDTA buffer for an average of 2 h. Band visualization for all products was accomplished by silver staining as described by Benbouza et al. (2006) with minor modifications, and sizes of the amplified bands were estimated by the reference DNA ladder.

Primer optimization

As the polymerase chain reaction is sensitive to reaction conditions, DNA detection protocols are best optimized for resolution on a capillary DNA genetic analyzer. Labelled primers for informative SSRs selected on the PAGE system were then optimized for the capillary ABI 3730 system. The forward primer was either modified directly with fluorochrome (6-FAM, VIC, NED or PET or HEX) at the 5' end or had a 19 bp M13 tail at the 5' end. In the latter case, a universal fluorescent - labelled M13 primer (5'-GCTACAGAGCATCTGGCTCACTGG) was added to the PCR reaction mix (Schuelke, 2000). Each SSR primer pair was optimized and then tested for polymorphism and segregation using parents and the six randomly selected progeny genotypes. The optimal annealing temperature, annealing time, number of amplification cycles, and final extension time were determined for each primer pair. The annealing temperature was established by running a temperature gradient PCR from 50 to 66°C using the TC-512 thermocycler. The quality of PCR amplification was assessed on 1.5% agarose gels before running on an ABI 3730 capillary genetic analyzer (Applied Biosystems). The size of amplified bands was determined against an internal standard (Gene Scan Liz 500; Applied Biosystems). The reaction condition that gave clear and unambiguous peaks with relative fluorescent units (rfu) above 2000 and minimum of stutter peaks on ABI 3730 was selected as the optimal PCR condition for that particular SSR primer. Samples were co-loaded in sets of 4 to 6 reactions according to fluor and allele size ranges. Allele calling was conducted using the software Genemapper V.3.7.

Screening of field samples and their cluster analysis

Eight highly informative SSR primers were selected to screen all

offspring in the field, to confirm genotype identity of individual plant and possible presence of mixtures. The transmission of each allele at each SSR locus was tracked from parents to progeny. The presence of each parental allele was scored as present *vs*. absent. Plant screening was conducted using polyacrylamide gel electrophoresis followed by silver staining. Cluster analysis was performed using Numeric Taxonomy Statistics System (NTSYS) computer package version 2.02 (Rohlf, 1998). Similarity between genotypes was estimated using the simple matching coefficient of similarity. The information generated from the similarity matrix was utilized to construct a dendrogram using Unweighted Pair Group Method with Arithmetic Mean (UPGMA) approach.

Population genetic structure analysis

139 selected individuals including 43 genotypes from the field that were consistent with regard to the labelled identity as well as 96 additional genotypes that were later added to the existing genotypes were assigned to subpopulations using the Bayesian algorithm implemented in the computer package Structure vs. 2.3 (http://pritch.bsd.uchicago.edu/structure.html). 20 selected SSRs were used to genotype these individuals using the capillary ABI 3730 instrument.Structure runs were performed with 10^6 iterations and a length burnin period of 50,000 and 100,000 MCMC replications. The admixture model and correlated allele frequency between populations were chosen to analyze data. The number of possible populations tested ranged from K=1 to K=10. Ten independent simulations were performed for each K. The true number of clusters was determined using the method of Evanno et al. (2005) based on the rate of change in the log probability of data between successive K values.

RESULTS

Screening of EST database and primer design

750 SSRs of 17 nt or longer were identified, of which 415 were suitable for primer design. The remaining 335 SSRs were not considered suitable for primer design for one of the following reasons: flanking sequences were too short or not unique for primer design, or data quality indicated problems with the sequence. Out of 415 SSR primer pairs designed, 413 primer pairs were synthesized. Of the 413 primer pairs, tri- (58%) and dinucleotide (33%) repeats were the most abundant followed by tetra-nucleotides (6%) and pentanucleotides (3%).

SSR marker screening and optimization

Genomic DNAs from the parents and randomly selected progenies were amplified with the newly developed EST-SSR primer pairs. Out of the 390 SSR primer pairs ordered and screened on polyacrylamide gel electrophoresis with parents and six randomly selected progenies, 37 (9.5%) did not give PCR products, 83 (21.3%) were monomorphic (uninformative), while 55 (14.1%) showed nonspecific amplification. 215 (55.1%) primer pairs were polymorphic between the parents and segregated in progeny genotypes of the diploid population. Labeled primers for 185 informative SSRs selected on the PAGE system were then optimized for the capillary ABI 3730 system. Out of the 185 informative SSRs selected on the PAGE system, 1 (0.5%) produced nonspecific amplicons between parents and selected individuals of the diploid population while 12 primers (6.5%) were monomorphic among progeny. There were 171 primer pairs (93.0%) that were polymorphic between parents and segregated among progeny genotypes. Combining the PAGE and capillary analysis, about 44% of the designed primers generated PCR amplicons informative in the diploid population. Table 1 provides information on the newly developed optimized markers. Other oligonucleotide sequences are available from the authors upon request (unoptimized, non-amplified, monomorphics and nonspecific primers).

The majority of microsatellites produced two alleles per locus which is congruent with the diploid nature of the crosses. However, 12 primer pairs showed 3 to 5 alleles. Some primers showed one allele per locus which is also expected (monomorphic, homozygous).

Screening of field samples and their cluster analysis

Gel scoring analysis revealed some discrepancies in hybrid identity and even parental identity. Some progeny bands were not derived from either parent. 146 plants had more than one non parental allele. The genetic similarity across the eight loci ranged from 0.63 to 1, with some hybrids showing identical profiles. Observation of the dendrogram showed that expected clonal relationships were often confirmed while in some cases, there was discrepancy in clustering of clones supposed to be identical, or inclusion of clones expected to be different (Figure 2). 128 plants from 43 genotypes were clearly differentiated, forming 43 unique clusters. However, another 130 plants distributed among 47 genotypes revealed genotype mixtures in the field as putatively identical clonal mats had differing molecular genotypes and putatively unique genotypes had identical molecular profiles (Figure 2). Genotypes that were consistent with regard to labeled identity and molecular fingerprints were selected for further analysis. One plant from each consistent genotype was selected.

Structure analysis

The optimal value of K based on the rate of change in the log probability of data between successive K value was revealed at K=2 (Figure 3). On the basis of this, we considered K=2 as being the supported number of populations. This result confirmed suspected identity mixtures in the field.

The existence of two main groups corresponding to two different half-sib populations indicated that this problem was initiated at the time of planting the crossing block.

Table 1. Information on simple sequence repeat primers optimized on ABI 3730 including flanking primer sequences, repeat motif, working annealing (Ta) temperature, expected and observed amplicon size.

