GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Development of highly polymorphic EST-SSR markers and segregation in F_1 hybrid population of *Vitis vinifera* L.

E. Kayesh[1], Y.Y. Zhang[2], G.S. Liu[3], N. Bilkish[1], X. Sun[1], X.P. Leng[1] and J.G. Fang

[1]College of Horticulture, Nanjing Agricultural University, Nanjing, China
[2]Agricultural College, Liaocheng University, Liaocheng, Shan Dong, China
[3]College of Horticulture, Qing Dao Agricultural University, Qing Dao, China

Corresponding author: J.G. Fang
E-mail: fanggg@njau.edu.cn

Genet. Mol. Res. 12 (3): 3871-3878 (2013)
Received January 29, 2013
Accepted July 28, 2013
Published September 23, 2013
DOI http://dx.doi.org/10.4238/2013.September.23.5

ABSTRACT. The objectives of this investigation were to develop and validate the expressed sequence tag (EST)-simple sequence repeat (SSR) markers from large EST sequences, and to study the segregation and distribution of SSRs within two grapevine parental lines. In total, 94 F_1 lines crossed between "Early Rose" and "Red Globe" were studied. Approximately 2100 EST-SSR sequences of *Vitis vinifera* L. were searched for SSRs and analyzed for the design of polymerase chain reaction (PCR) primers amplifying the SSR-rich regions. Trinucleotide repeats were found to be the most abundant, followed by other nucleotide repeats. A total of 182 SSR primer pairs were first developed for the study on the parental polymorphism. Among the 182 SSR primers, 142 primer pairs (78%) could amplify the anticipated PCR products, among which only 52 primer pairs (36.62%) showed polymorphism between the two parents. These polymorphic bands were further surveyed among the 94 F_1 lines, and the results showed that a total of 162 bands were amplified, and 98 of them were polymorphic in both parents (60.86%

polymorphism), with an average of 1.88 polymorphic DNA bands for each primer pair. After testing with the chi-square test, 33 of the clearly amplified polymorphic bands followed a 3:1 ratio, and 37 followed a 1:1 ratio. The rest showed distorted segregation ratios.

Key words: Grape vine; EST-SSR; Segregation; Polymorphism

INTRODUCTION

The *Vitis vinifera* L. is the most important fruit crop and is widely cultivated in the world. The grape species is representative of the genus *Vitis* and has a disomic inheritance with $2n = 38$ and a relatively small genome of 475 Mbp (Lodhi and Reisch, 1995; Aradhya et al., 2003). Fruit crop breeding is a time-consuming process, owing to the long reproductive cycles, the large plant size, and an evaluation period for productivity and quality of between 7 and 20 years.

Advancements in the field of biotechnology have provided a new technology called the molecular marker technology. It overcomes these difficulties and opens the way for new efficient breeding strategies (Morgante and Salamini, 2003; Varshney et al., 2006). In fact, marker applications to assist breeding are reported both for herbaceous (Huang et al., 1996) and woody plants (Gianfranceschi et al., 1994; Akkurt et al., 2007). In plants, genetic markers provide the framework in breeding programs via marker-assisted selection (Mazur and Tingey, 1995), map-based cloning (Tanksley et al., 1995), and anchoring physical maps (Mun et al., 2006; Troggio et al., 2007). Among the wood plants, poplar (*Populus trichocarpa* Torr. & Gray) has been the first to have its genome sequenced (Tuskan et al., 2006), whereas among the fruit trees, a draft genome sequence of grape vine (*Vitis vinifera* L.) has been published for both a near-homozygous line (Jaillon et al., 2007) and a highly heterozygous clone (Velasco et al., 2007).

Among the different types of molecular markers, microsatellite or simple sequence repeat (SSR) markers are highly prized as molecular markers owing to their co-dominance and high levels of polymorphism (Papa et al., 2005; Varshney et al., 2006). SSR markers have been introduced in the early nineties and are now being widely used in genetic characterization and diversity analyses in agricultural and horticultural crops.

The number of SSR markers available for grape vine genome mapping is still limited and only a low marker density has been achieved in the maps reported to date. A large number of polymorphic markers are required for studying diversity or linkage analyses in this crop. Moreover, the reported microsatellite sequences from grape vine have been isolated either by conventional genomic library screening procedures or generated from bacterial artificial chromosome (BAC) libraries. The presence of SSRs in the transcripts of genes suggests that they may have a role in gene expression or function (Cummings and Zoghbi, 2000). Therefore, the expansion and contraction of SSRs in genes of known function can be tested for their association with phenotypic variations or, more desirably, biological functions (Ayres et al., 1997). Although in earlier studies, SSRs were reported more in noncoding regions of eukaryotes (Hancock, 1995), a larger number of trinucleotide repeats have been reported in the coding regions of higher genomes (Li et al., 2004; Varshney et al., 2005).

