



Analysis of ‘Fuji’ apple somatic variants from next-generation sequencing

H.S. Lee¹, G.H. Kim¹, S.I. Kwon², J.H. Kim², Y.S. Kwon² and C. Choi¹

¹College of Agriculture and Life Science, Kyungpook National University, Daegu, Korea

²Apple Experiment Station, National Institute of Horticultural & Herbal Science, Rural Development Administration, Gunwi, Korea

Corresponding author: C. Choi
E-mail: cc31@knu.ac.kr

Genet. Mol. Res. 15 (3): gmr.15038185

Received December 2, 2015

Accepted March 11, 2016

Published August 12, 2016

DOI <http://dx.doi.org/10.4238/gmr.15038185>

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. The domesticated apple (*Malus x domestica* Borkh.) is a major fruit crop of temperate regions of the world. ‘Fuji’ apple (Ralls Genet x Delicious), a famous apple cultivar in Korea, has been very popular since its promotion in Japan in 1958. ‘Fuji’ and its bud mutant cultivars possess variable levels of genetic diversity. Nonetheless, the phenotypes of each group, which are classified into the bud mutation groups: early season, fruiting spur, and coloring, are similar. Despite attempts to identify these bud mutation cultivars, molecular markers, which were developed before the emergence of next-generation sequencing technology, have not been able to distinguish each cultivar easily. In this study, we adopted the resequencing technique using the ‘Golden Delicious’ (Grimes Golden x Unknown) apple genome as a reference. SNPs (single nucleotide polymorphisms) and InDels (insertions or deletions) of ‘Fuji’ apple and its bud mutant cultivar were detected and SNPs and unique InDels distinct to each cultivar were

identified. Data from this study may be used to identify bud mutant cultivars of ‘Fuji’ apples and be useful for further breeding of apples.

Key words: Apple; Next-generation sequencing; SNP; InDel; Resequencing; Molecular marker

INTRODUCTION

The domesticated apple (*Malus x domestica* Borkh.) is a major fruit crop of the temperate regions of the world (Velasco et al., 2010). A long juvenile period, self-incompatibility, and high heterozygosity are disadvantages, which make apple breeding difficult and time-consuming. To reduce the labor, time, orchard space, and cost for apple breeding, there is a need to identify practical traits in early-stage seedlings using accurate molecular marker-assisted selection (MAS).

In Korea, the ‘Fuji’ (Ralls Janet x Delicious) apple has been very popular after its promotion in Japan in 1958. Although ‘Fuji’ apple has high quality, preference, and vase life, genetic weaknesses require additional efforts such as bagging and the application of reflective film, which increase the cost of production. Moreover, biennial bearing, which is caused by rugged fruit shape, over-coloring, and unbalanced fertilization, as well as other drawbacks, limit high-quality fruit production. To overcome these problems, bud mutation cultivars, which have enhanced coloring among other factors, have been investigated. Consequently, numerous bud mutation cultivars have been selected and classified into three major bud mutation groups: early season group, fruiting spur group, and coloring group. The phenotypes of each group of bud mutations are similar. Furthermore, genetic variants of these breeding materials are also similar. Thus, without the use of molecular markers, identification of bud mutation cultivars is difficult.

Molecular markers are valuable tools for use in both basic and applied research for fingerprinting genotypes, analyzing genetic diversity, determining varietal identity, marker-assisted breeding, and phylogenetic analysis (Liu et al., 2013). Several markers exist, including restriction fragment length polymorphism (Botstein et al., 1980), random-amplified polymorphic DNA (Williams et al., 1990), and amplified fragment length polymorphism (Vos et al., 1995), which utilize variation in non-repetitive DNA. Other markers developed subsequently included microsatellites, or simple sequence repeats (SSR) (Tautz, 1989), which offer advantages such as reproducibility, co-dominant inheritance, and improved genome coverage. However, the complex and heterogeneous mutation patterns of SSRs can cause uncertainty in data analysis, and stutter bands, null alleles, false alleles, and homoplasy may lead to genotyping errors (Liu et al., 2013). Although MAS is essential, genome-wide polymorphism analyses and the development of reliable markers have been limited by the time-consuming and labor-intensive process. An attempt was made to identify ‘Fuji’ bud sports by retrotransposon-based sequence-specific amplified polymorphism markers (Zhao et al., 2010), but a lack of reproducibility indicated a methodological problem.