Locus	Forward primer (5'-3')	Reverse primer (5'-3')	Repeat motif	Expect size	Ta	Size
AF268391	gttgttctcgtcaccatggctg	acagcagacgcacacaaacc	(CT)13	104	58	101-317
DN238160	gatttgcctctgcctcgttcc	ggatccatggcgatgaccc	(AG)15	92	58	84-96
DN238509	tccgctgatgaactgtctgtctg	gctctgaggaaggccgtacc	(CACTG)4	188	58	206-212
DN239028	ggggcctcgtcgtagcat	tgctcaaaccaaaagtcgagagcttt	(CT)14	124	58	115-123
DN239472	cgtgttacagattgtattcccttgt	cggctgcgtcaacaagcc	(GA)16	106	58	84-108
DN239771	tcccctgtatcaccacagcag	tggaccatgcattactttgctgtgaa	(TC)9	150	58	147-165
DN239791	tggaggagaccgagctgc	gagcttgagatccacgatctccac	(GGA)6	128	60	114-126
DN239853	gggccttccattgggagaaag	tcgacatcctcccggatcc	(GA)8	113	58	132-150
DN240063	gcggatgattcatcgtggaccg	acagcagcaacaatctcgtcgt	(ATG)6	148	58	140-158
DN240179	cggctgcttcaagcttcgtg	cctgcttggcttggagaagg	(CAC)7	98	58	115-118
Ma513003884	tgggcgggttcaggtcgt	ggccgaggagtaggctagct	(GATC)5	140	58	166
Ma513007351	ccctggagcaacagtgctactg	tcgaaggaacgatggccatctc	(GGA)2N(GGA)4	113	61	110-113
Ma513019043	gttaacggccacctgcatgg	acggagcagtaacacgggattg	(GAG)3GNG(GAG)2NAGGA	160	58	181-193
Ma513026332	caactttctccaagatcag	tcccaacaagcagcccgt	(GTAG)5	148	58	168-176
Ma513030283	cgtgcagtgctgttgctgtg	ccgtcaccacccacaacac	(TCTGT)5	112	61	153
Ma513032586	tggttggtggcttgcaaacg	gcctcttcactgtgttaagtgcacaa	(CTACA)5	196	58	185-195
Ma513034073	ctccctgtactcgtccatgtgg	ccgtagatggcgcggagtc	(CGC)6	129	58	123-129
Ma513035997	gaggaccaatctgcgttcgc	acgcagcacaagtcgtcca	(GA)10	101	58	91-101
Ma513036168	cgcagtagcagcaggcag	gccacagcaggatccacc	(AG)11	145	58	136-174
Ma513036776	agataacgctcgagatcgccc	cgaagcacggcgagtgttc	(TC)10	105	58	100-108
Ma513037490	cttccgctccctttcaccc	cctttgcccttcctattccggtg	(TC)8	84	58	106-124
Ma513037689	tgggatgcagtgtatccagcc	gatggaccggtgctgctc	(AGAT)7	147	60.7	144-152
Ma513037803	tcgcaagctggatgtgcg	gggcttcaccctcgatgga	(TA)13	147	60.1	142-144
Ma513037972	tcgacggagaactgcgacc	cctggagattcagggttccgc	(CT)13	114	58	109-113
Ma513038451	gtcgatcgatgatcacttcgtcatgag	caatattgcaaccctcacggacg	(GAG)6	223	61.8	61
Ma513039300	tcgacggccaccgtgaac	ccagaggagcatcccagagtg	(AG)18	139	58	116-136
Ma513039645	agcagaagtgggaggtcagg	agctaaagctgaaaggatcggacc	(TC)12	165	58	191-203
Ma513041952	agcatggccacgagcgtc	ggacagtctccgagtcggttc	(CTA)6	104	54	114-135
Ma513041960	cgcgtacggcatcgactg	tgcgacccagaaagccgt	(GGA)10	134	58	130-133
Ma513042141	acggagttggacagtgcctag	gccacttgagtcgagcaatcc	(GA)5GC(GA)7	141	58	322
Ma513042326	tcagaaggcagatcgaacagcag	ggcatcacgcgactcgac	(CAG)6	143	61.8	163-168
Ma513042336	gtgaagaacatctttggtggcctct	tgcatggagatacaaaaagaatccaacagc	(CT)9	166	58	180-206
Ma513042741	gggatattctgcagaaggtctcagc	aaccctgtcgttgcgcag	(CGC)6	105	63	124-133
Ma513042855	agaacagcccctccacctc	agctgcgctacaggcgac	(CCT)6	94	58	83-125
Ma513043169	ccaagcaatcataagttccagagtccac	gactggtgagtactgcacaggttt	(TA)14	130	60	116-132

Table 1. Contd.

Ma513043179	tgggcaaagaccggaagc	tggaaacagcagcactccaga	(TA)12	191	58	212-228
Ma513043325	gctttggatcagcagtgctttgc	cccggatatcgcatcgaatccg	(CTC)6	150	58	152-155
Ma513043580	cagcctcctcggtctcgag	ggacgagctgctcaggctc	(GGC)6	98	58	92-95
Ma513043812	caagccggaattctaccttagggac	cgacggtggcttggatggt	(GCCTC)4	92	58	115
Ma513043835	gcatacaccgagacgctatcca	acaacaccacgtacttccgca	(AG)10	201	58	189-221
Ma513043899	ctgaccgctgacacctgg	tgcccagatcgcacggac	(CGA)3TGA(CGA)7	141	58	404-408
Ma513043994	tcctttcccaacgcgccc	ggtagtgccgttcgacaacg	(TCC)7	114	58	111
Ma513044673	ttcccgaccgtacctctttcc	tgctctcgcttatttctcttcggtg	(TC)8	87	58	105-121
Ma513044812	ggctaggacctcctctggag	tgcaacgttagccggagtcat	(GAG)6	95	58	87-189
Ma513044920	tcgctttgtgatgcgtgcac	cgttggcattgattgatatgcgtgg	(TC)13	137	64.1	154-173
Ma513044953	gttcgggtgatgatggcacc	ccaaacagcaccgtaggctg	(AGA)8	120	58	108-216
Ma513045018	cgtttccgtcaattacaacgtcagc	gtgcagctactgccacagc	(GAA)6	97	58	88-94
Ma513045065	gtggacggcccatacacg	accggattctccgccacc	(TCC)7	128	54	124-133
Ma513045122	cgctctgtggcaggactg	gcaccgattggtcgaattagcg	(TC)10	106	54	96-128
Ma513045645	gctgtgctgagtagattgctgc	aggaaggctgacagaacctcac	(TC)11	103	58	132-142
Ma513045753	catggcttctgctgccgag	cgcacaacggctttgcgg	(GCC)6	89	58	78-84
Ma513046038	actgtccctcatgagttgcttacg	cgctacgactgggctcga	(AT)10	97	58	98-124
Ma513046154	cggtgtattgacttatcacaatgcagct	tctcccacaaaaccaaggtcgt	(ACA)6	154	58	152-155
Ma513046253	gggttagcgctttacgaaccg	gccaaacagacatcatagattcacagaagc	(TA)20	116	58	87-93
Ma513046494	tggatgcggcgctccaag	gatgacgcagctgtggtcc	(CTC)5CC(CTC)3	133	61	133
Ma513046502	agccatggacgggctctc	agaaagaccccacttcgagcc	(TTCT)5	183	54	173-189
Ma513046802	tcggagtcgcttattcggtgg	gccagcaccttagccagca	(CCG)6	127	58	146
Ma513047166	tctcggtcgctttgtttggga	tcgttgcctcgattcgaggaatg	(GA)10	103	58	98-106
Ma513047251	accggtaaccaaatgcactgc	ggtcttcggtgttggcttgg	(GT)11	169	60	164-176
Ma513047331	ggccattcggtggcttcag	agaagagattgctgctctactaccgaac	(GAC)6	102	58	99-102
Ma513047439	gtaccaggcaacacccacc	tggcaacaccaacatcgctg	(ATC)7	165	58	161-164
Ma513047481	ggtgatctctagttcatcggtgttgg	agagcccaaagtcccaaggt	(CTC)7	98	61.8	88-97
Ma513048157	tgtaatcgcctcttgtttccgtcg	gagacaacccaggccaacc	(TGC)6	137	60	136-151
Ma513048504	tgcacggagagatctgctcc	tgtcagcaagatcttaaccctgcag	(GAA)6GAG(GAA)5	147	58	116-131
Ma513048534	ccagcggttaatgatttgtgtggac	tgtttggtggagagacacacactg	(TG)10	183	60	170-178
Ma513049034	aggccattcattccttaagggtgg	gctgcagctgacccaatcg	(GA)11	137	58	127-153
Ma513049196	gggcttctttcgttagcggga	tcacggcgacgagctgct	(ACC)6	84	58	106-112
Ma513049385	tcgtcgagcaaggcaaatgc	ggagaggtcaggcacgaagg	(TC)10	133	58	131-151
Ma513049407	gcaagtttcggctgcgttgg	tcgccagcgtcttcggag	(TA)8	107	60	103-107
Ma513049550	acctagacgatgctgctgctg	gccactccaaaaccagaagtcg	(GAA)6AGAA(GAA)2	150	66	151-163
Ma513049596	gtatctgaagtgggcgccac	ctgcgctagcaaactatggcc	(TC)10	193	58	186-210
Ma513050081	gtgcgctccatcgttgttgag	gccactacccaatgcatcgag	(AG)16	104	58	91-113

