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

Characterization of SNPs in strawberry cultivars in China

A.J. Ge[1], J. Han[2], X.D. Li[3], M.Z. Zhao[4], H. Liu[2], Q.H. Dong[1] and J.G. Fang[2]

[1]Beijing Key Laboratory of New Technology in Agricultural Application, Plant Science and Technology College, Beijing University of Agriculture, Beijing, P.R. China
[2]College of Horticulture, Nanjing Agricultural University, Nanjing, P.R. China
[3]Nanjing Fujiabian Science and Technology Ltd., Nanjing, P.R. China
[4]Institute of Horticulture, Jiangsu Academy of Agricultural Sciences, Nanjing, P.R. China

Corresponding author: Q.H. Dong
E-mail: dongbac@126.com

Genet. Mol. Res. 12 (1): 639-645 (2013)
Received November 7, 2012
Accepted January 25, 2013
Published March 7, 2013
DOI http://dx.doi.org/10.4238/2013.March.7.2

ABSTRACT. Single nucleotide polymorphisms (SNPs) occur at high frequencies in both plant and animal genomes and can provide broad genome coverage and reliable estimates of genetic relationships. The availability of expressed sequence tag (EST) data has made it feasible to discover SNPs. DNA analysis is crucial in genetic studies not only for strawberry breeding programs but also for characterization of hybrids and species. We cloned 96 EST sequences, and 116 SNPs were discovered by comparing 16 strawberry cultivars grown in the region of Nanjing, China. Sequence alignment of 6 group sequences derived from 16 sample cultivars yielded 116 SNPs, within a total genomic sequence length of 1755 bp. The SNPs were discovered with a mean frequency of one SNP per 15 bp. These SNPs were comprised of 57% transitions, 32.7% transversions, 8.6% InDels, and 1.7% others, based on which a phylogenetic tree was constructed. Among the 116 SNPs, 75% were located within the open reading frame (ORF), while 25% were located outside the ORF. All 16 cultivars scattered well in the

dendrogram derived from the SNP data, demonstrating that SNPs can be a powerful tool for cultivar identification and genetic diversity analysis in strawberries.

Key words: SNPs; Strawberry; EST; Genetic diversity

INTRODUCTION

Single nucleotide polymorphisms (SNPs) can be broadly defined as any single-base substitution/indel in the genome of an individual (Primmer et al., 2002). It is the most abundant form of genetic variation in most organisms (Cho et al., 1999). The importance of SNP markers has been recognized for genetic analysis in many research areas (Brookes, 1999). They are the most common class for detection of the smallest unit of genetic variation among individuals within a species and are usually biallelic variations between individuals that occur in genes (promoter, exons or introns) or between genes (intergenic) (Rafalski, 2002). SNP frequency is higher in certain regions of the genome. Cereon Genomics (Cambridge, USA) discovered 37,344 SNPs (one SNP every 3.3 kb) and 18,579 InDels (one SNP every 6.1 kb) between the *Arabidopsis* ecotypes, *Landsberg erecta* and *Columbia* (www.arabidopsis.org). The SNP frequency between bread wheat genes from the A, B and C genomes is one SNP per 20 bp (Wolters et al., 2000), and in the 3'-UTR of 30 maize lines, it is one SNP per 70 bp and one InDel per 160 bp (Bhattramakki et al., 2002). SNPs can be found in large numbers and are more prevalent in the genome, therefore increasing the probability of finding SNP markers near the locus of interest. Currently, there is great interest in genome-wide studies, such as gene mapping (Wang et al., 1998; Yang et al., 2004), evolution (Cargill et al., 1999; Stoneking, 2001; Primmer et al., 2002) and genetic variation assessment (Salmaso et al., 2004; Shamay et al., 2006). SNPs can be the best choice for markers in the studies mentioned. The utilization of SNPs in fruit tree genetic assessment is quite recent (Salmaso et al., 2004). The availability of expressed sequence tag (EST) data has made it feasible to discover putative SNPs *in silico* prior to experimental verification. The detection of SNPs *in silico* has been employed for large-scale generation of SNP markers in a number of organisms (Marth et al., 1999; Gupta et al., 2001), including animals and crops (Picoult-Newberg et al., 1999; Khlestkina and Salina, 2006; Twito et al., 2007). With the increase in ESTs and genomic sequence availability, the detection of SNPs *in silico* is relatively easy, quick, efficient, and cost-effective. Significant efforts have been initiated for the development of a new generation of markers showing higher accuracy in the detection of polymorphisms for genetic analysis (Xiong and Jin, 1999; Hoskins et al., 2001). Further improvements and application of this SNP discovery strategy would be useful for other organisms. The objective in this study was to identify candidate SNPs for biodiversity studies and to use the isolated SNP loci to analyze genetic variation and relationships between 16 strawberry cultivars.