The application of NGS (next-generation sequencing) technologies has led to the effective discovery of numerous genome-wide DNA polymorphisms, such as single nucleotide polymorphisms (SNPs), and insertions or deletions (InDels) (Varshney et al., 2009), which reduce the cost of marker development (Gao et al., 2012; Lü et al., 2015). SNPs are the most abundant type of DNA sequence polymorphism (Lijavetzky et al., 2007). The discovery of large numbers of SNPs, which have been studied extensively, provides an alternative approach

to the development of high-density markers (Liu et al., 2013). SNPs have more utility and are more stable than SSRs in genetic studies and in breeding applications, including cultivar identification, genetic map construction, genetic diversity estimation, detection of genotype/phenotype association, and marker-assisted breeding (Mammadov et al., 2012). Furthermore, the emergence of the NGS technique permits *de novo* and reference-based SNP discovery and application for diverse plant species (Kumar et al., 2012). A high-quality draft genome sequence of the domesticated apple (*Malus x domestica*) was reported in 2010 (Velasco et al., 2010) and was consequently available as a reference sequence. Lijavetzky et al. (2007) confirmed that the resequencing approach is the most direct method to identify SNPs in different plant species. Nevertheless, the various rapid alignment tools available may lack the sensitivity to detect InDels at varying frequencies from heterozygotes and from heterogeneous mixtures of genotypes, as occurs in tumor mixed with normal tissue. Furthermore, this requires additional post-alignment processing upon InDel discovery (Natsoulis et al., 2013). Therefore, we focused on both SNP and InDel polymorphisms to detect somatic variations of apple. In addition, according to Srivatsan et al. (2008), whole-genome resequencing is a reliable approach used to obtain sequence information of spontaneous mutants (Kanesaki et al., 2012).

Accordingly, the goal of this study was to characterize SNP and InDel markers in apple through a resequencing approach using the 'Golden Delicious (Grimes Golden x Unknown)' apple genome as a reference, and to identify four 'Fuji' apple bud mutant cultivars and 'Fuji' apple. The cultivars were 'Danhong' and 'Hangawi', which represent the early season group and the others were 'Benishogun' and 'Yataka Fuji', representing the coloring group. These results may enable the identification of early-stage seedlings and further study may permit unstable bud mutations to be distinguished, which are caused by mericlinal and sectorial chimeras based on the tunica-carpus theory. In summary, these stable SNP markers will be fundamental to apple bud mutation research and breeding.

MATERIAL AND METHODS

The workflow is summarized in Figure 1.

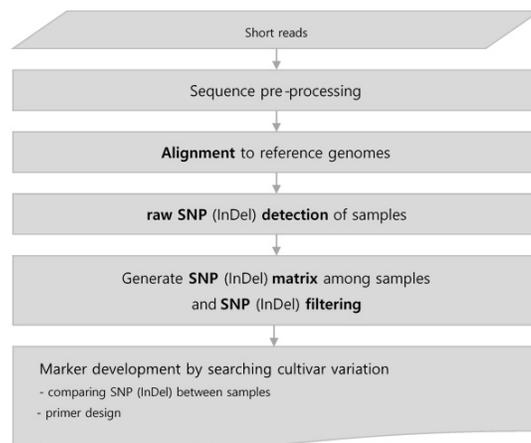


Figure 1. Summarized workflow for single nucleotide polymorphism (SNP), and insertion and deletion (InDel) detection.

Plant material and DNA extraction

We selected two cultivars from the enhanced coloring group and two from the early season group of ‘Fuji’ bud mutations (Table 1). Shoot tip tissues were collected from new shoots, for DNA extraction, at the Apple Experiment Station, National Institute of Horticultural & Herbal Science, Korea. Leaves collected from new shoots were stored in a deep freezer at -80°C until isolation. DNA was extracted using DNeasy Plant Mini Kit (QIAGEN Inc., USA) according to the manufacturer instructions. The quality and quantity of DNA were checked using the Qubit (Thermo Fisher Scientific Inc., Waltham, MA, USA) & Tecan F200-nanodrop.