Table 1. Contd.

Ma513050182	accaaatcccgtcaccgct	cgagcgctcatcactgagttcg	(CT)7CGC(CT)4	110	54	122-134
Ma513050212	ggctgcttcgttgccaagac	ggatcgcgagacatcgtgtacc	(GA)10	126	58	115-133
Ma513050755	ggctctttggtgggagggag	accctggtctgattcgatttggtg	(GA)14	129	58	123-139
Ma513050821	tccacatccatagaagatgttcctgca	ctttggcacagaggcccaatt	(TTCA)5	226	58	244-272
Ma513051273	caagggaagtgaacagaaacctctcc	agcttcctgtcgatgaggctg	(TCT)7	169	58	160-175
Ma513051385	ccaacggatatcaagggcgaaagac	gtggcttggaagtcctcctaca	(GAT)6	134	58	124-127
Ma513051490	ccgcctcttccatagctgc	atcacaaggcgcctgctg	(CT)11	201	64.1	193-213
Ma513051880	cagctatttaggcgaagatcatcggtg	tccaacaccagtcaaagctcca	(GA)9	146	58	148-186
Ma513052078	ccatggaccaaaccgtgctg	ccctctttcatcaccaacccatct	(CAG)8	180	58	160-177
Ma513052248	agcagtcgcctcttcggaa	tgcaggtgggttagccgt	(AGA)8	149	58	142-151
Ma513052458	cgtgtttgtcgtcggagctc	ggaagcacgattcaccgactcg	(CT)13	99	58	108-120
Ma513053096	cctccatcctttggccatcc	accctagtgacggcaacagag	(CT)10	93	59.3	93-109
Ta1043	gctaaactcattgaacgattgaatgcgct	tgcggttaaacaagctgcaacc	(TGA)6	137	58	159
Ta1069	agagaagcgactttgtcatgcctc	ggttcacaacaaagaggaatagaacgtctg	(CTT)11	143	54	133-148
Ta1080	cctatccgccgccaagac	ggagactctgaccgcgctc	(CTC)6	127	58	147-153
Ta1137	ggttggcagagttgtcggtg	agctcccatcattcatctgcagg	(CAG)CAT(CAG)2	131	58	129-138
Ta1138	ggctgtcttcgtgctgtgtg	ggcgggcatggtgttgtg	(CCT)7	147	58	145
Ta1349	tccgtttccctcagcggg	gctcgagttggagtcggaatcc	(CGT)6	142	61.5	252-266
Ta1384	aacttggcaacccacctgg	tgagtgcacggaaagcacatgtt	(GA)8CA(GA)3	167	58	189-201
Ta1386	cggtcggctcattgatctgctc	gcccagacccaagaatcccac	(AG)12TG(AG)4	137	63	141-163
Ta1401	acccgctattccgtttcgct	gagcatggaagaggcgttcc	(GA)10	104	61	95-109
Ta150	agagcagcagaccgcacc	cacagtggcttccgacaagc	(GAA)6	148	58	146-149
Ta1553	acgagacagatccctttcggtg	gctcatttcaccgacacgcac	(TTC)7	211	58	210-222
Ta160	ttgctaatatacgatgctgatgctgatgc	acccgtgttgatcgaacacca	(TGC)6	123	61	87-105
Ta1603	tcctcctcgcgagtgaca	cacgcgagcagtagctgc	(GA)11	132	58	124-132
Ta1693	gcaatctggtactccacctggtga	gaagcatgcatggctaaggagg	(CT)15	137	58	141-165
Ta1885	agcatatgcaaccacaacagttgc	tgcgtcataatttgagacctgcca	(GAG)7	148	58	144-147
Ta2102	tgctcgaccttcaggagtcc	cgtcgatgatatagttccgggcga	(ACC)8	150	60	138-150
Ta2139	ccgatggaagagctatccgagg	cgcctacctccatgcagagaag	(CTC)7	138	58	153-162
Ta2157	gggttgccgtcgtgagat	cgtcttctggcacgcttgg	(GA)7GG(GA)5	128	60	114-138
Ta2203	ggtgcccagatgccatgc	agacatttatccaccaaaggcttccag	(AG)9	105	58	152-155
Ta2267	ggtgcaagccaccgatgc	gatgcagtcgtgaaggccattc	(GCT)6	114	60	107-116
Ta248	ctcccaccgcgaacaatgg	acggagctgctcaccagc	(GGC)6	119	58	133-139
Ta253	ggacaaatcgacaaataagggatccatgc	acagtcatggtgggtgaggg	(CT)16	114	61	95-115
Ta2557	gactcgtctgcaaaccaacaacc	ttgcatcccacagccacg	(CTT)8	105	58	124-127
Ta276	accaggcactatcgcttaggtg	gggacgagatgtcgtcgaagc	(CGt)9	121	58	114-126
Ta2872	acgccgtcgaccttctcg	cggttctccatcaccatgacca	(CTG)6	150	58	166-175

Table 1. Contd.