The development of SSR markers requires a great deal of time, effort, and investment in the construction and screening of genomic libraries and sequencing of clones containing SSRs, primer development, and their validation. However, a large number of expressed se-

quence tags (ESTs) and genomic DNA sequences available in public databases provide an alternative method of microsatellite development. SSRs can be computationally mined from EST and genomic DNA databases (Varshney et al., 2005). EST-SSRs have been found to be extremely useful in this regard. Owing to their quality of following clear Mendelian inheritance, they can provide an anchor or reference point for specific regions of the genome. SSRs are extremely useful for the detection of polymorphisms between closely related genotypes (Sprague, 1966; Paterson et al., 1991; Troggio et al., 2007). With this background knowledge, the present study was undertaken with the following objectives: 1) to develop and validate EST-SSR markers from large EST sequences, and 2) to study the segregation ratio and distribution of SSRs within two grape vine parental lines.

MATERIAL AND METHODS

Search for microsatellites and primer design

Grape vine EST sequences were downloaded from the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/nucleotide) using MISA (Thiel et al., 2003), a Perl script able to detect perfect as well as compound microsatellites in nucleotide sequences. In considering the efficiency of the primer design and polymerase chain reaction (PCR), EST sequences of less than 100 bp were not included in the analysis. The identification of the microsatellites primer pairs was designed using the Primer3 (http://www.embnet.sk/cgi-bin/primer3_www.cgi) software, followed by oligo-6 and BLASTX (http://www.ncbi.nlm.nih.gov/blast) searches. The MISA identification tool was set with the following minimum length criteria for the extraction of repeated units (unit size/minimum number of repeats): at least six dinucleotides (2/6); at least five trinucleotides (3/5); and three tetranucleotides (4/3), pentanucleotides (5/3), and hexanucleotides (6/3).

Plant material

A total of 94 F_1 progenies crossed between “Early Rose” and “Red Globe” of grapevine were used for the polymorphism survey. Young fresh leaf samples were collected and frozen in liquid nitrogen. Finally, the leaf samples were stored at -40°C until use.

Genomic DNA isolation

Genomic DNA was extracted from the leaves using the modified cetyltrimethyl ammonium bromide method (Murray and Thompson, 1980; Bousquet et al., 1990). The extracted DNA was purified, and, after checking for quality by electrophoresis on a 0.8% agarose gel, was diluted to a final concentration of 30 ng/µL with 1X TE buffer and stored at -20°C.

PCR amplification and verification of genomic DNA

In total, 182 pairs of grape vine EST-SSRs were used to conduct the PCR amplification. The PCR amplification was carried out in a 20-µL reaction system containing 2 µL genomic DNA (30 ng/µL), 0.8 µL 10 pmol each primer, 0.1 µL *Taq* DNA polymerase (5 U/µL), 2

µL 10X buffer, 1.6 µL 25 mM $MgCl_2$, and 1.2 µL 2.5 mM dNTPs. Amplification of the reaction was performed in an Eppendorf Authorized Thermal Cycler using the following temperature cycling parameters: an initial denaturation for 5 min at 94°C; 35 cycles of denaturation at 94°C for 40 s, corresponding annealing temperature for 40 s, and extension at 72°C for 1 min; and a final extension step at 72°C for 10 min. The PCR products were resolved by non-denaturing polyacrylamide gel electrophoresis to check the DNA banding patterns.

EST analysis

Among the vector sequences, low-quality and redundant sequences were rejected with the cTrans (http://www.njau.edu.cn/down/ctrans/; Xu et al., 2007) and cap3 (http://seq.cs.iastate.edu/cap3.html; Huang and Madan, 1999) softwares.

SSR identification

Base on the established EST resources above, SSRs were identified using MISA. Compound microsatellites were defined as repeats interrupted by a non-repetitive sequence of a maximum 100 nucleotides. MISA was set with the following minimum length criteria for the extraction of repeated units (unit size/minimum number of repeats): at least six dinucleotides (2/6); at least five trinucleotides (3/5); and three tetranucleotides (4/3), pentanucleotides (5/3), and hexanucleotides (6/3).