Table 1. Four-bud mutation and ‘Fuji’ cultivars used in this study.

| Cultivars | Reported parentage | Year released |
|-------------------------|-------------------------|---------------|
| Fuji | Ralls Genet x Delicious | 1958 |
| Enhanced coloring group | Danhong Fuji | Sport of Fuji |
| | Hangawi Fuji | Sport of Fuji |
| Early season group | Benishogun Fuji | Sport of Fuji |
| | Yataka Fuji | Sport of Fuji |

Whole-genome resequencing and read mapping

For resequencing, a five paired-end sequencing library with an insert size of ~200 bp was constructed according to the Illumina protocol. Then, five DNA samples were sequenced on two lanes of the Illumina second-generation sequencing platform, and paired-end reads with an average length of 101 bp were produced. Next, the quality of raw sequencing reads was checked and they were trimmed using with SolexaQA package (Illumina Inc., USA). For sequencing quality and length trimming, the standard probability value was set 0.05 (between 0 and 1) and the Phred quality score was set to 20 (between 0 and 40) and the minimum read length (bp) was set to 25 (Cox et al., 2010).

The resulting short reads were subsequently aligned to the *Malus domestica* var. ‘Golden Delicious’ reference genome from National Center for Biotechnology Information (NCBI), using Burrows-Wheeler transform (BWA; 0.6.1-r104) using the default parameters, except that maximum number of gap extensions (-e) was set to 50, seed length (-l) to 30, maximum differences in the seed (-k) to 1, number of threads (-t) to 64, mismatch penalty (-M) to 6, gap open penalty (-O), and gap extension penalty (-E) was set to 15 and 8, respectively (Li and Durbin, 2009).

Discovery of SNPs and InDels by whole-genome alignment

After alignment through BWA, dependable mapped reads were deliberated for SNP and InDel calling. SNP and InDel positions from each aligned sample read relative to the reference sequence were identified using SAMtools (Trust Sanger Institute, Genome Research Limited, England) pileup2fq. SAMtools version 0.1.16 (r963:234) was adopted to detect raw SNPs and InDels between reference and test samples following the SAMtools varFilter default option with the following exceptions: the minimum mapping SNP quality (-Q) was set to 30, minimum mapping quality for gaps (-q) to 15, maximum read depth (-D) to 1000, minimum InDel score for nearby SNP filtering (-G) to 30, SNP within INT bp around a gap to be filtered (-w) was set to 15, and window size for filtering dense SNPs (-W) was input as 15 (Li et al., 2009).

To clarify the accuracy and reliability of likelihood SNPs and InDels, *SEEDERS in-house* script (SEEDERS Inc., Korea), which enables the depth, variation, and consensus quality of each SNP and InDel to be analyzed, was processed for validation. Using information on detected raw SNPs and InDels from sample read data mapped to the reference data, a matrix of total SNPs and InDels was generated with *SEEDERS in-house* script. Subsequently, miscalling values were filtered (Kim et al., 2014).

Primer design for detecting SNP and InDel variation

Polymorphic SNPs and InDels were selected by comparing homologous SNP and InDel loci among comparative varieties from the SNP and InDel matrix. Primers were designed using Primer-BLAST (NCBI), which selects primers based on the following criteria: product size 100-400 bp, primer size 20-24 bp (optimum 20 bp), and melting temperature 57°-63°C (optimum 60°C). Selected SNPs and InDel positions on the reference sequence were targeted (Untergasser et al., 2012).

Flanking region sequences for each sample were compared to not involve SNPs, InDels, or unmapped regions, but target loci. If a target region includes an un-mapping region, there will be a deletion in the base sequence; hence, primers of which target flanking sequence was not compared to validate accuracy have possibility of different result with prediction.

Amplification of InDel primers and fragment analysis

Polymerase chain reaction (PCR) amplification was conducted for selected InDel primers to identify 'Fuji' and somatic variants. PCR was performed with a final volume of 20 µL. The reaction mixture consisted of the following components: 20 ng template DNA, 10 µM both primers and 1X Hot Start *Taq* Master Mix (Philekorea Inc., Korea). PCR conditions were 5 min of pre-denaturation at 94°C, 30 cycles at 94°C for 30 s, 60°C for 30 s, and 70°C for 1 min, and 7 min of elongation at 72°C. Fragment Analyzer™ (Advanced Analytical Technologies Inc., USA) was used for the electrophoresis of PCR products. Data of analyzed InDel fragments were formatted using the PRO Size 2.0 software (Advanced Analytical Technologies). The Fragment Analyzer has a 2-bp resolution, and allele size determination for the PRO Size 2.0 software was based on regression analysis (Ban et al., 2014).