Ta2955	cactacgctaacaggatagcaagtcc	tgaagttgctagtgtgtttgctgactg	(CT)15	126	61.8	105-115
Ta2979	caggaaggtctgcagcgtg	acacagtccatcccatttggacg	(CT)10	118	58	104-128
Ta3054	tgccaacagcctataatacggcag	gtatcaggggacgatcgacagtc	(TGTTT)6	148	58	139-144
Ta3090	tctacagacgttccccccac	tggcagcctcatcaaccca	(TC)8	176	58	199-205
Ta3135	cccatttggccaacacttgca	gggaggcaacattcccctc	(GA)9	143	54	230-246
Ta3183	aaggccatccggctccag	tcgagcgctacgaggatgtc	(GCC)6	87	61.8	99-111
Ta3237	agatccttgcctcatggaactcc	agcctcctagcatattaaggtctgtagc	(GGT)11	120	58	109-124
Ta3320	gccgtcgaccgcaaaacc	tcgtgctcgatcaccgatttcg	(GA)10	115	58	103-111
Ta337	agcagtttacatgaactcatgcatggag	ggtgaagaacatctttggtggcctc	(AG)8	176	61	172-198
Ta3448	gccgtgagaggagttccgac	tggaggcgcctcaagcac	(CTT)6	146	54	170-173
Ta3454	ggcgtcttggttactgtccttgg	gcaacaacaatcactgtcgtgttcca	(TC)9	149	58	146-160
Ta3455	atgacgaggcgggctcac	ggagaggagtgagcgagaactg	(GGC)7	176	63.8	247-265
Ta3550	ccctgatcgatcccaatcggag	gacaacgcccgccacgaa	(AGA)7	169	58	152-179
Ta3625	gcctcctggaatccgaaacc	tcgaccggagattctcgctg	(AGA)7	118	58	138-141
Ta3709	gagcgtggcggtacacac	tcagaagcgcaaccctgc	(TC)8	105	58	218-242
Ta3738	gctgagtcgtataactttggatccatggac	gctcagcaaccggctacag	(CTG)7	155	58	150-156
Ta376	cgccattgcatttgctaatggct	tgttgatcgaaacagtagacagtacacgt	(GA)19	150	58	158-176
Ta3816	tgggttgggtgccatgac	agggcagtttctcgagacgtc	(GAA)8	122	58	117-147
Ta3938	tctggcccgcccactaaa	aaccatcacagagaactgtttggctt	(CTA)6	127	64.1	138-159
Ta4077	agccgtgctggatcaccg	gctgaacagagtaacatgtgccatcac	(TC)5TTTC(TC)9	144	58	129-147
Ta4134	cccattattgcacagaaattcggcac	aggatgaatctccaagtttacaggaacca	(CAG)9	136	60	172-190
Ta4184	tgggtgaacacacacacacct	tggggagacatgaggccattt	(TGT)7	155	58	147-159
Ta4187	tggatcaacctgtcctccaagg	caaggtgagcatgtcacagcg	(CCT)6	97	58	109-118
Ta4501	cctccgcatttcgcaagcg	tgggggattcttggagtttcg	(AAG)7	117	61.8	139-142
Ta4634	ggttgttgtgcaaatgcttgatgc	aagaagactaccacaagcccataagc	(GA)10	153	61.5	150-156
Ta4708	tgccggtgcaagacctcg	gcgtggatgttgcgtttgtgg	(AGA)6	138	58	159-171
Ta4750	acagtgtagcgattctactgcatcga	ggtccctcttacccgcaagac	(TATGT)5	170	58	169-173
Ta4757	ggaccccgaagagtcgtcc	tcacccagtagaagtaggccct	(GGC)6	146	58	154-169
Ta4941	cgcccttcttgagaatctgctgg	accaccagaacgaacctcacc	(GGA)6	127	58	146-155
Ta5104	tcctccgtcgccgtcttg	ggtcatcaccaaatcaccgcg	(CGT)6	100	60	94-100
Ta5199	gcaccaaatcctataagcatagaggcctt	tccgacttcgaaggctgagac	(TTC)9TGC(TTC)4	115	58	96-138
Ta5282	gaaaaggcaatcgttcctggagg	tgtttttccatgcagcagaccc	(AG)11	111	60	106-120
Ta5540	ccatgctgtgaatgcatcggag	cgcaggctgtagaagtaccacac	(TCT)7	178	58	175-183
Ta562	cgcctcgtgtttcaacgagc	agaggcaggtcacggcac	(AGC)7	134	60	125-131
Ta578	acttaccaggtcctggtgcag	actgaaccactacatcggccag	(GCA)7	118	64.1	108-114
Ta5917	accctgagggcaacggtg	ggtggctgaggaagctcctc	(GAC)7	148	58	148-151
Ta6025	gtggtgaagccgctcaagtg	cactggagttctggtgcagc	(ATCT)9	165	58	149-169

Table 1. Contd.

Ta6083	ggtcagtttgtttctgacactgggta	acgacgtacagtaaccatctcgct	(GA)10	165	60	164-176
Ta6186	cgagaacgaatcctgcgcac	gcggattgcaagacacacaaca	(TGT)9	121	58	107-122
Ta6203	ggagaagacgagagacccgct	agccgccatcaaccaacg	(CTG)6CAG(CTG)2	123	58	103-115
Ta6377	cgacggagctcaaagtccct	tgaccagccggcaaatcc	(CG)9	200	58	215-231
Ta6415	gaaccgatcggctgggtc	acggagaccacagctatggc	(CTC)6TTC(CTC)3	179	58	176-179
Ta6456	ggctgctgaaggcaagaactg	cagggcttggaaggcaagg	(AG)12	95	60.1	88-102
Ta6591	cagcttcgtgatctcacccagaa	acaccgaggatgcgctgc	(GAA)10TAA(GAA)2	148	61	122-152
Ta6670	gcattccgcctatcaagtcgctg	tgttgccaacgtagatacctgctg	(GA)13	144	61.8	111-133
Ta6799	acggcggtgtattcatcggag	tcgtcatggctacctccagg	(CA)4C(CA)9	167	58	158-164
Ta6833	gcaccactagttcttccaccacc	ggatccgggatgcagctc	(AAG)7	125	58	112-124
Ta6838	gggtcatctcgccagggat	ggagcggtcactaccaccg	(GGC)6	112	62.4	110-122
Ta6942	ctgcaaggagctggaccc	cgagaggacgacacgacgtc	(AG)13	129	54	121-135
Ta7223	gggaacaccatcaacaccgg	tccacatcagagctctcggg	(CAG)6	112	59	110-112
Ta7279	ccctaccagcacagaacaagc	caagtgctctgcgtgatcagc	(CAG)6	138	58	122-134
Ta7514	gctcagctgtccaggttgac	tgctgctgagtgaccgga	(CTG)8	145	58	135-148
Ta7568	gaggggaagcttccagactacg	tgcgccgttgccgtagac	(AG)7GG(AG)2(AG)9	148	60.8	142-154
Ta7577	tcatggaaggccaaccgc	tgatgaagaggtaccatgactcctctg	(AGC)7	105	58	84-108
Ta7676	acgaggccacccagtgat	tccacgcatgcacacagg	(CT)11	186	60	193-203
Ta775	catctgcacctgtggttgagg	tgcacgctcttcagctgc	(GAT)7	146	58	355-364
Ta7774	aaggcctgcctgctgctg	gcagggaacaaaatcctagagccac	(AGA)4	93	58	112-118
Ta7780	tgacagccaccctagaggc	gcggaatgtgggcatctctc	(TGG)5CGG(TGG)3	149	58	152
Ta7796	agagatgctggagctctgtgg	tgcattattactcgcacattacaggcag	(TCC)6	130	60	130-178
Ta7918	attgccggcgctagggtt	tgttctgccatgaccgatcagg	(CT)16	141	54	127-141
Ta7966	gctggttctggatgtggttgac	cctggatccgatccgatggac	(TGG)7	85	60.7	80-86
Ta834	accgagatccgacggagg	gagaaccaggtcctctccgag	(GAG)7	94	58	110-116