MATERIAL AND METHODS

Plant material

Leaf samples were obtained from Jiangsu Academy of Agricultural Sciences, Nanjing, China. The 16 strawberry cultivars used in this study are shown in Table 1.

Table 1. Cultivar names and properties of strawberry used in this study.

No.	Name of cultivar	Place of origin	Main characters
1	Honeye	America	Large fruit with a beautiful red color, firm, red clear through and is great for pies, ripe early
2	Midway	America	Large fruit with a firm, smooth texture and prominent seeds
3	Sunrise	America	Conical fruit with moderate hardness, luminous red, little acerbic taste
4	Redchief	America	Moderate fruit with sweet taste and luster
5	All Star	America	Moderate fruit, red, firm and ripe early
6	Sachinoka	Japan	Near conical fruit, deep red and ripe early
7	Chandler	America	Near conical fruit, deep red with sweet taste
8	Harunoka	Japan	Big fruit and juicy, sweet taste
9	Love berry	Japan	Big fruit with sweet taste, ripe early
10	Meiho	Japan	Moderate fruit with pink pulp, pericarp
11	Hokowase	Japan	Moderate fruit with red pulp and juicy, sweet taste
12	Marshall	America	Low growing, hardy perennial, large, deep dark red fruit in late June and has exceptional flavor
13	Anna	Espana	Big fruit and moderate in taste
14	Cartuno	America	Big fruit, near conical, orange red, firm and moderate in taste
15	Shimei4hao	China	Big fruit with conical appearance, ripe early
16	Gongsimei1hao	China	Deep red, luster, several deep groove on fruit face

DNA extraction

Young leaves of strawberry samples were selected and used for DNA extraction, using the modified CTAB method (Saghai-Maroof et al., 1984). The extracted DNA was diluted 1:10 with ddH_2O and visually inspected for quality on a 0.8% agarose gel.

ESTs

The availability of EST data has made it feasible to discover putative SNPs *in silico* prior to experimental verification. The ESTs of commercial and wild strawberry were obtained from NCBI (http://www.ncbi.nlm.nih.gov/) on March 25, 2011, and then compared by BLAST-N for candidate SNP loci discovery (Table 2).

Table 2. Ideal primers used in 16 grapevine cultivars for single nucleotide polymorphism (SNP) detection and the SNPs in 6 strawberry accessions.

No.	Primer sequence (5'-3')		Annealing temperature (°C)	No. of SNP	Size (bp)	
	Left	Right			Target	Used
2	AGACGCGCAAATAGAGCAAT	TTGGACAAGACGACGATCTG	60	11	156	156
21	TGTGATCCTTCTCAGCATGG	CATGCAGGCCTATTGGAAAT	60	10	261	261
22	GTGCTATGATTGGGCAGGTT	CTTCTCCTTGGAGGCAGTTG	60	26	288	287
23	TAATGCCCACCACAAGTTCA	CACAAAATGCACTCGGTAGG	60	23	368	365
24	GTGCTATGATTGGGCAGGTT	CACAAAATGCACTCGGTAGG	60	24	429	426
25	TACTCGCTACCGGTTCTGCT	ATGCGTGGAACAAAATCTCC	60	22	260	260

Primer design

After the BLAST results, we selected 30 EST groups that showed the presence of some SNP loci, and they were aligned for the location of conserved regions flanking two ends of a sequence, which were used as templates for designing specific primers by PRIMER3 (Rozen and Skaletsky, 2000) (Table 2). All primers were used to amplify the fragments from the 16 strawberry cultivars using genomic DNA.