Scoring, data organization, and analysis

In order to analyze the gene variations of two parental lines of grape vine, EST-SSR polymorphic bands were visually scored as either present (1) or absent (0) and were used to create a binary data set in which only clear unambiguous bands in non-denaturing polyacrylamide gels were chosen and scored. Data were entered into Microsoft® Excel (Microsoft Corp.) spreadsheets.

For each parental data set, all polymorphic bands were individually evaluated by the chi-square method to detect any deviation of gametic segregation from the expected Mendelian ratio ($P < 0.05$). Two independent data sets were generated that separately contained the meiotic segregation information from each parent. In the absence of phase information, each segregating locus was paired with a dummy locus, resulting in a double data set (double pseudo-test cross strategy; Weeden et al., 1994).

RESULTS AND DISCUSSION

Molecular markers are identifiable DNA sequences, found at specific locations of the genome and transmitted by the standard laws of inheritance from one generation to the next. They can be thought of as constant landmarks in the genome, since their presence (which can be detected by DNA assay) predicts the presence of the specific genes with which they are associated. From the point of view of a plant breeder, the presence of a particular marker in a sample of plant tissue predicts (with varying degrees of probability) that the trait associated with that marker will be present in the plant sampled and in its progeny. The main benefit of incorporating marker-assisted selection into a traditional plant breeding program is to increase

the efficiency of such a program and bring forward the time of varietal release, primarily by reducing or eliminating the need for costly and time-consuming field evaluation, in order to select superior lines. Grape breeding programs are increasingly using molecular tools to enhance the efficiency and speed of developing productive cultivars. Among different types of molecular markers, microsatellite or SSR markers are highly prized as molecular markers owing to their co-dominance and high levels of polymorphism. SSR markers are useful for a variety of applications in plant genetics and breeding because of their reproducibility, multi-allelic nature, co-dominant inheritance, relative abundance, and good genome coverage. They have been useful for the integration of genetic, physical, and sequence-based physical maps in plant species and simultaneously have provided breeders and geneticists with an efficient tool to link phenotypic and genotypic variations. The number of SSR markers for grape vine has been very limited.

In fact, many SSRs harboring ESTs show homology to the known genes when used for searches with BLASTX (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cg). SSR technology offers the potential of more cost-effective data acquisition than other marker technologies. SSRs amplification profiles vary in different varieties of grape vine, and this variability may be used to develop molecular markers for mapping important genes and traits in grape vine. These primers, being the derivative of gene sequences, are expected to be of immediate use in molecular marker-assisted breeding. The present study used the MISA program to identify SSR sequences from 2100 EST sequences that had been submitted to the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/Genbank). Tandem Repeat Occurrence locator (TROl) is a light-weight SSR finder based on a slight modification of the Aho-Corasick algorithm. It is a kind of dictionary-matching algorithm that locates elements of a finite set of strings (the "dictionary") within an input text. Different primer parameters, such as GC content, annealing, etc., had been analyzed while designing primers. SSR primers were designed to have an annealing temperature of between 50 and 65°C, to minimize formation of undesired hairpin or slipped duplex conformations. The percentage of GC content in primers was kept between 40 and 60% and the product sizes were set at 100 to 350 bp.

A total of 182 EST-SSR primers were used to study the segregation and distribution of SSRs within two grape vine parental lines, and DNA amplification was obtained by 142 (78%) primers. Among the 142 primers, 52 (36.62%) showed polymorphism between the parents. The list of polymorphic primers (25) is given in Table 1. These polymorphic markers were surveyed among the 94 F_1 lines. The 52 polymorphic primers produced a total of 161 bands, among which 98 were polymorphic bands in both parent (60.86% polymorphism) with an average of 1.88 polymorphic DNA bands each. When used on the progeny, among 161 amplified bands, 44 polymorphic bands were segregated into progenies and 19 were non-polymorphic bands. After testing with the chi-square test, 33 of the clearly amplified polymorphic bands followed a 3:1 ratio, along with 33 Early Rose and Red Globe that amplified 19 and 14 polymorphic bands, respectively. On the other hand, 37 clearly amplified polymorphic bands followed a 1:1 ratio, among which 35 Early Rose and Red Globe amplified 20 and 17 polymorphic bands, respectively. The rest of them showed distorted segregation ratios (Table 2). The segregation patterns of the 94 F_1 hybrid population with Parent 1 (Early Rose) and Parent 2 (Red Globe) genotypes within the species *Vitis vinifera* obtained with primer E10 are shown in Figure 1. In this study, 112 SSR loci were identified, with an efficiency of 52 polymorphic primers out of 2100 EST sequences. Another possible advantage of using EST-derived SSR markers is that once mapped, they will always be associated with the genes carrying them.