Suspect genotypes were consistent with regard to alleles from the male parent but had alleles not present in the female parent. This observation indicates that we were probably dealing with half-sib progenies sharing a common male parent. The analysis at K=2 split the sample into two groups. The first group included the suspect genotypes and the second group contained the expected set of genotypes from authentic parents (Figure 4).

At K=3 the two main groups were maintained, however group 2 split further reflecting the complex background of individuals of this group (data not presented). Investigation of individual parents from which seed was collected identified one individual female parent ("6142-1-S") that was improperly labelled and contained unique alleles not found in true 6142-1 individuals but present in suspect progeny. This parent was later identified by molecular analysis as maternal grandparent Long Tavoy, which was a parent of intended female 6142-1. Therefore, the sub-populations shared common alleles from both the common (male) parent and the two related female parents (Figure 5).

DISCUSSION

Screening of EST database and primer design

Proportion of SSR types and motifs varied in this study. Trinucleotide repeats were the most abundant. Similar observation was reported by Gong et al. (2010) and Mishra et al. (2011).

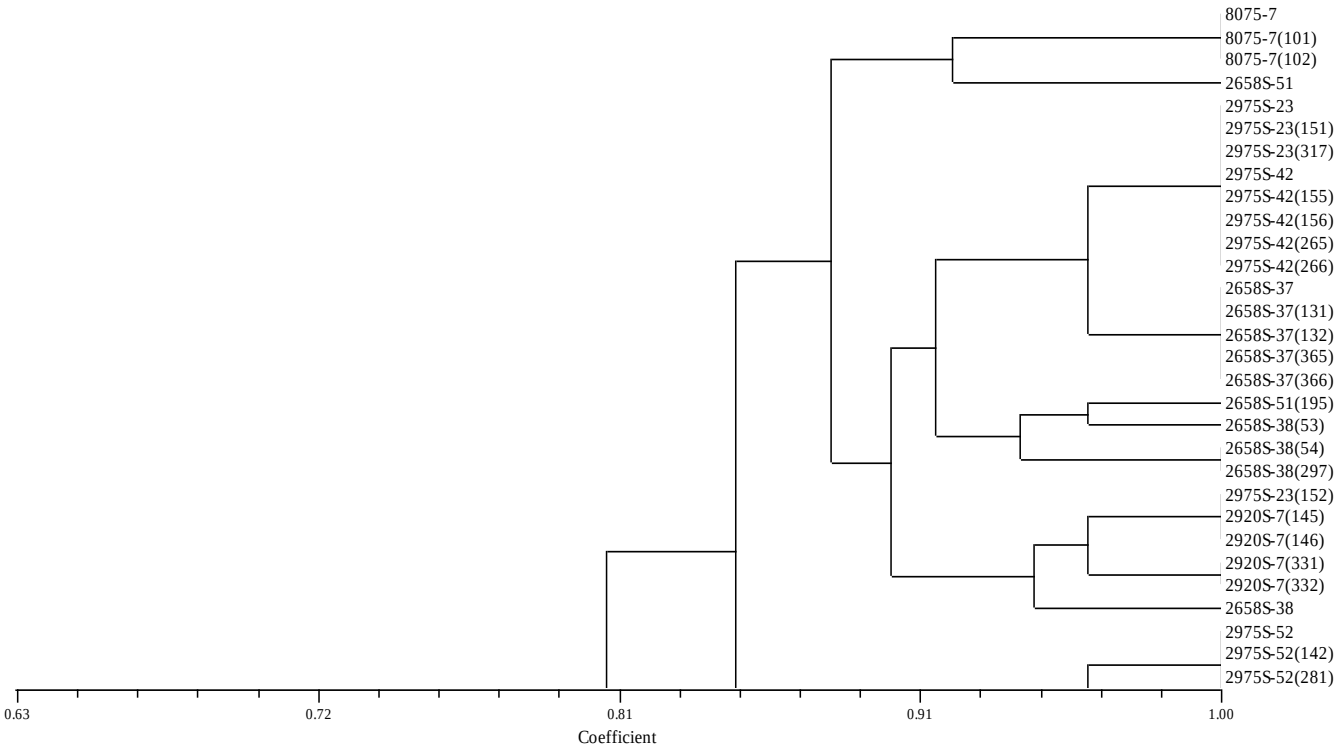

Figure 2. Clustering of some samples using UPGMA, based on simple matching coefficient of similarity. The figure in bracket represents the original recorded genotype identity of each field sample.