RESULTS AND DISCUSSION

DNA sequencing of 'Fuji' and bud mutation

Each lane of a 68-Gb paired-end read of a DNA sequence from 273,000,101 bp reads was generated for 'Fuji', each two cultivars of enhanced coloring group and early seasoning group. The total number of reads per variant for each lane ranged from 45,000,000 for 'Yataka Fuji' to 61,000,000 for 'Fuji', respectively. Genome coverage that represent value for division of reference genome size with total read length of each sample was revealed as 17.20-23.71 X for 'Yataka Fuji' and 'Fuji', individually. Trimmed sequencing data obtained with the SolexaQA package is presented (Table 2).

Table 2. Summary of trimmed sequenced data for ‘Fuji’ and somatic variants.

| No. | Cultivar | No. of reads | Avg. length (bp) | Total length (bp) | Trimmed/raw ^a | Genome coverage ^b |
|-----|-----------------|--------------|------------------|-------------------|--------------------------|------------------------------|
| 1 | Fuji | 53,810,550 | 96.98 | 5,218,727,134 | 83.67% | ≅ 19.31 X |
| | | 53,810,550 | 91.86 | 4,942,949,683 | 79.25% | |
| 2 | Danhong Fuji | 50,113,936 | 96.79 | 4,850,701,708 | 84.03% | ≅ 17.94 X |
| | | 50,113,936 | 91.60 | 4,590,472,470 | 79.53% | |
| 3 | Hangawi Fuji | 49,560,182 | 95.53 | 4,734,631,547 | 81.15% | ≅ 17.45 X |
| | | 49,560,182 | 89.73 | 4,447,202,246 | 76.23% | |
| 4 | Benishogun Fuji | 46,178,820 | 96.61 | 4,461,503,837 | 85.11% | ≅ 16.58 X |
| | | 46,178,820 | 92.34 | 4,264,118,068 | 81.35% | |
| 5 | Yataka Fuji | 38,247,413 | 92.43 | 3,535,187,339 | 78.13% | ≅ 12.90 X |
| | | 38,247,413 | 85.02 | 3,251,681,370 | 71.86% | |

^aTrimmed/raw: total length of trimmed read/total length of raw read. ^bGenome coverage: total read length of each sample/reference genome size (about 526 Mb).

The average genome coverage of raw sequencing data was 20.99 X; therefore, we decided that sufficient depth was retained. Overall, 80% trimmed/raw value and 16.84 X sequencing coverage arose from strong criteria for trimming may cause potential ambiguity, but sufficient number of ~92 bp reads and even trimmed/raw values were worthwhile for further mapping on reference.

Read mapping and SNP/InDel calling

Trimmed reads of each cultivar were individually aligned to the ‘Golden Delicious’ reference sequence for SNP and InDel calling by BWA (Table 3). The number of total reads was reduced, compared with the total number of reads for each lane, because overlapping reads from each lane were counted twice. Although the number of reads for ‘Yataka Fuji’ was the lowest with a noticeable gap, the ratio of the number of mapped reads was the highest at 68.81%, which was 1% above mean value. This may represent short-reads of ‘Yataka Fuji’, which possess the lowest mapped region located evenly on the chromosome with even sequencing coverage. We decided that the overall regular mapped region ratio would enable downstream analysis. Raw SNPs and InDels detected from read mapping by SAMtools were validated to generate a total SNP and InDel matrix with *SEEDERS in-house* script. Subsequently, filtered SNPs were sorted into homozygous SNPs, heterozygous SNPs, and all others (Table 4); filtered InDels were sorted in the same way (Table 5).

Table 3. Summary of mapping reads.

| Sample | No. of total reads | No. of mapped reads | Mapped region ^a (%) |
|-----------------|--------------------|---------------------|--------------------------------|
| Fuji | 107,621,100 | 73,417,529 (68.22%) | 361,212,858 (68.65%) |
| Danhong Fuji | 100,227,872 | 68,424,087 (68.27%) | 360,500,064 (68.51%) |
| Hangawi Fuji | 99,12,364 | 68,153,601 (68.76%) | 360,548,947 (68.52%) |
| Benishogun Fuji | 92,357,640 | 63,724,260 (69.00%) | 359,891,057 (68.39%) |
| Yataka Fuji | 76,494,826 | 53,403,408 (69.81%) | 357,529,432 (67.95%) |

^aMapped region: regions containing reads mapped to the reference genome.