PCR conditions

The SNPs detected by computer analysis were verified by PCR. Initial PCR conditions were: 2 µL DNA as the template, 15 µL ddH_2O, 2.5 µL 10X PCR buffer Mg(-), 2.0 µL $MgCl_2$, 1.5 µL dNTP, 1 µL each of forward and reverse primers, and finally 0.7 U Taq DNA polymerase, giving a total volume of 25 µL. The following describes the cycling parameters: 94°C for 3 min, and 35 cycles of 94°C for 30 s, 60°C for 30 s and 72°C for 1 min, followed by extension at 72°C for 10 min.

PCR product purification and sequencing and SNP analysis

All PCR products were purified with the AxyPrep DNA Gel Extraction kit (Axygen Scientific Inc., Union City, CA, USA). These products were cloned, and then sequenced by the Sequencing Institute (Jinsite Boitechnology, Nanjing, China). All sequences were verified by chromatogram sequence information and only the reliable sequence regions shared among all the 16 cultivars were used for determining SNP location by manual alignment using the DNAMAN® software (Lyon BioSoft, Quebec, Canada). Cluster analysis of SNP data was carried out with MAGE 4.1, as shown in Figure 1.

RESULTS

Sequence homologies

From these sequences, we designed 30 primer pairs for amplification. All 30 primer pairs designed based on the alignment of the corresponding groups of ESTs were used to run PCR using strawberry genomic DNA as template, among which 6 pairs yielded good PCR products and resulted in good sequences. After sequencing, 116 SNPs were discovered in the 96 ESTs (1755 bp) (Table 3). Thus, the SNP frequency was about one SNP per 15 bp.

Table 3. Summary of single nucleotide polymorphism (SNP) characterization.

Location	Sequence length (bp)	Transition	Tranversion	Insertion/deletion (InDel)	Others		Total	Frequency (bp/SNP)
					Transition and tranversion	Tranversion and deletion		
Total	1755	66	38	10	2	0	116	15
%		57%	32.7%	8.6%	1.7%			
ORF	1280	50	29	6	2	0	87	14.7
%		57.5%	33.3%	6.9%	2.3%			

ORF = open reading frame.

SNP identification and frequencies

After amplification and sequencing, six groups of 96 reliable sequences were aligned manually for discovery of SNP loci. A total of 116 SNPs were discovered in the 6 groups of sequences making a total length of 1755 bp. The SNP frequency was about one SNP per 15

bp and one InDel per 175.5 bp. The distribution was: 66 transitions (57%), 38 transversions (32.7%), 10 InDel (8.6%), and 2 others (1.7%) (Table 3). Eighty-seven (75%) of the SNPs were located in the open reading frame (ORF) while 29 (25%) outside the ORF. In the ORF, the SNP frequency was about one SNP per 14.7 bp and in the non-coding regions the frequency was about one SNP every 16.4 bp. It seemed that the frequencies of SNPs in the ORFs and non-coding sequences of strawberry were similar.

Genetic analysis

The inferences of relationships between the different cultivars were done through the clustering analysis. SNPs can be used to saturate genetic maps in plants (Bhattramakki and Rafalski, 2001). Cluster analysis of the 16 cultivars using the 116 SNPs was carried out as described in Material and Methods. A consensus dendrogram was constructed (Figure 1). The cultivars Meiho, Sachinoka and Hokowase showed a close genetic variation, which was consistent with the fact that they are from the same area. Chandler and Honeye also showed a close genetic variation. Although the 16 cultivars belong to different areas, they were grouped into three small sub-clusters. This revealed that the genetic variation between the cultivars was not much as expected.