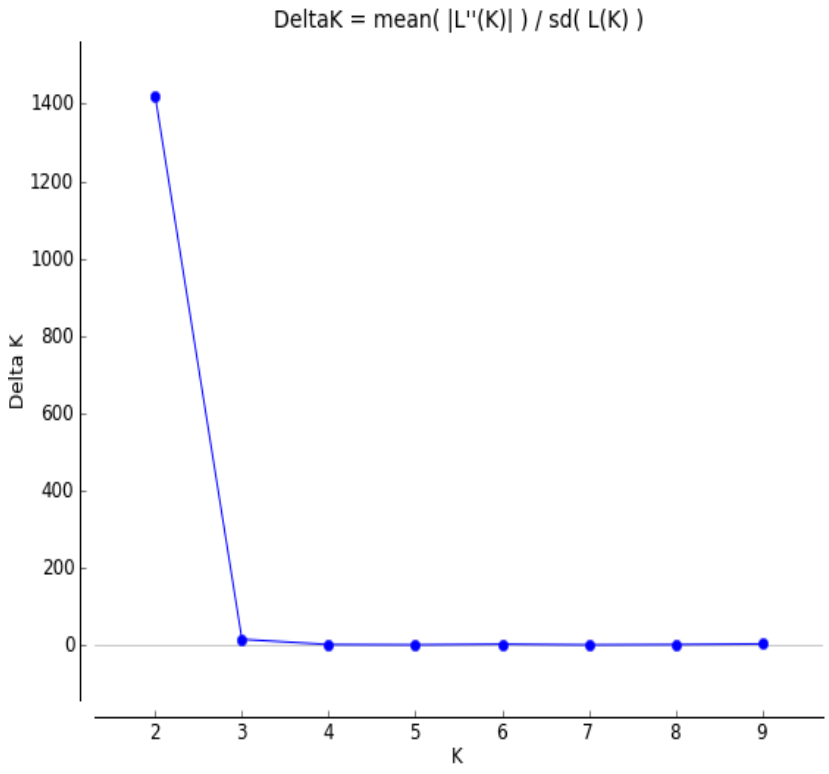

Figure 3. Determination of the optimal K value.

Figure 4. Structure analysis on the diploid population based on SSR markers. The ancestry of each population is represented by different colors.

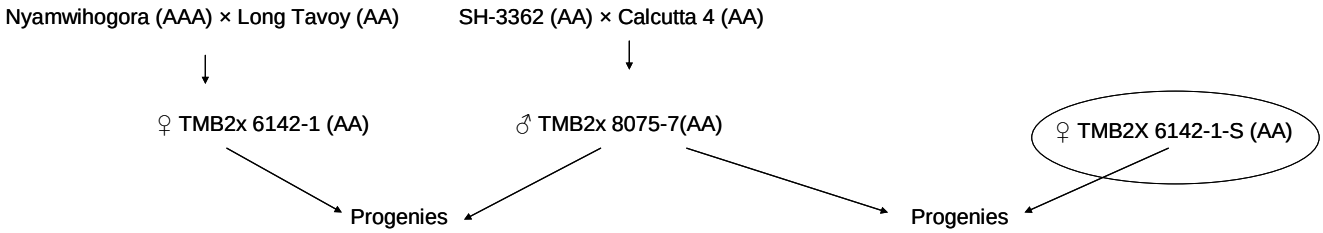

Figure 5. The final pedigree of the segregating population after hybrid verification shows the structure is two segregating half-sib populations instead of one as previously thought.

Selection against frame shift mutation may explain the high occurrence of trinucleotide repeats in EST sequences (Metzgar et al., 2000).

PCR Optimization and primer screening

Optimal PCR conditions are essential for specificity, efficiency, and sensitivity of PCR. PCR optimization reduces commonly encountered problems such as undetectable products, low yield of desired products, or the presence of non-specific PCR products (Wang et al., 2010). A few of the SSRs recorded as polymorphic after PAGE showed as monomorphic on the capillary system; showing the greater accuracy of this detection system compared to PAGE. Indeed, it can be challenging to accurately score alleles on PAGE gels. Furthermore, allele sizing for PAGE gels can be difficult due to differences in migration between lanes in the gel (Wang et al., 2009). Capillary-based microsatellite analysis provides accuracy, consistency and sensitivity for microsatellite detection essential in genetic analysis (Vemireddy et al., 2007).Therefore, SSRs that were recorded as monomorphic after PAGE should be reassessed on the capillary system. Although the advantages offer by capillary electrophoresis technology, most small to medium size laboratories cannot afford the device. Polyacrylamide gel electrophoresis remains a convenient and an affordable approach for routine SSR analyses.

Across all SSR primer pairs screened, 44% of those that amplified were polymorphic between parents and segregated in the diploid population. This percentage is quite high; considering the fact that EST-SSRs have been reported as less polymorphic than genomic SSRs in previous studies (Blair et al., 2003; Simko, 2009). A high level of polymorphism in full-sib *Musa* breeding population was reported by Crouch et al. (1999). The polymorphism rate in *Musa* seems higher compared to that observed in common beans (Ferreira et al., 2010) or tomato (Williams and Clair, 1993; Kulkarni and Deshpande, 2010). This high rate of polymorphisms could be associated with the outcrossing mating system which influences the extent of polymorphism (Glémin et al., 2006). However, our result aligns with studies conducted by Okogbenin et al. (2006) and Feingold et al. (2005), who obtained a high proportion of success and polymorphic loci with EST-derived SSRs from cassava and potato, respectively. The low level of polymorphism generally observed with EST-derived SSR has been associated to the conserved nature of coding sequences and their lower mutation rate, yet ESTs also contain non-coding sections (Feingold et al., 2005). This high level of polymorphism for outcrossing crops like banana and potato means that EST-SSRs are useful for such crops given their transferability across species and sometimes genera (Simko, 2009; Wen et al., 2010; Mishra et al., 2011).

Most primers (78.5%) produced suitable amplification owing to the stable annealing of the primers synthesized

from EST (Torada et al., 2006). This result is similar to a study conducted by Feingold et al. (2005) in potato (*Solanum tuberosum* L.), who found that 85% of the primers designed gave products of expected size. The difference between that study and the present one may be related to the poorer quality and reduced depth of the banana EST database used in developing the assembly. Some SSR amplicons were larger than expected. The unexpected product sizes obtained with some micro-satellites may be explained by the presence of introns within the amplicon (Kota et al., 2001; Feingold et al., 2005). The non-specific amplification observed for some primers may be the result of multiple priming sites along the genome (Rallo et al., 2000), for example, because of tandem repeats or larger duplications. Some SSR primer pairs did not generate PCR amplicons, possibly because of very large introns, poor source sequence or the possibility of a primer site spanning an exon/intron junction. Possible occurrence of duplicated loci with more than two alleles per diploid genotype was reported in the study. A similar observation was made for banana genotypes Calcutta 4 and Obinol`Ewai (Crouch et al., 1999). The SSRs with more than two alleles might represent closely-related members of gene families.

Relatively small number of *Musa* SSR markers has been developed to date in comparison with other crop species (Miller et al., 2010). Some of the already available SSR were developed from express sequence tags (Wang et al., 2008). Development of further molecular tools such as SSRs from transcribed regions constitutes an additional research resource for the *Musa* community. Many more SSRs will be identified with the imminent release of the reference *Musa* genome sequence (The Global Musa Genomics Consortium, 2002). The current set of SSRs could also be useful in improving the *Musa* genome assembly.