We defined homozygous SNPs as the detected SNPs from each read that were over 90% identical. Conversely, heterozygous SNPs were those from each read that were 40-60% identical and 40-60% identical to the reference. The other SNPs were those SNPs that were

ambiguous, whether homozygous or heterozygous. The mean percentage of homozygous SNPs (820,101) was 42.52% and the mean percentage of heterozygous SNPs (562,043) was 29.14%. Because of sequencing coverage, heterozygous SNPs have ambiguity when select SNP variation so that a high percentage of homozygous SNPs has more advantages.

Table 4. Summary of single nucleotide polymorphism (SNP) detection.

| Sample | No. of total SNPs | No. of homozygous SNPs ^a | No. of heterozygous SNPs ^b | No. of remaining SNPs ^c |
|-----------------|-------------------|-------------------------------------|---------------------------------------|------------------------------------|
| Fuji | 1,744,187 | 710,516 | 529,921 | 503,750 |
| Danhong Fuji | 1,890,460 | 762,835 | 563,648 | 563,977 |
| Hangawi Fuji | 1,881,325 | 825,225 | 543,278 | 512,822 |
| Benishogun Fuji | 2,056,036 | 898,575 | 584,487 | 572,974 |
| Yataka Fuji | 2,072,148 | 903,355 | 588,883 | 579,910 |

^aNo. of homozygous SNPs: number of SNP type when detected SNPs from each mapped read to reference are consensus above 90%. ^bNo. of heterozygous SNPs: number of SNPs of this type when detected SNPs from each mapped read to reference are consensus in 40-60%. ^cNo. of remaining SNPs: indistinguishable SNP type, homozygous or heterozygous.

Table 5. Summary of InDel detection.

| Sample | No. of total InDels | No. of homozygous InDel ^a | No. of heterozygous InDels ^b | No. of reset | Homo/heterozygous InDel | | |
|-----------------|---------------------|--------------------------------------|---|--------------|-------------------------|----------|--------|
| | | | | | Insertion | Deletion | Total |
| Fuji | 149,931 | 39,814 | 39,895 | 70,222 | 31,822 | 47,887 | 79,709 |
| Danhong Fuji | 182,136 | 42,802 | 46,840 | 92,494 | 35,847 | 53,795 | 89,642 |
| Hangawi Fuji | 153,705 | 43,860 | 38,519 | 71,326 | 33,147 | 49,232 | 82,379 |
| Benishogun Fuji | 192,440 | 49,975 | 47,191 | 95,274 | 39,074 | 58,092 | 97,166 |
| Yataka Fuji | 195,216 | 49,505 | 47,895 | 97,816 | 38,946 | 58,454 | 97,400 |

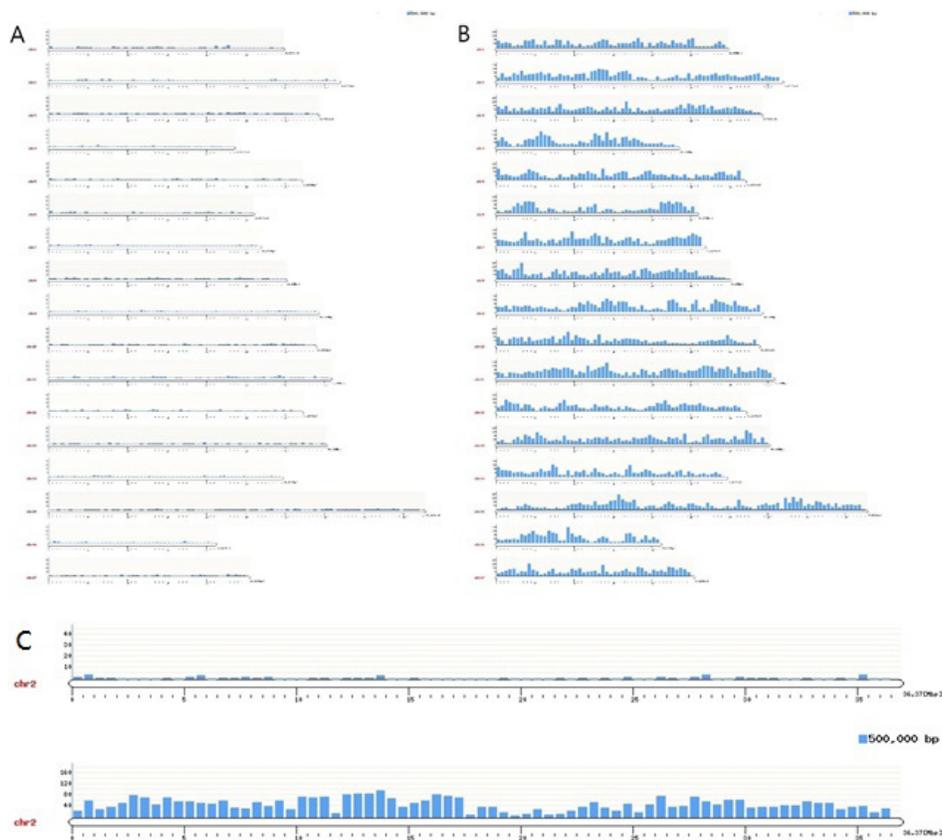
^aNo. of homozygous InDels: number of InDel type when detected InDels from each mapped read to reference are consensus above 90%. ^bNo. of heterozygous InDels: number of InDel type when detected InDels from each mapped read to reference are consensus in 40-60%.