In conclusion, efforts are needed to generate EST databases for the development of

Table 1. List of expressed sequence tag SSR markers showing polymorphism in "Early Rose" x "Red Globe" parents.

SL No.	Primer No.	Primer sequences(5'→3')	Repeat motif	Tm (°C)	Expected product size (bp)	No. of total bands	No. of polymorphic bands	Polymorphism (%)	NCBI GI No.
1	E2	GGAAGCAGAAACAGCAGAGG(F) GGTGGTGTGCGGATAGACTT(R)	$(CT)_8$	58	320	3	2	66.66	161717677
2	E3	TTTTCTCGTCTTGGGGTCTG ACTGTTCGGAGGTTGACGAC	$(CT)_9$	57	169	3	3	100	161721399
3	E4	TCATCATGCAAAAACCCATC CAGCCCATATGCAAAAACCT	$(CT)_8$	58	230	3	2	66.66	161718130
4	E10	ACCGCTTCTTTGCCTCTTCT GATAAACCCCCTCCAGCAAT	$(CTC)_7$	55	140	3	2	66.66	161721396
5	E13	CTTCTCCCCTCTCCAAATCC TTTACTCCGTTCTGGCGACT	$(GGA)_{11}$	58	274	4	4	100	161720028
6	E17	GGGACGCTTTTTCAGAGATG TGGCCTTTCTTTGTCAATCC	$(GGT)_7$	58	250	3	1	33.33	161721480
7	E20	GAGATGGCTGTGGGATCATT TGCCTTTTCCTTGCACTTTT	$(ATT)_{12}$	58	169	6	4	66.66	254915138
8	E23	CAGAAGCCCAAGAAAGATCG CTTCTTTGGAGCTGGTGGAC	$(AGA)_7$	58	189	5	2	40	161718390
9	E27	CCGATGCACTTCAAACACTG TGGATTCGGCTCAGCTACTT	$(TTCTT)_3$	58	239	5	2	40	254914379
10	E30	GACCATGTTCTCTCCGCTTC CGGATGTACTCGTCCTCCAT	$(AATCCA)_3$	58	237	2	2	100	161717492
11	E39	TGGGCTCTTGTTGGGTTTAG TTCCCGTGATTCGTCTTACC	$(GTTAGG)_3$	55	270	3	1	33.33	161717962
12	E61	TGCCAAAGTTGTTCATGGAG TATGGAGTCGGGTAGCAAGG	$(TGAATC)_3$		345	2	1	50	156727817
13	E105	TACAACCCCTTCTCCTGTGG CTTCTGGTCCGACCTCTCAG	$(TTTTA)_3$	58	310	3	2	66.66	110731918
14	E117	GTCCGTACAGGAGCTTGAGG GCTAGTGACTTGCGCAACAG	$(GGTT)_3$	55	335	2	1	50	110732248
15	E119	TGGAAGCGAGAATGTCAATG GGCACACTTGCTTAGGCTCT	$(AAG)_{11}$	55	230	6	5	83.33	110732353
16	E127	GACCATGTTCTCTCCGCTTC CGGATGTACTCGTCCTCCAT	$(AATCCA)_3$		230	3	2	66.66	110732806
17	E130	CCAATGAGGGCAGCAATAAC TCAGGAACAACGCACTCAAC	$(GCA)_4$	59	250	5	3	60	110732828
18	E137	CGAGCCCATCTACTCACCTC TGTGCCGCTCCTTCTATTCT	$(TGA)_5$	60	230	3	2	66.66	110733208
19	E142	TCAGGTACGACCCTCTCAGC CGAGAATTCCCGCACATAGT	$(TAGT)_3$	60	269	2	1	50	110733143
20	E150	GGATGAAGGGCAACACATCT GAACCAATCAACCGAGCATT	$(GAAAAA)_3$	60	289	3	2	66.66	122689074
21	E155	GGTGTGGAGTGTTGGGAGAT TGGTCGCAAGTGCAACTTAT	$(TTTA)_3$	59	245	2	1	50	122689350
22	E157	CTCTGGACAACAACCCATCC GGAGGTGCAGAACAAGAAGC	$(AAAG)_3$	59	250	2	1	50	122689538
23	E166	GCAAATTGTTTCCGCAAAGT GCATTTAACATTAAGGGCCTGT	$(AAAC)_5$	59	325	3	2	66.66	122689784
24	E176	CAACGTCTCCCTTGCTTCTC TCCACACTCTGATTCGTTGC	$(TTCTCT)_3$	60	230	4	3	75	122690179
25	E182	CAAGAAGCTCCAAACCAAGC CGGCGACTTTCAAAGAGAAC	$(CTTTTT)_3$	60	245	3	2	66.66	122690385