Screening of breeding material

This study demonstrates the applicability of SSRs in screening breeding material for identification of possible genotype identity mix-ups, parent-offspring analysis, and other crossing errors. They can confirm pedigree records for some genotypes, and highlight discrepancies with others. In this case, they enabled identification of a population structure problem in a diploid population and subsequent identification of the authentic parent based on inherited alleles. Non-parental bands observed with some individuals were the result of crossing error or admixture from unrelated parents. This resulted in two half-sib populations sharing a common male parent instead of one full-sib population as previously thought. Likewise, cluster analysis revealed incongruence of field plants, both mixtures within a clonal "genotype" and duplicates among "unique" genotypes. Some individuals had to be eliminated or re-named. The clonal population had been transplanted twice, so it is impossible to determine whether the identity problems dated back to the original planting or subsequent transplanting. The previously reported effectiveness of SSRs has been confirmed regarding their utility in a banana breeding program. They have successfully been used to detect off-type genotypes and confirm true hybrids (Takrama et al., 2005; Dongre et al., 2011). Brunings et al. (2010) showed that SSR could be used for high confidence authentication of genotypes and were able to differentiate even full-siblings. Also, Sartie and Asiedu (2011) confirmed successful hybridization of yam mapping population parents and the true identity of offspring using SSR markers. As expected, no selfed plants were observed in the present study. However, banana flowers are rarely perfect and flowers are routinely bagged to prevent random pollination. Genotype mix-ups are a common problem encountered in breeding programs. In the present study, the mixed genotype identity may have arisen from mislabeling at any stage from seed labeling to planting or transplanting in the field, and included progeny and at least one parent in a crossing block. Therefore, careful management of banana fields is essential to avoid intermingling of genotypes. While some cases of mixed genotype may be visually obvious, errors are often not noticeable by visual observation and a method should be in place to address such problem before errors are perpetuated as suggested by Brunings et al. (2010).

In this study, the problem was discovered 12 years after the population was established and impacted an intended mapping project. The potential of DNA markers, especially codominant multi-allelic SSR markers, to identify and correct structural/organizational problems is high. The limited number of markers used seems to be sufficient for detecting off-types in a highly polymorphic species such as banana, but for species with less polymorphism, the number of SSR primers used would probably need to be increased. Even in the present study, more markers are required to unambiguously discriminate some closely related genotypes.

Conclusion

We developed a new set of EST-SSR primers using an EST database. This approach is an efficient way to identify markers, although for many crops is being superseded by the availability of reference genomes. The newly developed markers are useful for screening hybrids, and for detecting off-types and crossing errors. They will be useful for other molecular studies, including molecular breeding.

Other possible applications include development of genetic linkage maps, quantitative analysis of economically important traits, and marker- assisted selection. These newly developed SSRs will enrich the molecular resources for *Musa*.

REFERENCES

Acquah WE, Quain DM, Twumasi P (2011). Genetic relationships between some released and elite Ghanaian cassava cultivars based on distance matrices. Afr. J. Biotechnol. 10(6):913-921.

Al-Doss AA, Elshafei AA, Moustafa KA, Saleh M., Barakat MN (2011). Comparative analysis of diversity based on morphoagronomic traits and molecular markers in durum wheat under heat stress. Afr. J. Biotechnol. 10(19):3671-3681.

Ali MA, Seyal MT, Awan SI, Niaz S, Ali S, Abbas A (2008). Hybrid authentication in upland cotton through RAPD analysis. Aust. J. Crop Sci. 2(3):141-149.

Benbouza H, Jacquemin JM, Baudoin JP, Mergeai G (2006). Optimization of a reliable, fast, cheap and sensitive silver staining method to detect SSR markers in polyacrylamide gels. Biotechnol. Agron. Soc. Environ. 10(2):77–81.

Blair MW, Pedraza F, Buendia HF, Gaiten-Solis E, Beebe SE, Gepts P, Tohme J (2003). Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. 107:1362–1374.

Brunings AM, Moyer C, Peres N, Folta KM (2010). Implementation of simple sequence repeat markers to genotype Florida strawberry varieties. Euphytica, 173:63–75.

Buhariwalla HK, Jarret RL, Jayashree B, Crouch JH, Ortiz R (2005). Isolation and characterization of microsatellite markers from *Musa balbisiana*. Mol. Ecol. Notes 5:327–330.

Creste S, Benatti TR, Orsi MR, Risterucci AM, Figueira A (2006). Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata*. Mol. Ecol. Notes 6:303–306.

Crouch JH, Crouch HK, Ortiz R, Jarret RL (1997). Microsatellite markers for molecular breeding of *Musa*. InfoMusa 6(1):5–6.

Crouch JH, Crouch HK, Tenkouano A, Ortiz R (1999). VNTR-based diversity analysis of 2x and 4x full-sib *Musa* hybrids. Electron. J. Biotechnol. 2(3):130-139.

Dochez C (2004). Breeding for resistance to *Radopholus similis* in East African highland bananas (*Musa* spp.). PhD Thesis, Katholieke Universiteit Leuven, Belgium.

Dochez C, Tenkouano A, Oritz R, Whyte J, De Waele D (2009). Host plant resistance to *Radopholus similis* in a diploid banana hybrid population. Nematology, 11(3):329-335.

Dongre AB, Raut MP, Bhandarkar MR, Meshram KJ (2011). Identification of genetic purity testing of cotton F_1 hybrid using molecular markers. Indian J. Biotechnol. 10:301-306.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611-2620.

Evans KM, Patocchi A, Rezzonico F, Mathis F, Durel CE, Fernández-Fernández F, Boudichevskaia A, Dunemann F, Stankiewicz-Kosyl M, Gianfranceschi L, Komjanc M, Lateur M, Madduri M, Noordijk Y, Van de Weg WE (2011). Genotyping of pedigreed apple breeding material with a genome-covering set of SSRs: trueness-to-type of cultivars and their parentages. Mol. Breed. 28(4):535-547.

FAOSTAT (2009). http://faostat.fao.org.

Feingold S, Lloyd J, Bonierbale M, Lorenzen J (2005). Mapping and characterization of new EST-derived microsatellites for potato (*Solanum tuberosum L*.). Theor. Appl. Genet. 111:456-466.

Fen LJ, Guo-Bin M, Ling X (2008). SSR markers for identification of purity of melon hybrids. Chin. J. Agr. Biotechnol. 5(3):223-229.

Ferreira LG, Buso GSC, Brondani RPV, Brondani C, Melo LC, Del Peloso MJ, Bassinello PZ, Sibov ST, Carneiro MS (2010). Genetic map of the common bean using a breeding population derived from the Mesoamerican gene pool. Crop Breed. Appl. Biotechnol. 10:1-8.

Glémin S, Bazin E, Charlesworth D (2006). Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc. R. Soc. B. 273:3011-3019.

Gomez SM, Denwar NN, Ramasubramanian T, Simpson CE, Burow G, Burke JJ, Puppala N, Burow MD (2008). Identification of peanut hybrids using microsatellite markers and horizontal polyacrylamide gel electrophoresis. Peanut Sci. 35:123-129.