Compared with the reference sequence, the total number of total SNPs in 'Fuji' was the lowest at 1,744,187, despite the number of total reads and the ratio of mapped regions being the highest. Conversely, the total number of SNPs for 'Yataka Fuji' was the highest at 2,072,148, which recorded the lowest number of total reads and mapping regions. Therefore, we anticipated that SNP variation for 'Yataka Fuji' would have higher sequencing depth than of other somatic variants. Furthermore, InDels were classified in the same way as SNPs. Compared with the reference sequence, and in line with the SNP data, the number of total InDels for 'Fuji' was the lowest, at 149,931. SNPs and InDel polymorphism of 'Fuji' were set as the control to select for polymorphic SNPs and InDels that detect variation within somatic variants, utilizing *SEEDERS in-house* script (Table 6).

Previous SNP data for each cultivar revealed a higher number of homozygous than heterozygous SNPs. However, homozygous SNPs derived by comparing the data with that derived from 'Fuji' were remarkably less. Despite the lowest number of total reads for 'Yataka Fuji', the highest number of homozygous and heterozygous SNPs implied that 'Yataka Fuji' has the highest level of variation with 'Fuji'. The physical distribution of SNPs specific for the identification of 'Fuji' from 'Yataka Fuji' is described in Figure 2.

Table 6. SNPs and InDels from ‘Danhong Fuji’, ‘Hangawi Fuji’, ‘Benishogun Fuji’, and ‘Yataka Fuji’ compared with those from ‘Fuji’.

| Sample | No. of homozygous SNPs | No. of heterozygous SNPs | No. of heterozygous InDels |
|-------------------------|------------------------|--------------------------|----------------------------|
| Fuji vs Danhong Fuji | 670 | 50,089 | 8053 |
| Fuji vs Hangawi Fuji | 723 | 50,162 | 8025 |
| Fuji vs Benishogun Fuji | 753 | 51,197 | 7908 |
| Fuji vs Yataka Fuji | 960 | 57,710 | 8313 |

**Figure 2.** Distribution of homozygous and heterozygous SNPs; homozygous SNPs (A) and heterozygous SNPs (B) on chromosomes that can discriminate between ‘Fuji’ and ‘Yataka Fuji’. C. Distribution of homozygous and heterozygous SNPs on chromosome 2. The y-axis represents the chromosome number and the x-axis represents the total length of chromosomes.

Selection of unique SNPs and InDels for identification

We combined SNP and InDel comparisons for the identification of variants. To detect unique SNPs and InDels for each cultivar, those from one cultivar were set as the control and used to compare those from the others, with the control changing for each comparison.

The combinations prepared were as follows: 1. 'Fuji' vs 'Danhong Fuji', 'Hangawi Fuji', 'Benishogun Fuji', and 'Yataka Fuji'; 2. 'Danhong Fuji' vs 'Fuji', 'Hangawi Fuji', 'Benishogun Fuji', and 'Yataka Fuji'; 3. 'Hangawi Fuji' vs 'Danhong Fuji', 'Fuji', 'Benishogun Fuji', and 'Yataka Fuji'; 4. 'Benishogun Fuji' vs 'Danhong Fuji', 'Hangawi Fuji', 'Fuji', and 'Yataka Fuji'; 5. 'Yataka Fuji' vs 'Danhong Fuji', 'Hangawi Fuji', 'Benishogun Fuji', and 'Fuji'. As a result, the SNPs and InDels derived from each combination were cultivar specific (Table 7). This result may reduce at least one step to discriminate 'Fuji' and other four somatic variants, and would increase the utility of application to other bud sport type to classify before specific marker test.