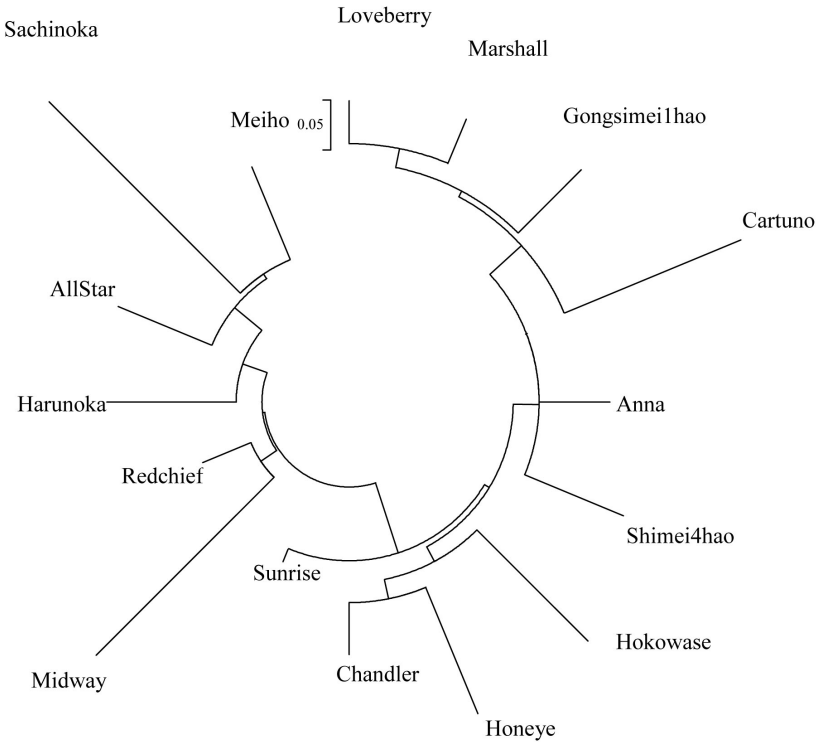

Figure 1. Unrooted phylogenetic consensus tree of the 16 strawberry cultivar analyses based on the single nucleotide polymorphism markers.

DISCUSSION

SNPs are suitable for association studies and are more common in coding sequences. Therefore, they can be utilized as molecular markers to study strawberry evolution, genetic assessment, and QTL mapping. In this study, SNPs were discovered by comparison of 96 sequences cloned from 16 strawberry cultivars as shown in Table 1. A total of 116 SNPs were discovered in a total length of 1755 bp with an SNP frequency of one SNP per 15 bp. The frequency is higher than those reported in other plant species (Jander et al., 2002; Shamay et al., 2006). SNPs could result from either transition or transversion. Theoretically, transversions should be twice as frequent as transitions; however, it has been found that in various species, transitions are much more common than transversions (Vignal et al., 2002). Among the SNP loci, transition SNPs (57%) were more than transversion SNPs (32.7%). This agrees with the findings of Yang et al. (2004) in *Lycopersicon esculentum* and Shamay et al. (2006) in *Anemone*.

Even though SNP discovery was usually carried out by aligning genomic sequence and/or ESTs from two cultivars, the frequency of SNPs could only show what happened in these two genomic samples. However, if the genomic sequences from more cultivars were aligned, more SNP loci could be found for some other different point mutations that might have happened in various cultivars other than in only two. This suggests that the frequency of SNPs may vary with the number of cultivars employed in the study. This was the reason why a much higher frequency of SNPs (one SNP every 15 bp) was found in the strawberry genomic sequences shared in all the 16 cultivars compared to those reported in most other plant species.

This study therefore indicates that SNP analysis could be a powerful method for revealing genetic diversity. Moreover, SNP detection can be easily automated and applied in the characterization and mapping of genes and haplotypes (Rafalski, 2002). The increasing number of gene sequences available in databases facilitates the development of a high number of SNP markers in model as well as in crop plants. The genetic variability in strawberry can be limited by sequencing methods. Nevertheless, we must consider that this study may not be representative of the entire strawberry group, since a limited number of individuals were evaluated for the different species. Analysis of more genotypes could possibly change some of the relationships observed. This study supplied consistent information at the level of polymorphism for the strawberry analyzed, suggesting that such markers have great potential for use in characterization and breeding programs as well as phylogenetic studies.