Gong Y, Xu SC, Mao W, Hu Q, Zhang G, Ding J, Li Y (2010). Developing new SSR markers from ESTs of pea (*Pisum sativum* L). J. Biomed. Biotechnol. 11(9):702-707.

Hashemi SH, Mirmohammadi-maibody SAM, Ali G, Arzani A (2009). Identification of rice hybrids using microsatellite and RAPD markers. Afr. J. Biotechnol. 8(10):2094-2101.

Kalendar R, Lee D, Schulman AH (2009). FastPCR software for PCR primer and probe design and repeat search. Genes Genomes Genomics 3(1):1-14.

Kanazawa K, Sakakibara H (2000). High content of dopamine, a strong antioxidant, in cavendish banana. J. Agric. Food Chem. 48(3):844-848.

Khasa DP, Nadeem S, Thomas B, Robertson A, Bousquet J (2003). Application of SSR markers for parentage analysis of *Populus* clones. Forest Genet. 10(4):73-281.

Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A (2001). Generation and comparison of EST-derived SSRs and SNPs in barley *(Hordeurn vulgare L*.). Hereditas 135:145-151.

Kulkarni M, Deshpande U (2010). RAPDs based fingerprinting of tomato genotypes for identification of mutants and wild cherry specific markers. J. Plant Sci. 5(3):273-281.

Kumar P, Gupta VK, Misra AK, Modi DR, Pandey BK (2009). Potential of molecular markers in plant biotechnology. Plant Omics J. 2(4):141-162.

Kwon Y, Lee J, Yi G, Yi S, Kim K, Soh E, Bae K, Park E, Song I, Kim B (2005). Use of SSR markers to complement tests of distinctiveness, uniformity, and stability (DUS) of pepper (*Capsicum annuum* L.) varieties. Mol. Cells 19(3):428-435.

Mbanjo EGN, Tchoumbougnang F, Mouelle AS, Oben JE, Nyine M, Dochez C, Ferguson ME, Lorenzen J (2012). Molecular marker-based genetic linkage map of a diploid banana population (*Musa acuminata* Colla). Euphytica, DOI 10.1007/s10681-012-0693-1

Metzgar D, Bytof J, Wills C (2000). Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res. 10(1):72-80.

Miller RNG, Passos MAN, Menezes NNP, Souza MT, Costa MMC, Azevedo VCR, Amorim EP, Pappas GJ, Ciampi AY (2010). Characterization of novel microsatellite markers in *Musa acuminata* subsp. *burmannicoides*, var. Calcutta 4. BMC Res. Note 3:148.

Mishra RK, Gangadhar BH, Yu JW, Kim DH, Park SW (2011). Development and characterization of EST based SSR markers in Madagascar periwinkle (*Catharanthus roseus*) and their transferability in other medicinal plants. Plant Omics J. 4(3):154-162.

Mohapatra D, Mishra S, Satur N (2010). Banana and its by-product utilization: an overview. J. Sci. Ind. Res. 69:323-329.

Okogbenin E, Marin J, Fregene M (2006). An SSR-based molecular genetics map of cassava. Euphytica, 147:433–440.

Rabello E, Nunes de Souza A, Saito D, Tsai SM (2005). *In silico* characterization of microsatellites in *Eucalyptus* spp: abundance, length variation and transposon associations.Genet. Mol. Biol. 28(3):582-588.

Rallo P, Dorado G, Martín A (2000). Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). Theor. Appl. Genet*.,* 101: 984-989.

Rohlf FJ (1998) NTSYS-pc. Numerical taxonomy and multivariate analysis system, Version 2.0. Applied Biostatistics, New York, US.

Roux N, Baurens FC, Dolezel J, Hriboya E, Heslop-Harrison P, Town C, Sasashi T, Matsumoto T, Aert R, Remy S, Souza M, Lagoda P (2008).Genomics of banana and plantain (*Musa* spp.) major staple crops in the tropics. In Moore PH, Ming R (eds), Genomics of tropical crop plants. Springer, pp. 83-111.

Sartie A, Asiedu R (2011). Development of mapping populations for genetic analysis in yams (*Dioscorea rotundata* Poir. and *Dioscorea alata* L.). Afr. J. Biotechnol. 10(16):3040-3050.

Schuelke M (2000). An economic method for the fluorescent labelling of PCR fragments. Nat. Biotechnol. 18:233-234.

Semagn K, BJørnstad A, Ndjiondjop MN (2006). An overview of molecular marker methods for plants. Afr. J. Biotechnol. 5(25):2540-2568.

Sharrock S, Lusty C (2000). Nutritive value of banana. INIBAP, Montpellier, pp. 28-31.

Simko I (2009). Development of EST-SSR markers for the study of population structure in Lettuce (*Lactuca sativa* L.). J. Hered. 100(2):256–262.

Takrama JF, Cervantes-Martinez C, Phillips-Mora W, Brown JS,

Motamayor JC, Schnell RJ (2005). Determination of off-types in a cacao breeding program using microsatellites. Ingenic Newsletter. 10:2-8.
The Global Musa Genomics Consortium (2002). A strategy for the Global Musa Genomics Consortium. Report of a meeting held in Arlington, USA, 17-20 July 2001. INIBAP, Montpellier, pp. 1-44.
Torada A, Koike M, Mochida K, Ogihara Y (2006). SSR-based linkage map with new markers using an intraspecific population of common wheat. Theor. Appl. Genet. 112:1042-1051.
Tóth G, Gáspári Z, Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res. 10:967–981.
Vemireddy LR, Archak S, Nagaraju J (2007). Capillary electrophoresis is essential for microsatellite marker based detection and quantification of adulteration of basmati rice (*Oryza sativa*). J. Agric. Food Chem. 55:8112-8117.
Wang J, Lu W, Jin G, Bai Y, Li W, Hang Y (2010). Optimization of PCR system in EST-SSR analysis of *Phytophthora infestans*. J. NE Agric. 17(1):20-26.
Wang JY, Chen YY, Liu WL, Wu YT (2008). Development and application of EST-derived SSR markers for bananas (*Musa nana Lour.*). Hereditas 30(7):933-940.
Wang JY, Huang BZ, Chen YY, Feng SP, Wu YT (2011). Identification and characterization of microsatellite markers from *Musa balbisiana*. Plant Breed. 130(5):584-590.
Wang JY, Zheng LS, Huang BZ, Liu WL, Wu YT (2010). Development, characterization and variability analysis of microsatellites from a commercial cultivar of *Musa acuminata*. Genet. Res. Crop Evol. 57(4):553–563.
Wang X, Rinehart TA, Wadl PA, Spiers JM, Hadziabdic D, Windham MT, Trigiano RN (2009). A new electrophoresis technique to separate microsatellite alleles. Afr. J. Biotechnol. 8(11):2432-2436.
Wen M, Wang H, Xia Z, Zou M, Lu C, Wang W (2010). Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha Curcas* L., BMC Res. Notes 3:42.
Williams CE, Clair DA (1993). Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. Génome 36(3):619-630.