Table 7. SNPs and InDels used for cultivar identification.

| Sample | No. of homozygous SNPs | No. of heterozygous SNPs | No. of heterozygous InDels |
|---|------------------------|--------------------------|----------------------------|
| Fuji vs Danhong, Hangawi, Benishogun, Yataka Fuji | 5 | 1045 | 155 |
| Danhong vs Fuji, Hangawi, Benishogun, Yataka Fuji | 3 | 1264 | 190 |
| Hangawi vs Danhong, Fuji, Benishogun, Yataka Fuji | 3 | 1386 | 156 |
| Benishogun vs Danhong, Hangawi, Fuji, Yataka Fuji | 4 | 1310 | 165 |
| Yataka Fuji vs Danhong, Hangawi, Benishogun, Fuji | 7 | 2484 | 261 |

Identification of 'Fuji' and somatic variants

We preferentially performed PCR with InDel primers, since InDel polymorphisms are more convenient and advantageous to convert to molecular markers than SNPs. A total of 94 primer sets were designed using Primer-BLAST (NCBI) and were used for amplification. Through capillary analysis with Fragment Analyzer™ (Advanced Analytical Technologies Inc.), we confirmed two primers that identify 'Fuji', 'Benishogun' (Figure 3) and the rest of the primer sets were not able to distinguish sport-type mutants. Each marker specifically differentiates each cultivar from the four other somatic variants that involve 'Fuji'.

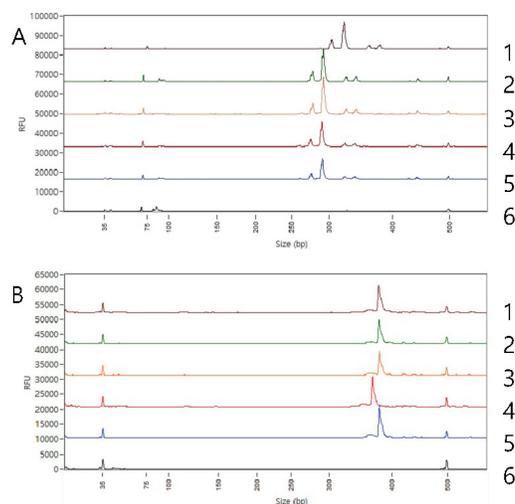


Figure 3. Identification of the InDel peak pattern generated by Fragment Analyzer. **A.** InDel peak pattern used for the identification of 'Fuji'; **B.** InDel peak pattern used for the identification of 'Benishogun'. 1: 'Fuji', 2: 'Danhong', 3: 'Hangawi', 4: 'Benishogun', 5: 'Yataka', 6: negative control.

Because of the high heterozygosity of the fruit tree, which reproduces via vegetative propagation, short-reads derived from Illumina Hi-seq may not generate haplotypes, making it difficult to discriminate somatic variants with selected InDels. When designing primers, selected InDels with ~2-bp InDel variations were excluded from fragment analysis, because according to the manufacturer instructions, the analyzer has only 2-bp resolution. Furthermore, InDel variations under 10 bp and exceeding 10 bp were also excluded, if sequencing coverage was under 8 depth, due to ambiguity. These strong selection criteria reduced the number of final primers, but also reduced ambiguity. To select more exact candidates, enhanced sequencing coverage may be needed. Without written combinations, selected SNP and InDel data that ‘Fuji’ was set as a control to compare with each somatic variants still show enormous chance to select candidate primers, and additional application of previously selected SNPs, which are converted into markers, would enable more accurate and specific identification.

In summary, although numerous candidate InDel markers were observed, 94 candidate InDel combination primers were tested to identify somatic variants of ‘Fuji’ apple. In this analysis, two InDel markers that distinguish ‘Fuji’ and its somatic variant ‘Benishogun’ were developed (Table 8).