ACKNOWLEDGMENTS

Research supported by the Opening Foundation of Beijing Key Laboratory of New Technology in Agricultural Application (#NYXJ-2010-07) and Special Foundation of Department of Agriculture Breed Conservation (#NB2010-2130135-7).

REFERENCES

Bhattramakki D and Rafalski A (2001). Discovery and Application of Single Nucleotide Polymorphism Markers in Plants. In: Plant Genotyping: The DNA Fingerprinting of Plants (Henry RJ, ed.). CABI Publishing, Oxon, 179-191.

Bhattramakki D, Dolan M, Hanafey M, Wineland R, et al. (2002). Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. *Plant Mol. Biol.* 48: 539-547.

Brookes AJ (1999). The essence of SNPs. *Gene* 234: 177-186.

Cargill M, Altshuler D, Ireland J, Sklar P, et al. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* 22: 231-238.

Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, et al. (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* 23: 203-207.

Gupta PK, Roy JK and Prasad M (2001). Single nucleotide polymorphism: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr. Sci.* 80: 524-535.

Hoskins RA, Phan AC, Naeemuddin M, Mapa FA, et al. (2001). Single nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster*. *Genome Res.* 11: 1100-1113.

Jander G, Norris SR, Rounsley SD, Bush DF, et al. (2002). Arabidopsis map-based cloning in the post-genome era. *Plant Physiol.* 129: 440-450.

Khlestkina EK and Salina EA (2006). SNP markers: methods of analysis, ways of development, and comparison on an example of common wheat. *Genetika* 42: 725-736.

Marth GT, Korf I, Yandell MD, Yeh RT, et al. (1999). A general approach to single-nucleotide polymorphism discovery. *Nat. Genet.* 23: 452-456.

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, et al. (1999). Mining SNPs from EST databases. *Genome Res.* 9: 167-174.

Primmer CR, Borge T, Lindell J and Saetre GP (2002). Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Mol. Ecol.* 11: 603-612.

Rafalski A (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5: 94-100.

Rozen S and Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365-386.

Saghai-Maroof MA, Soliman KM, Jorgensen RA and Allard RW (1984). Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 81: 8014-8018.

Salmaso M, Faes G, Segala C, Stefanini M, et al. (2004). Genome diversity and gene haplotypes in the grapevine (*Vitis vinifera* L.), as revealed by single nucleotide polymorphisms. *Mol. Breed.* 14: 385-395.

Shamay A, Fang J, Pollak N, Yonash N, et al. (2006). Discovery of c-SNPs in *Anemone coronaria* and assessment of genetic variation. *Genet. Resour. Crop Evol.* 53: 821-829.

Stoneking M (2001). Single nucleotide polymorphisms. From the evolutionary past. *Nature* 409: 821-822.

Twito T, Weigend S, Blum S, Granevitze Z, et al. (2007). Biodiversity of 20 chicken breeds assessed by SNPs located in gene regions. *Cytogenet. Genome Res.* 117: 319-326.

Vignal A, Milan D, SanCristobal M and Eggen A (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* 34: 275-305.

Wang DG, Fan JB, Siao CJ, Berno A, et al. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280: 1077-1082.

Wolters P, Powell W, Lagudah E, Snape J, et al (2000). Nucleotide Diversity at Homologous Loci in Wheat. In: Plant and Animal Genome VIII Conference, San Diego, 9-12.

Xiong M and Jin L (1999). Comparison of the power and accuracy of biallelic and microsatellite markers in population-based gene-mapping methods. *Am. J. Hum. Genet.* 64: 629-640.

Yang W, Bai X, Kabelka E, Eaton C, et al. (2004). Discovery of single nucleotide polymorphisms in *Lycopersicon esculentum* by computer aided analysis of expressed sequence tags. *Mol. Breed.* 14: 21-34.