Table 2. Summary of the segregation pattern in an F_1 hybrid population of "Early Rose" x "Red Globe".

EST-SSR analysis	Total	Early Rose	Red Globe
Polymorphic markers	52	-	-
Polymorphic band	98	53	45
Segregating ratio 1:1	37	20	17
Segregating ratio 3:1	33	19	14
Distorted segregation ratio	28	14	14

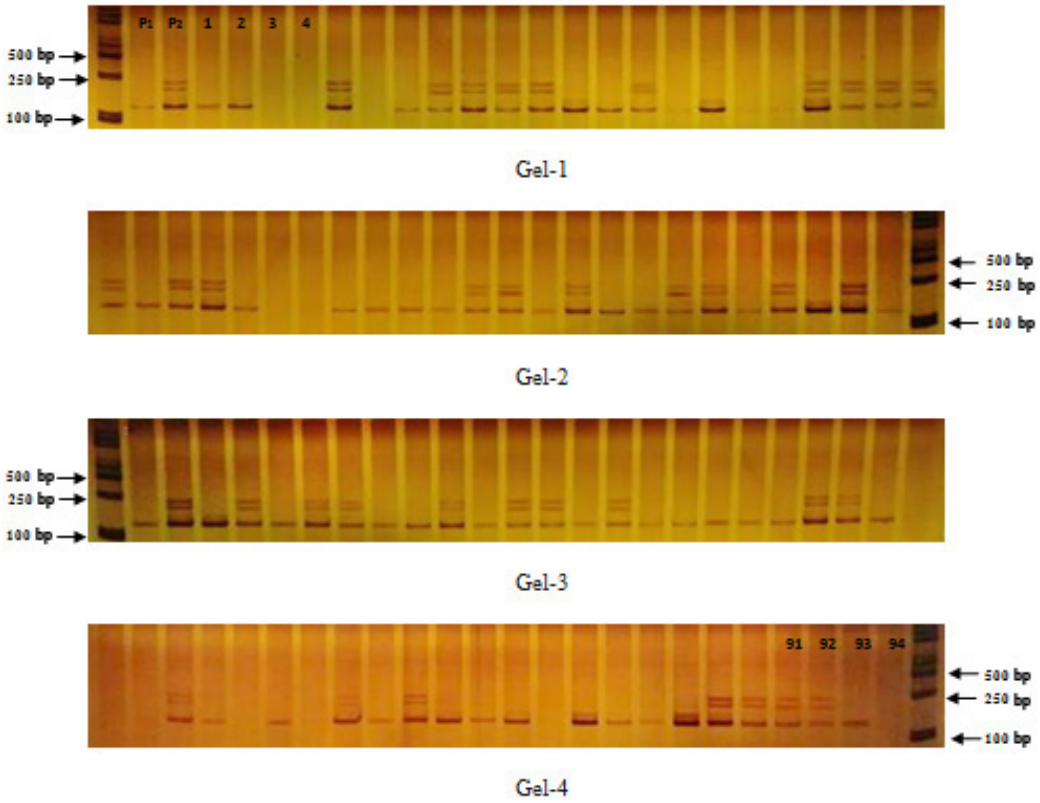

Figure 1. Polyacrylamide gel electrophoresis segregation pattern of 94 F_1 hybrids with Parent 1 (Early Rose) and Parent 2 (Red Globe) genotypes within the genus *Vitis vinifera* obtained with primer E10. *Lane M* = DL2000 plus marker.

genomic microsatellite markers that will accelerate the advancement of genomic resources and overcome the paucity of polymorphic markers in grape vine. The polymorphic markers identified will be explored both for mapping of the next generation and genetic diversity analysis.