Table 8. InDel primers used for the identification of ‘Fuji’ and ‘Benishogun’.

| ID | Classification | Chromosome No. | Reference position | Primer size | InDel variation | Product size (bp) |
|--------------------|----------------|----------------|--------------------|----------------------|-----------------|-------------------|
| Fuji | | 16 | 1855749 | F: 22mer R: 22mer | 20-bp insertion | 293 |
| Benishogun Fuji | | 15 | 40542280 | F: 24mer R: 22mer | 11-bp deletion | 394 |

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

- Ban SH, Yun WH, Kim GH, Kwon SI, et al. (2014). Genetic identification of apple cultivars bred in Korea using simple sequence repeat markers. *Hortic. Environ. Biotechnol.* 55: 531-539. <http://dx.doi.org/10.1007/s13580-014-0017-6>
- Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.
- Cox MP, Peterson DA and Biggs PJ (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11: 485. <http://dx.doi.org/10.1186/1471-2105-11-485>
- Gao Q, Yue G, Li W, Wang J, et al. (2012). Recent progress using high-throughput sequencing technologies in plant molecular breeding. *J. Integr. Plant Biol.* 54: 215-227. <http://dx.doi.org/10.1111/j.1744-7909.2012.01115.x>
- Kanesaki Y, Shiwa Y, Tajima N, Suzuki M, et al. (2012). Identification of substrain-specific mutations by massively parallel whole-genome resequencing of *Synechocystis* sp. PCC 6803. *DNA Res.* 19: 67-79. <http://dx.doi.org/10.1093/dnares/dsr042>
- Kim JE, Oh SK, Lee JH, Lee BM, et al. (2014). Genome-wide SNP calling using next generation sequencing data in tomato. *Mol. Cells* 37: 36-42. <http://dx.doi.org/10.14348/molcells.2014.2241>
- Kumar S, Banks TW and Cloutier S (2012). SNP discovery through next-generation sequencing and its applications. *Int. J. Plant Genomics* 2012: 831460. <http://dx.doi.org/10.1155/2012/831460>
- Li H and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754-1760. <http://dx.doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, et al.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078-2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>

- Lijavetzky D, Cabezas JA, Ibáñez A, Rodríguez V, et al. (2007). High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* 8: 424. <http://dx.doi.org/10.1186/1471-2164-8-424>
- Liu B, Wang Y, Zhai W, Deng J, et al. (2013). Development of InDel markers for *Brassica rapa* based on whole-genome re-sequencing. *Theor. Appl. Genet.* 126: 231-239. <http://dx.doi.org/10.1007/s00122-012-1976-6>
- Lü Y, Cui X, Li R, Huang P, et al. (2015). Development of genome-wide insertion/deletion markers in rice based on graphic pipeline platform. *J. Integr. Plant Biol.* 57: 980-991. <http://dx.doi.org/10.1111/jipb.12354>
- Mammadov J, Aggarwal R, Buyyarapu R and Kumpatla S (2012). SNP markers and their impact on plant breeding. *Int. J. Plant Genomics* 2012: 728398. <http://dx.doi.org/10.1155/2012/728398>
- Natsoulis G, Zhang N, Welch K, Bell J, et al. (2013). Identification of insertion deletion mutations from deep targeted resequencing. *J. Data Mining Genomics Proteomics* 4: 4. <http://dx.doi.org/10.4172/2153-0602.1000132>
- Srivatsan A, Han Y, Peng J, Tehranchi AK, et al. (2008). High-precision, whole-genome sequencing of laboratory strains facilitates genetic studies. *PLoS Genet.* 4: e1000139 <http://dx.doi.org/10.1371/journal.pgen.1000139>.
- Tautz D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17: 6463-6471. <http://dx.doi.org/10.1093/nar/17.16.6463>
- Untergasser A, Cutcutache I, Koressaar T, Ye J, et al. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 40: e115. <http://dx.doi.org/10.1093/nar/gks596>
- Varshney RK, Nayak SN, May GD and Jackson SA (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 27: 522-530. <http://dx.doi.org/10.1016/j.tibtech.2009.05.006>
- Velasco R, Zharkikh A, Affourtit J, Dhingra A, et al. (2010). The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat. Genet.* 42: 833-839. <http://dx.doi.org/10.1038/ng.654>
- Vos P, Hogers R, Bleeker M, Reijans M, et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23: 4407-4414. <http://dx.doi.org/10.1093/nar/23.21.4407>
- Williams JG, Kubelik AR, Livak KJ, Rafalski JA, et al. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18: 6531-6535. <http://dx.doi.org/10.1093/nar/18.22.6531>
- Zhao G, Dai H, Chang L, Yue M, et al. (2010). Isolation of two novel complete Ty1-copia retrotransposons from apple and demonstration of use of derived S-SAP markers for distinguishing bud sports of *Malus domestica* cv. Fuji. *Tree Genet. Genomes* 6: 149-159. <http://dx.doi.org/10.1007/s11295-009-0236-7>