ACKNOWLEDGMENTS

Research supported by a project "Investigation and Collection of Indigenous Varieties of Deciduous Fruit Trees in South Predominance Region (Grant #2012FY 110100-3)" of China.

REFERENCES

Akkurt M, Welter L, Maul E, Töpfer R, et al. (2007). Development of SCAR markers linked to powdery mildew (*Uncinula necator*) resistance in grapevine (*Vitis vinifera* L. and *Vitis* sp.). *Mol. Breed.* 19: 103-111.

Aradhya MK, Dangl GS, Prins BH, Boursiquot JM, et al. (2003). Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. *Genet. Res.* 81: 179-192.

Ayres NM, McClung AM, Larkin PD, Bligh HFJ, et al. (1997). Microsatellites and a single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germ plasm. *Theor. Appl. Genet.* 94: 773-781.

Bousquet J, Simon L and Lalonde M (1990). DNA amplification from vegetative and sexual tissues of trees using polymerase chain reaction. *Can. J. Res.* 20: 254-257.

Cummings CJ and Zoghbi HY (2000). Trinucleotide repeats: mechanisms and pathophysiology. *Annu. Rev. Genomics Hum. Genet.* 1: 281-328.

Gianfranceschi L, McDermott JM, Seglias N, Koller B, et al. (1994). Towards a marker assisted breeding for resistance against apple scab. *Euphytica* 77: 93-96.

Hancock JM (1995). The contribution of slippage-like processes to genome evolution. *J. Mol. Evol.* 41: 1038-1047.

Huang N, Angeles ER, Domingo J, Magpantay G, et al. (1996). Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. *Theor. Appl. Genet.* 95: 313-320.

Huang X and Madan A (1999). CAP3: A DNA sequence assembly program. *Genome Res.* 9: 868-877.

Jaillon O, Aury JM, Noel B, Policriti A, et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463-467.

Li YC, Korol AB, Fahima T and Nevo E (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21: 991-1007.

Lodhi MA and Reisch BI (1995). Nuclear DNA content of *Vitis* species, cultivars, and other genera of the vitaceae. *Theor. Appl. Genet.* 90: 11-16.

Mazur BJ and Tingey SV (1995). Genetic mapping and introgression ofgenes of agronomic importance. *Curr. Opin. Biotechnol.* 6: 175-182.

Morgante M and Salamini F (2003). From plant genomics to breeding practice. *Curr. Opin. Biotechnol.* 14: 214-219.

Moser C, Segala C, Fontana P, Salakhudtinov I, et al. (2005). Comparative analysis of expressed sequence tags from different organs of *Vitis vinifera* L. *Funct. Integr. Genomics* 5: 208-217.

Mun JH, Kim DJ, Choi HK, Gish J, et al. (2006). Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. *Genetics* 172: 2541-2555.

Murray MG and Thompson WF (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8: 4321-4325.

Papa R, Troggio M, Ajmone-Marsan P and Nonnis MF (2005). An improved protocol for the production of AFLP markers in complex genomes by means of capillary electrophoresis. *J. Anim. Breed. Genet.* 122: 62-68.

Paterson AH, Tanksley SD and Sorrells ME (1991). DNA markers in plant improvement. *Adv. Agron.* 46: 39-90.

Sprague GF (1966). Quantitative Genetics in Plant Improvement. In: Plant Breeding (Frey KJ, ed.). Lowa State Universty Press, Ames, 315-354.

Tanksley SD, Ganal MW and Martin GB (1995). Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends Genet.* 11: 63-68.

Thiel T, Michalek W, Varshney RK and Graner A (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106: 411-422.

Troggio M, Malacarne G, Coppola G, Segala C, et al. (2007). A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring Pinot Noir bacterial artificial chromosome contigs. *Genetics* 176: 2637-2650.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604.

Varshney RK, Sharma PC, Gupta PK, Balyan HS, et al. (2006). Low level of polymorphism detected by SSR probes in bread wheat. *Plant Breed.* 117: 182-184.

Varshney RK, Graner A and Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23: 48-55.

Velasco R, Zharkikh A, Troggio M, Cartwright DA, et al. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PloS One* 2: e1326.

Weeden NF, Hemmatt M, Lawson DM, Lodhi M, et al. (1994). Development and application of molecular marker linkage maps in woody fruit crops. *Euphytica* 77: 71-75.

Xu H, Yang L, Xu P, Tao Y, et al. (2007). cTrans: generating polypeptide databases from cDNA sequences. *Proteomics* 7: 177-179.